CHEMICAL REVIEWS

REVIEW

pubs.acs.org/CR

Use of ^{13}C NMR Chemical Shift as QSAR/QSPR Descriptor

Rajeshwar P. Verma* and Corwin Hansch

Department of Chemistry, Pomona College, 645 North College Avenue, Claremont, California 91711, United States

CONTENTS

1. INTRODUCTION

The main target of the quantitative structure−activity/property relationship (QSAR/QSPR) paradigm (one of the well-established disciplines in computational chemistry) is to develop a mathematical model that can be used to predict the desired biological activity, chemical reactivity, or physical properties of new, untested, and even forecast compounds from the knowledge of their molecular structures. The QSAR/QSPR technique began with the seminal work of Hansch et al.[1] in 1962 that changed the ways of assessment of chemical−biological interactions. A large set of molecular descriptors, which characterizes the electronic, hydrophobic, steric, and topological (structural) properties of the molecules, is routinely utilized in this methodology. QSAR modeling has become increasingly helpful in understanding mechanisms of action of physiologically active substances, to enhance their effectiveness and reduce the cost of development of the final medicinal products; that is, the QSAR technique can be considered as a low-cost/high-return technique.[2] QSAR models thus can be considered to play a critical role in drug discovery and development by being involved in both the opening and the endgame phases of lead optimization.[3] The selection/utilization of an appropriate statistical methodology and structural descriptors is always vital in the development of a predictive QSAR/QSPR model that encodes the relationship between the structure of a molecule and its biological activity, chemical reactivity, or physical characteristics. It should be borne in mind that the predictive QSAR/QSPR

Received: April 27, 2010
Published: March 25, 2011

models based only on the descriptors calculated from molecular structure are preferable for an early stage of screening in the drug discovery/development processes.

Nuclear magnetic resonance (NMR) spectroscopy is one of the most important techniques, which has long been frequently used not only for the structural elucidation of simple and complicated molecules but also for the detailed explanation of some important processes, including structural configuration, reaction mechanisms, molecular dynamics, chemical equilibrium, structural genomics, and even three-dimensional structures of proteins in aqueous solution.[4−7] In addition, NMR is a nondestructive technique; that is, compounds may always be recovered. Since the chemical information in NMR spectroscopy is encoded in the form of chemical shift, intensity, and multiplicity, these data can be used in the development of gigantic databases.[8] In NMR spectroscopy, the tensor of chemical shift is mainly composed of two terms, such as diamagnetic and paramagnetic, for which, the diamagnetic term is directly related to the electrostatic potential at the nucleus, whereas the paramagnetic term is mainly dependent upon the orbital configuration.[9]

In the case of ^{13}C NMR spectra, there are sufficiently large differences between these two diamagnetic and paramagnetic terms. This suggests that the spectral regions of the orbital configurations for different carbon atoms can be separated very easily from each other. The ^{13}C NMR of a compound provides a specific pattern of frequencies that corresponds directly to the quantum mechanical properties of the nucleus of each carbon atom present in that molecule. Thus, the attached and adjacent carbon atoms must have to show a significant effect on the ^{13}C NMR chemical shift. It is important to note that the ^{13}C NMR chemical shifts have been successfully used to predict the chemical structure of compounds and vice versa.[9−11] A number of computational programs are also commercially available, which can easily and accurately predict the ^{13}C NMR spectra of the chemicals. By considering the features of ^{13}C NMR spectroscopy as well as the recent advancement in this area, its chemical shifts can be used successfully as a molecular descriptor for QSAR/QSPR modeling.

The development of QSAR/QSPR models using ^{13}C NMR chemical shifts as descriptor is usually referred to as the quantitative spectrometric data−activity relationship (QSDAR). Thus, models developed between the ^{13}C NMR spectral data of a set of molecules and their biological activity/chemical reactivity or physical characteristics are considered as QSDAR models. In the past two decades, the potential use of ^{13}C NMR chemical shifts as input parameters to develop models for biological activities, chemical reactivities, and physical properties has successfully been demonstrated by extensive publications. It has been observed from the previous results that the QSDAR model using chemical shifts of ^{13}C NMR works very well when attempted on a set of compounds with a large proportion of carbon nuclei or on similar structural motifs.[12] The recent development of a three-dimensional quantitative spectrometric data−activity relationship (3D-QSDAR) by combining the ^{13}C NMR spectral data and the structural information of compounds in a 3D-connectivity matrix has further enhanced the importance of using the ^{13}C NMR chemical shifts in QSDAR modeling. In 3D-QSDAR modeling, the 3D-connectivity matrix has been built by displaying all the possible assigned chemical shifts of the ^{13}C NMR, such as carbon-to-carbon connections and distances between carbons. This technique is also called comparative structural connectivity spectra analysis (CoSCoSA) modeling.[12,13]

The major two advantages of using the ^{13}C NMR chemical shifts as input parameters in the development of QSAR/QSPR (QSDAR) models are the following: (i) the spectra are taken in liquid solution, similar to the cases of biological systems, and (ii) it is a very sensitive technique, such that a very small change in the molecular shape can be clearly observed as major changes in the spectra. The importance of the QSDAR methodology has further been promoted by some patents.[14−16] Different methods for using the ^{13}C NMR spectral data in QSDAR studies include the followings: (i) sum of the chemical shifts, (ii) averaged chemical shift, (iii) substituent's chemical shift, (iv) chemical shift of a C-atom common to all compounds, (v) difference of chemical shifts between two carbon atoms, (vi) difference of chemical shift of a C-atom common to all compounds from that of the parent, (vii) sum of carbon chemical shift differences in two solvents, (viii) the spectral bins' intensities, and (ix) the principal components.

In the present review, we discuss the following QSAR/QSPR models using the ^{13}C NMR spectral data: classical QSAR/QSPR, spectroscopic data−activity relationship (SDAR), comparative spectral analysis (CoSA), comparative structurally assigned spectral analysis (CoSASA), and CoSCoSA (3D-QSDAR) models, with special emphasis on the classical QSAR/QSPR modeling in order to understand chemical−biological interactions, which may provide strategies that might aid in the drug development processes.

2. MATERIALS AND METHODS

The literature was surveyed to collect all the data (see individual QSAR/QSPR models for their respective references). Although most of the QSAR/QSPR models of this review are taken from the literature, some newly developed models are also added, for which details have been given at the appropriate places. K is the rate constant of a compound for specific reaction types; log K is the subsequent dependent variable that defines the chemical reactivity parameter for the development of the QSAR model and explains where it is used. IC_{50} is the concentration of a compound that causes 50% inhibition toward the specified biological targets, as explained with respective models. Similarly, LD_{50} and MIC represent, respectively, 50% toxicity and the minimal inhibition concentration of a compound toward specific biological responses (see individual QSAR models for their respective biological activities). In the literature, the IC_{50}, LD_{50}, and MIC values are normally given in either μM or nM concentration. For the comparison point of view, all those values of IC_{50}, LD_{50}, and MIC were converted into molar concentrations and mentioned in the appropriate places. In QSAR studies, we often prefer to convert the concentration of a desired effect C to an activity parameter A using the equation $A = -\log C = \log 1/C$. This transformation represents that the more effective compound always has a higher "activity" and vice versa.[2] Thus, IC_{50}, LD_{50}, and MIC in molar concentrations were further transformed into their negative logarithm, i.e., log $1/IC_{50}$, log $1/LD_{50}$, and log 1/MIC, respectively, which were then used subsequently as dependent variables to define the biological parameters for QSAR models. The C-QSAR program was used to develop new QSAR/QSPR models by utilizing multiregression analyses (MRA).[17] In this program, descriptors are autoloaded and selection of descriptors is made by using permutation and correlation matrices among the descriptors in order to avoid the collinearity problems. Details about the C-QSAR program, the search engine,

Table 1. Correlations of the Measured δ_C(C═N) with σ_F(X) and σ_R(X) (Using eq 4) or σ(X) (Using eq 5) for Different Substituted Benzylidene Anilines *p*-X-C_6H_4—CH═N—C_6H_4-*p*-Y (I), and Their Respective ρ_F(X), ρ_R(X), ρ(X), and Correlation Coefficient (r) Values[a]

		correlations of δ_C(C═N) with σ_F(X) and σ_R(X) using eq 4			correlations of δ_C(C═N) with σ(X) using eq 5	
no.	Y	ρ_F(X)	ρ_R(X)	r	ρ(X)	r
1	NO_2	−3.71 ± 0.15	0.12 ± 0.20	0.9889	−1.4 ± 0.5	0.7614
2	CN	−3.81 ± 0.13	0.06 ± 0.18	0.9893	−1.4 ± 0.4	0.7772
3	F	−4.35 ± 0.13	−0.83 ± 0.18	0.9939	−2.0 ± 0.4	0.8706
4	Cl	−4.20 ± 0.13	−0.62 ± 0.18	0.9930	−1.9 ± 0.4	0.8534
5	H	−4.14 ± 0.11	−0.66 ± 0.15	0.9948	−1.9 ± 0.4	0.8609
6	CH_3	−4.42 ± 0.13	−0.94 ± 0.18	0.9940	−2.1 ± 0.4	0.8780
7	OCH_3	−4.82 ± 0.15	−1.46 ± 0.21	0.9943	−2.5 ± 0.5	0.9037
8	$N(CH_3)_2$	−5.68 ± 0.20	−2.54 ± 0.28	0.9939	−3.4 ± 0.5	0.9324

[a] Adapted with permission from ref 38. Copyright 2006 American Chemical Society.

the choice of parameters, and their use in the development of QSAR models have already been discussed in previous publications.[18,19] The details about the physicochemical descriptors are given at the appropriate places. The ^{13}C NMR of the compounds was either taken from the literature or calculated with the ChemBioDraw Ultra 12 program.[20]

In QSAR/QSPR models, n is the number of data points, r^2 is the square of the correlation coefficient (the goodness of fit), q^2 is the cross-validated r^2 (a measure of the quality of the model), and s is the standard deviation. The cross-validated r^2 (q^2) is obtained by using the leave-one-out (LOO) procedure as described by Cramer et al.[21] F is the Fischer statistics and represented by the following equation: $F = fr^2/[(1 - r^2)m]$, where f is the number of degrees of freedom $[f = n - (m + 1)]$, n is the number of data points, m is the number of variables, and r is the correlation coefficient. The F value indicates a true relationship or the significance level for MLR models and is valid at the 95% level.[22] Outliers were filtered on the basis of their deviation between observed and calculated activities from the equation (obsd − pred > 2s).[23,24] All the new QSAR/QSPR models reported here are derived by us that include 95% confidence limits for each term in parentheses.

3. QSPR RESULTS AND DISCUSSION

In this section, a comparative study of the ^{13}C NMR chemical shift with a large set of traditionally and routinely utilized molecular descriptors (including physical properties of the molecules), which characterizes the electronic, hydrophobic, steric, and topological (structural) properties of the compounds, is carried out to illustrate the importance of using the ^{13}C NMR chemical shift as QSPR descriptors.

3.1. Electronic Descriptors

3.1.1. Hammett Electronic Parameters (σ, σ^+, and σ^-) and Its Extended Form (σ_I, σ_R, σ_F, and σ^*). σ, σ^+, and σ^-, the Hammett electronic parameters, are applied to determine the substituent effects on the aromatic systems. The normal σ is used for the substituents on aromatic systems where strong resonance interaction between the substituent and the reaction center does not exist. On the contrary, σ^+ and σ^- are employed where there is a strong resonance interaction between the substituent and the reaction center.[25]

The Hammett equation, which correlates the rate or equilibrium constants of side-chain reactions of *meta*- and *para*-substituted aromatic compounds, is one of the most typical and useful linear free energy relationships to date. It is represented by the following general equation:

$$\log K_X - \log K_H = \rho\sigma \quad (1)$$

where K_H is the rate constant for an unsubstituted aromatic compound, K_X is that for a *meta*- or *para*-substituted aromatic derivative, σ is the substituent constant for the substituent in question, and ρ is the reaction constant.[26,27] The value of unity for the ρ constant in the ionization equilibrium of substituted benzoic acids in water at 25 °C was ascertained on the basis of the experimental results.[26] The term σ is mainly defined as to measure the size of the electronic effect for a given substituent, which represents a measure of the electronic charge distribution in the benzene nucleus. Positive values of σ represent electron withdrawal by the substituent from the aromatic ring while negative values of σ suggest electron release (relative to H) toward the ring. Thus, a positive value of ρ signifies the rate enhancement of a reaction by electron-withdrawal at the reaction site. In contrast, a negative value of ρ suggests that a reaction is assisted by the electron-releasing substituents.[23,28] To see the influence of resonance on the reaction rates, two other constants, i.e. σ^+ (electrophilic substituents) and σ^- (nucleophilic substituents), were introduced and the linear free-energy relationship was then modified as follows:[29,30]

$$\log K_X - \log K_H = \rho^+\sigma^+ \text{ or } \rho^-\sigma^- \quad (2)$$

In order to tune up the substituent effects, the Hammett σ term was then partitioned into two components, i.e. sigma inductive (σ_I) and sigma resonance (σ_R), which is based on Taft's earlier work in aliphatic systems where the resonance effects were absent. This partition led to the development of the following dual-substituent parameter (DSP) equation:[31,32]

$$\log K_X - \log K_H = \rho_I\sigma_I + \rho_R\sigma_R \quad (3)$$

where ρ_I and ρ_R are the sensitivity of a reaction to inductive and resonance effects, respectively. The other electronic parameter, σ_F, was also developed for the substituent field effect. It must be noted that there are no large differences in the σ scale of field/inductive effects for common dipolar

Table 2. Correlations of the Measured δ_C(C═N) with σ_F(Y) and σ_R(Y) (Using eq 4) or σ(Y) or σ^+(Y) (Using eq 5) for Different Substituted Benzylidene Anilines *p*-X—C_6H_4—CH═N—C_6H_4-*p*-Y (I), and Their Respective ρ_F(Y), ρ_R(Y), ρ(Y), ρ^+(Y), and Correlation Coefficient (*r*) Values[a]

		correlations of δ_C(C═N) with σ_F(Y) and σ_R(Y) using eq 4			correlations of δ_C(C═N) with σ(Y) using eq 5		correlations of δ_C(C═N) with σ^+(Y) using eq 5	
no.	X	ρ_F(Y)	ρ_R(Y)	*r*	ρ(Y)	*r*	ρ^+(Y)	*r*
1	NO_2	2.81 ± 0.50	8.49 ± 0.65	0.9857	5.23 ± 0.42	0.9809	3.44 ± 0.05	0.9995
2	CN	2.79 ± 0.47	8.20 ± 0.62	0.9865	5.08 ± 0.40	0.9819	3.34 ± 0.05	0.9994
3	CF_3	3.17 ± 0.47	7.90 ± 0.51	0.9873	4.96 ± 0.43	0.9820	3.07 ± 0.06	0.9991
4	F	2.45 ± 0.35	6.53 ± 0.46	0.9886	4.13 ± 0.90	0.9856	2.71 ± 0.05	0.9988
5	Cl	2.57 ± 0.38	6.92 ± 0.50	0.9882	4.37 ± 0.31	0.9851	2.86 ± 0.05	0.9990
6	H	2.50 ± 0.35	6.52 ± 0.45	0.9890	4.14 ± 0.28	0.9864	2.71 ± 0.06	0.9986
7	CH_3	2.40 ± 0.33	6.19 ± 0.43	0.9888	3.94 ± 0.27	0.9866	2.58 ± 0.05	0.9987
8	OCH_3	2.29 ± 0.31	5.91 ± 0.41	0.9891	3.76 ± 0.25	0.9870	2.46 ± 0.05	0.9986
9	$N(CH_3)_2$	1.99 ± 0.28	5.20 ± 0.37	0.9885	3.30 ± 0.22	0.9868	2.16 ± 0.04	0.9987

[a] Adapted with permission from ref 38. Copyright 2006 American Chemical Society.

substituents.[33] The Hammett approach was found successful only for the substituent effects in aromatic systems and in aliphatic systems does not lead to the development of Taft's aliphatic polar constant, σ^*, which applies to the electronic effects in aliphatic systems.[34]

It is now well established that a substituent can affect a system to which it is bonded by two basic mechanisms of electronic transmission, i.e. through polar and resonance effects.[32,35] In particular, the substituent-induced changes in the chemical shift, SCS, depend strictly on the inductive ($\rho_I\sigma_I$ or $\rho_F\sigma_F$) and resonance ($\rho_R\sigma_R$) parameters as shown in the following DSP equation (eq 4):

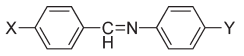
$$\text{SCS} = \rho_I\sigma_I \text{ (or } \rho_F\sigma_F) + \rho_R\sigma_R \quad (4)$$

where SCS is the ^{13}C NMR chemical shift (in ppm) of substituted compound relative to unsubstituted one. Although excellent correlations have always been obtained by the DSP equation (eq 4), single-parameter correlation (eq 5) may also provide good or acceptable results, and thus, it cannot be ignored.[36−38]

$$\text{SCS} = \rho\sigma + \text{constant} \quad (5)$$

In eq 5, the term σ can be replaced by any one of the six different resonance scales (σ^+, σ^-, σ_I, σ_R, σ_F, and σ^*) depending upon the electron demand.

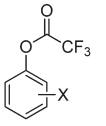
To determine the substituent cross-interaction effects on the electronic character of the C═N bridging group, ^{13}C NMR chemical shifts δ_C(C═N) were measured in $CDCl_3$ for a reasonably large set of substituted benzylidene anilines *p*-X-C_6H_4—CH═N—C_6H_4-*p*-Y (**I**). The substituent dependence of δ_C(C═N) was then used as a tool to determine the electronic substituent effects on the azomethine unit. The correlations of δ_C(C═N) with electronic parameters (σ, σ^+, σ_F, and σ_R) of X or Y substituents were developed using both the dual-substituent parameter and single-parameter correlations (eqs 4 and 5), and results are summarized in Tables 1 and 2.[38] Excellent correlations of experimental δ_C(C═N) with σ (r = 0.7614−0.9324) and σ_F and σ_R (r = 0.9889−0.9948) of X substituents (see Table 1) as well as with σ (r = 0.9809−0.9870), σ^+ (r = 0.9986−0.9995), and σ_F and σ_R (r = 0.9857−0.9891) of Y substituents (see Table 2) illustrate the possibility that the ^{13}C NMR chemical shifts may be used as the significant electronic parameters in the QSAR/QSPR analyses.



I

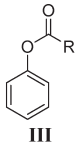
X = NO_2, CN, CF_3, F, Cl, H, CH_3, OCH_3 or $N(CH_3)_2$
Y = NO_2, CN, F, Cl, H, CH_3, OCH_3 or $N(CH_3)_2$

The ^{13}C NMR chemical shifts of the carbonyl carbon δ_C(C═O) for phenyl trifluoroacetates (**II**), acyl-substituted phenyl acetates (**III**), and methyl acetates (**IV**) were measured by the same research group (Neuvonen et al.[36]), and δ_C(C═O) data was then correlated with the substituent constants σ or σ^* of the respective compound series (**II**, **III**, and **IV**). Statistical data for these correlations are summarized in Table 3.



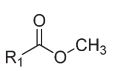
II

X = 4-OCH_3, 4-CH_3, H, 4-Br, 4-Cl, 4-CN, 4-NO_2, 3-Cl or 3-NO_2



III

R = CH_3CH_2, CH_3, $PhCH_2$, CH_3OCH_2, $PhOCH_2$, CH_2Cl, $CHCl_2$ or CCl_3



IV

R_1 = CH_3, CH_3OCH_2, CH_2Cl, $CHCl_2$ or CCl_3

Correlations of δ_C(C═O) with the electronic parameters σ or σ^* for these compound series **II**, **III**, and **IV** are excellent (r^2 = 0.9495, 0.9488, and 0.9883, respectively) (see Table 3).

An excellent correlation between the experimental ^{13}C NMR chemical shift of the cyano carbon δ_C(C≡N) for

Table 3. Statistical Data for Different Correlations of the Measured ^{13}C NMR Chemical Shifts of Carbonyl Carbon [$\delta_C(C{=}O)$] and Substituent Constants σ or σ^* for Phenyl Trifluoroacetates (II), Acyl-Substituted Phenyl Acetates (III), or Methyl Acetates (IV)[a]

no.	compd series	correlations	n	r^2	s	slope
1	II	$\delta_C(C{=}O)$ vs σ	9	0.9495	0.08	−1.03
2	III	$\delta_C(C{=}O)$ vs σ^*	8	0.9488	0.40	−4.00
3	IV	$\delta_C(C{=}O)$ vs σ^*	5	0.9883	0.20	−3.50

[a] Adapted with permission from ref 36. Copyright 2002 American Chemical Society.

Table 4. Experimental ^{13}C NMR Chemical Shift (ppm rel to TMS) of Cyano Carbon $\delta_C(C{\equiv}N)$ for X−CH_2−CN (V) and σ_I Values for Their X Substituents Used To Derive eq 6

		σ_I (eq 6)[b]			
no.	X	obsd	pred	Δ	$\delta_C(C{\equiv}N)$[c]
1[a]	F	0.52	0.05	0.47	118.63
2	Cl	0.47	0.43	0.04	114.22
3	Br	0.44	0.44	0.00	114.08
4	I	0.39	0.31	0.08	115.56
5	OCH_3	0.27	0.29	−0.02	115.79
6	OC_2H_5	0.28	0.26	0.02	116.18
7	SCH_3	0.25	0.33	−0.08	115.41
8	SC_2H_5	0.25	0.25	0.00	116.36
9	$N(CH_3)_2$	0.06	0.04	0.02	118.77
10[a]	$N(C_2H_5)_2$	0.02	0.49	−0.47	113.47
11	CH_3	−0.04	−0.09	0.05	120.24
12	C_2H_5	−0.01	0.09	−0.10	118.11

[a] Not included in the derivation of QSPR eq 6. [b] Calculated/data from the C-QSAR program (ref 17). [c] Data from ref 39.

α-monosubstituted acetonitriles (**V**) and their substituent constants σ_I (after omitting outliers, r = 0.964) was graphically represented by Reis and Rittner.[39] By using their data, we developed eq 6 (Table 4).

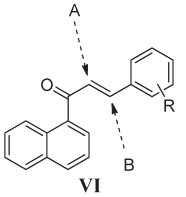
V

$$\sigma_I = -0.09(\pm 0.02)\delta_C(C{\equiv}N) + 10.27(\pm 2.67) \quad (6)$$

where n = 10, r^2 = 0.903, s = 0.059, q^2 = 0.842, and $F_{1,8}$ = 74.474. Outliers: X = F and $N(C_2H_5)_2$.

The negative coefficient of $\delta_C(C{\equiv}N)$ in eq 6 suggests that the σ_I value of the X substituent decreases with an increase in the $\delta_C(C{\equiv}N)$ value, i.e. more deshielded of the −C≡N carbon. Since the shielding effect on the −C≡N carbon depends on the inductive effect of the X substituent, an electron-withdrawing X substituent will cause a decrease in the $\delta_C(C{\equiv}N)$ value (more shielded −C≡N carbon) while an electron-releasing X substituent will cause an increase in the $\delta_C(C{\equiv}N)$ value (more deshielded −C≡N carbon). Two compounds (X = F and $N(C_2H_5)_2$) were omitted on the basis of their deviation (>2s). As expected, the calculated $\delta_C(C{\equiv}N)$ values of halogen derivatives (X = F, Cl, Br, and I) are almost the same, i.e. 114.9, 115.5, 114.9, and 114.9, respectively.[20] But the experimental $\delta_C(C{\equiv}N)$ value of one halogen derivative (X = F) is unexpectedly higher than that of the others (X = Cl, Br, and I). Due to this, the analogue (X = F) has predicted a very much lower σ_I value than that observed, by ∼8 times the standard deviation, and this value has been assigned as an outlier. On the other hand, it is expected that X = $N(CH_3)_2$ and $N(C_2H_5)_2$ will cause a similar downfield shift $\delta_C(C{\equiv}N)$ due to their identical inductive effect. But the experimental $\delta_C(C{\equiv}N)$ value of the compound with X = $N(C_2H_5)_2$ is unexpectedly lower than that of the analogue with X = $N(CH_3)_2$. This is the reason the analogue with X = $N(C_2H_5)_2$ has predicted a very much higher σ_I value than that observed, by ∼8 times the standard deviation, which is deemed another outlier. Since the experimental $\delta_C(C{\equiv}N)$ values are used in this correlation, the exact reason behind the outliers is not very clear.

It was Thirunarayanan[40] who showed that the experimental ^{13}C NMR chemical shifts of α and β carbons of substituted styryl 1-naphthyl ketones (**VI**) are well correlated with the Hammett substituent constants (σ_I and σ_R) as represented by eqs 7 and 8, respectively, with high correlation coefficients (r = 0.998 and 0.999). The assignment of chemical shifts for the ethylene carbons was based on substituted styrene; such as, C-α and C-β represent the α-carbon (A) and β-carbon (B), respectively.



VI

R = H, 3-NH_2, 4-NH_2, 3-Br, 3-Cl, 4-Cl, 4-$N(CH_3)_2$, 4-OH, 4-OCH_3, 4-CH_3, 2-NO_2, 3-NO_2, and 4-NO_2

$$\delta_C(C\text{-}\alpha) = 3.79(\pm 1.37)\sigma_I + 5.18(\pm 2.38)\sigma_R + 123.64(\pm 5.37) \quad (7)$$

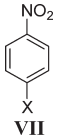
where n = 13, r^2 = 0.996, and $F_{2,10}$ = 1245

$$\delta_C(C\text{-}\beta) = 8.73(\pm 2.57)\sigma_I + 2.10(\pm 3.22)\sigma_R + 144.45 \quad (8)$$

where n = 13, r^2 = 0.998, and $F_{2,10}$ = 2495.

The high correlations between the ^{13}C NMR chemical shift and Hammett substituent constants (σ_I and σ_R) in eqs 7 and 8 suggest that the substituent effect is well distributed between both α and β carbons.

In order to understand the significance for the use of ^{13}C NMR chemical shifts in place of frequently used Hammett electronic parameters (σ, σ^+, and σ^-) in QSAR/QSPR analysis, a comparative QSPR study was carried out using a large data set of 4-X-nitrobenzene (**VII**). Results from those QSPR studies are listed in eqs 9−11 (Table 5). Although good correlations were obtained between the calculated ^{13}C NMR chemical shift of the nitro carbon $\delta_C(C{-}NO_2)$ of 4-X-nitrobenzene (**VII**) and any of the three Hammett electronic parameters (σ, σ^+, or σ^-) of the X substituents, the best correlation was with σ^- (see eq 11).



VII

$$\sigma = 0.07(\pm 0.02)\delta_C(C-NO_2) - 10.53(\pm 2.16) \quad (9)$$

where n = 16, r^2 = 0.890, s = 0.138, q^2 = 0.855, and $F_{1,14}$ = 113.273

$$\sigma^+ = 0.11(\pm 0.02)\delta_C(C-NO_2) - 16.52(\pm 2.95) \quad (10)$$

Table 5. Calculated ^{13}C NMR Chemical Shift (ppm rel to TMS) of Nitro Carbon $\delta_C(C-NO_2)$ for 4-X-Nitrobenzene (VII) as Well as σ, σ^+, and σ^- Values for Their X Substituents Used To Derive eqs 9, 10, and 11

		σ (eq 9)[a]			σ^+ (eq 10)[a]			σ^- (eq 11)[a]			
no.	X	obsd	pred	Δ	obsd	pred	Δ	obsd	pred	Δ	$\delta_C(C-NO_2)$[b]
1	$COOCH_3$	0.45	0.51	−0.06	0.49	0.55	−0.06	0.75	0.82	−0.07	152.20
2	COOH	0.45	0.57	−0.12	0.42	0.66	−0.24	0.77	0.92	−0.15	153.10
3	CHO	0.42	0.62	−0.20	0.47	0.72	−0.25	1.03	0.98	0.05	153.70
4	COC_6H_5	0.43	0.46	−0.03	0.51	0.49	0.02	0.83	0.76	0.07	151.60
5	NO_2	0.78	0.64	0.14	0.79	0.76	0.03	1.27	1.01	0.26	154.00
6	NH_2	−0.66	−0.53	−0.13	−1.30	−1.05	−0.25	−0.63	−0.69	0.06	137.90
7	OCH_3	−0.27	−0.36	0.09	−0.78	−0.79	0.01	−0.26	−0.45	0.19	140.20
8	Cl	0.23	0.06	0.17	0.11	−0.14	0.25	0.19	0.17	0.02	146.00
9	Br	0.23	0.12	0.11	0.15	−0.04	0.19	0.25	0.26	−0.01	146.90
10	CH_3	−0.17	−0.02	−0.15	−0.31	−0.26	−0.05	−0.17	0.05	−0.22	144.90
11	OH	−0.37	−0.34	−0.03	−0.92	−0.76	−0.16	−0.37	−0.41	0.04	140.50
12	F	0.06	−0.12	0.18	−0.07	−0.42	0.35	−0.03	−0.10	0.07	143.50
13	CF_3	0.54	0.43	0.11	0.61	0.44	0.17	0.65	0.72	−0.07	151.20
14	CH_2OH	0.00	0.12	−0.12	−0.04	−0.05	0.01	0.08	0.25	−0.17	146.80
15	C_6H_5	−0.01	0.12	−0.13	0.18	−0.05	−0.13	0.02	0.25	−0.23	146.80
16	CN	0.66	0.51	0.15	0.66	0.55	0.11	1.00	0.82	0.18	152.20

[a] Calculated/data from the C-QSAR program (ref 17). [b] Calculated from the ChemBioDraw Ultra 12 program (ref 20).

where $n = 16$, $r^2 = 0.912$, $s = 0.188$, $q^2 = 0.880$, and $F_{1,14} = 145.091$

$$\sigma^- = 0.11(\pm 0.02)\delta_C(C-NO_2) - 15.27(\pm 2.35) \quad (11)$$

where $n = 16$, $r^2 = 0.936$, $s = 0.150$, $q^2 = 0.916$, and $F_{1,14} = 204.750$

Equations 9−11 suggest that the C1 carbon of 4-X-nitrobenzene (**VII**) is most sensitive to the X substituent and $\delta_C(C-NO_2)$ could explain the electronic substituent effect. The positive coefficient of $\delta_C(C-NO_2)$ indicates that the Hammett electronic parameters (σ, σ^+, or σ^-) of the X substituent increase with a more deshielded C1 carbon. Since the shielding effect on the C1 carbon will depend on both the inductive and resonance effects of the X substituent, an electron-withdrawing substituent will cause a downfield shift (more deshielded) while an electron-donating substituent will cause an upfield shift (more shielded) at the C1 carbon.

3.1.2. Swain−Lupton Constants (*F* and *R*). *F* and *R*, the field and resonance constants, were proposed by Swain and Lupton as significant independent variables for correlating or predicting substituent effects.[41] They evaluated the field effect of the substituent, *F*, in terms of Hammett's parameters (σ_m and σ_p) according to the following equation (eq12):

$$F = a\sigma_m + b\sigma_p \quad (12)$$

In this equation, the coefficients *a* and *b* are evaluated via the least-squares method. σ_m and σ_p are the Hammett's σ constants for the *meta*- and *para*-substituents, respectively. To calculate the value of *R*, Swain and Lupton[41] proposed eq 13 with the assumption that $R = 0$ for $N^+(CH_3)_3$.

$$\sigma_p = \alpha F + R \quad (13)$$

By substituting the following values of $F = 0.89$, $\sigma_p = 0.82$, and $R = 0$ for $N^+(CH_3)_3$ in eq 13, α was found to be 0.921.[42] *F* and *R* were then calculated for any substituent whose σ_m and σ_p values were known.

Excellent correlations were observed by Thirunarayanan[40] between the experimental ^{13}C NMR chemical shifts of either the C-α or C-β carbons of substituted styryl 1-naphthyl ketones (**VI**) and the Swain and Lupton constants (*F* and *R*) of the substituents as shown in eqs 14 and 15.

$$\delta_C(C\text{-}\alpha) = 5.71(\pm 1.32)F + 1.70(\pm 2.04)R + 123.44(\pm 5.03) \quad (14)$$

where $n = 13$, $r^2 = 0.986$, and $F_{2,10} = 352.143$

$$\delta_C(C\text{-}\alpha) = 2.48(\pm 1.25)F + 2.00(\pm 0.79)R + 121.94(\pm 0.63) \quad (14a)$$

where $n = 13$, $r^2 = 0.925$, $s = 0.382$, $q^2 = 0.875$, and $F_{2,10} = 61.667$

$$\delta_C(C\text{-}\beta) = 6.30(\pm 2.87)F + 2.19(\pm 1.19)R + 143.95(\pm 192.00) \quad (15)$$

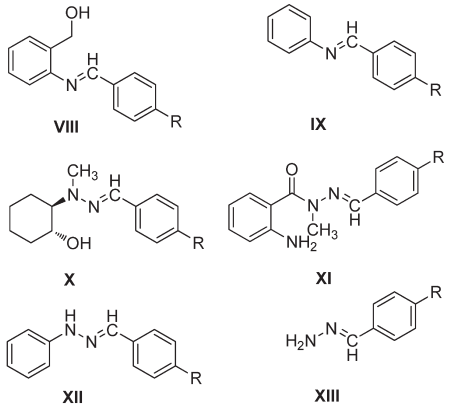
where $n = 13$, $r^2 = 0.996$, and $F_{2,10} = 1245$ (Note: Since eq 14 from the original publication contains one parameter (*R*) with coefficient considerably smaller than their corresponding standard deviation, the data of Thirunarayanan[40] was examined and developed eq 14a.)

The high correlations between the ^{13}C NMR chemical shift and the Swain and Lupton constants (*F* and *R*) of the substituents in eqs 14a and 15 suggest the possibility the high degree of transmission of substituent effects in both α and β carbons.

3.1.3. Atomic Charge (*q*). Although NMR shielding is not determined only by the electron density, satisfactory linear correlations with positive slopes between the atomic charges and the ^{13}C NMR chemical shifts were observed in several systems when the substitution was varied.[36,43−45] This similarity supports the idea that the substituent effects on both the atomic charges and the chemical shifts reflect the electronic substituent effects.

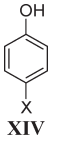
Statistical data for the different correlations of the experimental ^{13}C NMR chemical shifts of the C═O carbon of

compounds **III** or the C═N carbon of compounds **VIII**−**XIII** and the atomic charges of their respective carbons are summarized in Table 6.[36,43]



R = NO_2, CN, CF_3, F, Cl, Br, H, CH_3, OCH_3 or $N(CH_3)_2$

3.1.4. Bond Dissociation Energy (BDE). Bond dissociation energy (BDE), the measure of bond strength in a chemical bond, is defined as the standard enthalpy change when a bond is cleaved by homolysis at a specified temperature. It was observed in our earlier work[46] that the calculated OH homolytic bond dissociation energies can serve as an electronic parameter for phenols, which is further supported by an excellent correlation between BDE and the σ^+ of the X-substituents (n = 19, r = 0.982) for a series of 4-X-phenols (**XIV**) (see Table 7).



XIV

The same data set of 4-X-phenols (**XIV**) (Table 7) was then used to develop a correlation (eq 16) between the calculated BDE and the calculated ^{13}C NMR chemical shift of the phenolic carbon δ_C(C−OH).

$$\mathrm{BDE} = 0.76(\pm 0.07)\delta_C(\mathrm{C-OH}) - 120.95(\pm 10.54) \quad (16)$$

where n = 18, r^2 = 0.972, s = 0.368, q^2 = 0.967, and $F_{1,16}$ = 555.429. Outlier: X = NH_2.

In eq 16, the strong correlation between BDE and δ_C(C−OH) suggests that the ^{13}C NMR chemical shift of the phenolic carbon δ_C(C−OH) can be considered as an electronic parameter for 4-X-phenols. Since the BDE has already shown an excellent correlation with σ^+ for this data set, the idea may generate that the ^{13}C NMR chemical shift can be used as an electronic descriptor in either the place of the σ^+ or BDE parameter in QSAR/QSPR modeling. One compound (X = NH_2) was deemed to be an outlier because this analogue has a far smaller BDE value than that expected by 4.8 times the standard deviation. The exact reason for this deviation is not very clear, but it may be due to the calculated values of either BDE or δ_C(C−OH).

3.1.5. E_{HOMO}, E_{LUMO}, and E_{HOMO} − E_{LUMO} (H−L Gap). According to the frontier orbital theory, the energy of the highest occupied molecular orbital (E_{HOMO}) and the lowest unoccupied molecular orbital (E_{LUMO}) can determine the kinetics of the chemical reaction.[47] In recent years, the E_{HOMO}, E_{LUMO}, and E_{HOMO} − E_{LUMO} (H−L gap) have been utilized successfully as the electronic descriptors in various QSAR/QSPR models.

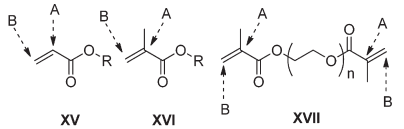
Table 6. Statistical Data for Different Correlations of the Measured ^{13}C NMR Chemical Shifts of the Carbonyl Carbon [δ_C(C═O)] of Acyl-Substituted Phenyl Acetates (III) or the C═N Carbon [δ_C(C═N)] of Compounds VIII−XIII and the Atomic Charges of Their Respective Carbons (q_C)[a]

no.	compd series	correlations	n	r^2	slope
1	III[b]	δ_C(C═O) vs q_C(C═O)	8	0.9443	343
2	VIII	δ_C(C═N) vs q_C(C═N)	10	0.6711	52 ± 14
3	IX	δ_C(C═N) vs q_C(C═N)	10	0.7597	69 ± 14
4[c]	X	δ_C(C═N) vs q_C(C═N)	9	0.9620	95 ± 9
5	XI	δ_C(C═N) vs q_C(C═N)	10	0.8614	82 ± 13
6	XII	δ_C(C═N) vs q_C(C═N)	10	0.8339	76 ± 17
7	XIII	δ_C(C═N) vs q_C(C═N)	10	0.9276	99 ± 16

[a] Adapted with permission from ref 43. Copyright 2003 American Chemical Society. [b] Data from ref 36. [c] One compound (R = CN) was excluded.

The E_{HOMO} is associated with the capacity of a molecule to donate electrons to an acceptor with an empty molecular orbital. Thus, the molecule with higher E_{HOMO} will have the higher electron donating ability and tends to be more reactive. On the contrary, the E_{LUMO} indicates the ability of the molecule to accept electrons. Thus, the molecule associated with a low E_{LUMO} value will be a more powerful electron acceptor than a molecule with a higher value. This is the reason the E_{LUMO} value increases by the influence of electron donating groups (EDG) such as NH_2 or OH and decreases by the presence of electron withdrawing groups (EWG) such as NO_2, SO_3H, or CN.[48,49] The E_{HOMO} − E_{LUMO} is used as a descriptor to define the chemical hardness, which is an important stability and reactivity index. The hardness of a molecule is also qualitatively related to their polarizability, since a decrease of the energy gap usually leads to easier polarization of the molecule. Thus, the large E_{HOMO}–E_{LUMO} value suggests a stable and less reactive molecule, while the low value indicates an unstable and highly reactive molecule.[48,50]

In a very recent publication, a reasonably large data set of aliphatic acrylates (**XV**), methacrylates (**XVI**), and dimethacrylates (**XVII**) was published by Ishihara and Fujisawa[51] along with their ^{13}C NMR spectra, biological activities, and different physicochemical parameters (either collected from the literature[52,53] or generated by them). These authors[51] also published some QSPR/QSAR models either for very limited compounds with excellent correlations or highly sacrificed in terms of correlation coefficient. The best correlation between the experimental ^{13}C NMR and calculated E_{HOMO} from their publication was eq 17.



XV **XVI** **XVII**

R = CH_3, C_2H_5, n-C_3H_7, n-C_4H_9, i-C_4H_9, $(CH_2)_{11}CH_3$, CH_2CH_2OH, $CH_2CH(OH)CH_3$, and $CH_2C_6H_5$; n = 1, 2, and 3

$$\delta_C(\mathrm{C}\text{-}\alpha) - \delta_C(\mathrm{C}\text{-}\beta) = 17.83(\pm 0.70)E_{HOMO} + 198.72(\pm 0.63) \quad (17)$$

where n = 13, r^2 = 0.983, and $F_{1,11}$ = 636.059

dx.doi.org/10.1021/cr100125d |*Chem. Rev.* 2011, 111, 2865–2899

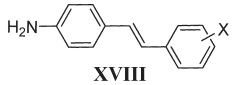
Table 7. Calculated ^{13}C NMR Chemical Shifts (ppm rel. to TMS) of Phenolic Carbon δ_C(C−OH) and C4 Carbon δ_C(C−X), Calculated BDE Values of OH Bonds, $E_{HOMO} - E_{LUMO}$, and Electronegativity (χ) Values for 4-X-Phenols (XIV) Used To Derive Equations 16, 19, and 21 as Well as σ^+ Values for X-Substituents

		BDE (eq 16)			$E_{HOMO} - E_{LUMO}$ (eq 19)			χ (eq 21)					
no.	X	obsd[a]	pred[b]	Δ	obsd[c]	pred[b]	Δ	obsd[d]	pred[b]	Δ	δ_C(C−OH)[e]	δ_C(C−X)[e]	$(\sigma^+)^b$
1	OCH_3	−6.01	−5.71	−0.30	8.96	8.95	0.01	4.17	4.21	−0.04	150.80	153.20	−0.78
2	OC_2H_5	−6.16	−6.24	0.08	8.95	8.90	0.05	4.14	4.15	−0.01	150.10	152.00	−0.81
3	OC_3H_7	−6.23	−6.24	0.01	8.95	8.90	0.05	4.13	4.15	−0.02	150.10	152.00	−0.83
4	OC_4H_9	−6.27	−6.24	−0.03	8.95	8.90	0.05	4.13	4.15	−0.02	150.10	152.00	−0.81
5	OC_6H_{13}	−6.30	−6.24	−0.06	8.95	8.90	0.05	4.13	4.15	−0.02	150.10	152.00	−0.81
6	H	0.00	0.18	−0.18	9.51	8.54	−0.03	4.36	4.25	0.11	158.50	121.30	0.00
7	F	−1.99	−2.88	0.89	9.15	9.23	−0.08	4.52	4.47	0.05	154.50	155.50	−0.07
8[f]	NH_2	−9.25	−7.47	−1.78	8.71	8.78	−0.07	3.92	3.90	0.02	148.50	141.00	−1.30
9	OH	−6.04	−5.48	−0.56	8.95	8.98	−0.03	4.25	4.20	0.05	151.10	151.10	−0.92
10	CH_3	−2.22	−2.12	−0.10	9.32	9.31	0.01	4.22	4.20	0.02	155.50	131.00	−0.31
11	C_2H_5	−1.90	−1.96	0.06	9.31	9.33	−0.02	4.20	4.29	−0.09	155.70	136.90	−0.30
12	OC_6H_5	−4.55	−5.10	0.55	8.91	9.01	−0.10	4.34	4.21	0.13	151.60	149.60	−0.50
13[g]	$C(CH_3)_3$	−1.54	−2.19	0.65	9.36	9.30	0.06	4.22	4.37	−0.15	155.40	143.90	−0.26
14	C_3H_7	−2.01	−1.89	−0.12	9.34	9.33	0.01	4.23	4.26	−0.03	155.80	134.60	−0.29
15	C_4H_9	−2.08	−1.96	−0.12	9.34	9.33	0.01	4.24	4.27	−0.03	155.70	135.20	−0.29
16	C_5H_{11}	−2.13	−1.96	−0.17	9.34	9.33	0.01	4.24	4.27	−0.03	155.70	135.20	−0.29
17	C_7H_{15}	−2.17	−1.96	−0.21	9.34	9.33	0.01	4.23	4.27	−0.04	155.70	135.20	−0.29
18	C_8H_{17}	−2.17	−1.96	−0.21	9.34	9.33	0.01	4.24	4.27	−0.03	155.70	135.20	−0.29
19	C_9H_{19}	−2.17	−1.96	−0.21	9.35	9.33	0.02	4.24	4.27	−0.03	155.70	135.20	−0.29

[a] Data from ref 46. [b] Calculated/Data from the C-QSAR program (ref 17). [c] Data from ref 55. [d] Calculated by eq 20 using data from ref 55. [e] Calculated using the ChemBioDraw Ultra 12 program (ref 20). [f] Not included in the derivation of QSPR eq 16. [g] Not included in the derivation of QSPR eq 21.

δ_C(C-α) and δ_C(C-β) represent the ^{13}C NMR chemical shifts of the ethylene carbons (A) and (B), respectively, of compounds (**XV**−**XVII**). Equation 17 is an excellent correlation between the difference between the ^{13}C NMR chemical shifts of the ethylene carbons [δ_C(C-α) − δ_C(C-β)] and E_{HOMO}. The positive coefficient of E_{HOMO} suggests that the E_{HOMO} of the compounds (**XV**−**XVII**) increases with an increase in the ^{13}C NMR chemical shifts difference of the ethylene carbons. Since the E_{HOMO} is the electron donating capacity of the molecules, eq 17 supports the idea that the substituent effects on both the α and β carbon chemical shifts reflect the electronic substituent effects and their difference can estimate the electron donating capacity of the molecules for this data set. Thus, the [δ_C(C-α) − δ_C(C-β)] value of the molecules can either be used to calculate E_{HOMO} or be used as an electronic parameter for this data set.

The published data of You et al.[54] for the E_{LUMO} of 4-amino-X-stilbenes (**XVIII**) and the calculated ^{13}C NMR chemical shifts of their substituted carbon δ_C(C−X) in Table 8 were used to develop QSPR eq 18.



XVIII

$$E_{LUMO} = 0.018(\pm 0.007)\delta_C(C - X) - 2.715(\pm 0.933) \quad (18)$$

where $n = 8$, $r^2 = 0.864$, $s = 0.107$, $q^2 = 0.771$, and $F_{1,6} = 50.824$
outlier: X = 4′-NO_2
E_{LUMO} vs σ^-; $r = 0.970$
δ_C(C−X) vs σ^-; $r = 0.875$

Since E_{LUMO} is associated with the molecular ability to accept electrons, eq 18 suggests that the substituent effects on the chemical shifts of substituted carbon δ_C(C−X) can measure the electron accepting capability of the molecules (**XVIII**). The high correlations between E_{LUMO} and σ^- and between δ_C(C−X) and σ^- suggest that δ_C(C−X) can measure the substituent electronic effects. The positive coefficient with δ_C(C−X) in eq 18 suggests that the E_{LUMO} of the compounds (**XVIII**) increases with an increase in the value of δ_C(C−X); that is, the ^{13}C NMR chemical shifts of the substituted carbon will be more deshielded. One outlier (X = 4′-NO_2) is more predicted than observed by about 10 times the standard deviation. The exact reason for this deviation is not very clear, but it may be due to the unexpected calculated values of either δ_C(C−X) or E_{LUMO} for the analogues (X = 4′-NO_2).

The data of 4-X-phenols (**XIV**) in Table 7 was further used to develop a correlation (eq 19) between the (E_{HOMO}−E_{LUMO})[55] and the calculated ^{13}C NMR chemical shift of the phenolic carbon δ_C(C−OH).

$$E_{HOMO} - E_{LUMO} = 0.076(\pm 0.006)\delta_C(C - OH) - 2.493(\pm 1.208) \quad (19)$$

where $n = 19$, $r^2 = 0.961$, $s = 0.047$, $q^2 = 0.950$, and $F_{1,17} = 418.897$
$E_{HOMO} - E_{LUMO}$ vs σ^+; $r = 0.909$
$E_{HOMO} - E_{LUMO}$ vs BDE; $r = 0.963$
BDE vs σ^+; $r = 0.982$

Equation 19, an excellent correlation between E_{HOMO}−E_{LUMO} and δ_C(C−OH), suggests that the chemical shifts of the C1 carbon can measure the $E_{HOMO} - E_{LUMO}$ of 4-X-phenols. The positive coefficient of δ_C(C−OH) indicates that

Table 8. Calculated ^{13}C NMR Chemical Shift (ppm rel to TMS) for the Substituted Carbon δ_C(C−X) of 4-Amino-X-stilbenes (XVIII) and Their Calculated E_{LUMO} Used To Derive eq 18 as Well as σ^- Values for X- Substituents

		E_{LUMO} (eq 18)				
no.	X	obsd[a]	pred[b]	Δ	[δ_C(C−X)][c]	(σ^-)[b]
1	4′-NH_2	−0.03	−0.09	0.06	147.60	−0.63
2	4′-CH_3	−0.24	−0.27	0.03	137.60	−0.17
3	3′-NH_2	−0.13	−0.06	−0.07	149.30	−0.16
4	4′-H	−0.26	−0.44	0.18	127.90	0.00
5	4′-Cl	−0.44	−0.34	−0.10	133.50	0.19
6	3′-Cl	−0.40	−0.33	−0.07	134.20	0.37
7	3′-CN	−0.67	−0.72	0.05	112.50	0.56
8	4′-CN	−0.82	−0.73	−0.09	111.80	1.00
9[d]	4′-NO_2	−1.24	−0.10	−1.14	147.10	1.27

[a] Data from ref 54. [b] Calculated/Data from C-QSAR program (ref 17). [c] Calculated by using ChemBioDraw Ultra 12 program (ref 20). [d] Not included in the derivation of QSPR eq 18

the $E_{HOMO} - E_{LUMO}$ of 4-X-phenols increases with an increase in the value of δ_C(C−OH) and subsequently the reactivity of the molecule decreases. The high mutual correlations between $E_{HOMO} - E_{LUMO}$ and σ^+, $E_{HOMO} - E_{LUMO}$ and BDE, and BDE and σ^+ support the idea that δ_C(C−OH) can be considered as an electronic parameter for 4-X-phenols.

3.1.6. Electronegativity (χ). The experimental electronegativity, χ, is found to be in good agreement with the negative averaged value of the calculated HOMO/LUMO energies as shown in eq 20.[49,51]

$$\chi = -\frac{E_{HOMO} + E_{LUMO}}{2} \quad (20)$$

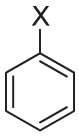
Equation 20 was used to calculate the electronegativity (χ) of 4-X-phenols (**XIV**) from the published data of their E_{HOMO} and E_{LUMO} values[55] (see χ values in Table 7). A correlation (eq 21) was then developed between electronegativity and the calculated ^{13}C NMR chemical shifts.

$$\chi = 0.061(\pm 0.017)\delta_C(C-OH) + 0.014(\pm 0.005)\delta_C(C-X) - 7.087(\pm 3.236) \quad (21)$$

where n = 18, r^2 = 0.796, s = 0.058, q^2 = 0.678, and $F_{2,15}$ = 29.265
outlier: X = $C(CH_3)_3$
χ vs σ^+; r = 0.788
χ vs BDE; r = 0.711
δ_C(C−OH) vs σ^+; r = 0.935
δ_C(C−X) vs σ^+; r = 0.590

δ_C(C−OH) and δ_C(C−X) are the ^{13}C NMR chemical shifts of the C1 and C4 carbons, respectively. Equation 21 suggests that the electronegativity (χ) of 4-X-phenols (**XIV**) can be calculated by using the chemical shifts of their C1 and C4 carbons. The high mutual correlations between σ^+ and χ, σ^+ and δ_C(C−OH), σ^+ and δ_C(C−X), and χ and BDE suggest the idea that δ_C(C−OH) and δ_C(C−X) together can be used as an electronic parameter for 4-X-phenols and may be the substitute for the electronegativity descriptor of 4-X-phenols. One molecule (X = $C(CH_3)_3$) was deemed to be an outlier because this analogue has a lower χ value than expected by 2.6 times the standard deviation. The reason for this deviation is due to the more downfield shift of the C4 carbon attached with the $C(CH_3)_3$ group than that of the n-C_4H_9 ($\Delta\delta$ = 143.90 − 135.20 = 8.70). This may be due to a large effect coming from electrons in the p-orbital with the branching $C(CH_3)_3$ group. Since this outlier (X = $C(CH_3)_3$) is well predicted by eqs 16 and 19, the reason due to the calculated χ value cannot be ruled out.

3.1.7. Ionization Potential (IP). Ionization potential (IP) is an important property of the molecules for the quantitative estimation of their energy state in a redox process as well as in other physical, chemical, and biochemical processes. An excellent linear correlation between the first ionization potential (IP)[56−58] and the experimental ^{13}C NMR chemical shift of the para-carbon δ_C(C-4)[59] for a series of monosubstituted benzenes (**XIX**) was graphically demonstrated by Jørgensen and Dalgaard.[60] By using the same data in Table 9, we developed eq 22.



XIX

$$IP = 0.11(\pm 0.02)\delta_C(C\text{-}4) - 5.46(\pm 2.27) \quad (22)$$

where n = 15, r^2 = 0.935, s = 0.204, q^2 = 0.914, and $F_{1,13}$ = 187.000
IP vs σ^+; r = 0.953
δ_C(C-4) vs σ^+; r = 0.967

Anilines (X = NH_2, $NHCH_3$, and $N(CH_3)_2$) with relatively low chemical shifts δ_C(C-4) suggest that the C-4 position is relatively more shielded. On the contrary, the C-4 position of Ph-X (**XIX**; X = NO_2, CN, CHO, COOH, and COCl) is more deshielded and subsequently increases the value of δ_C(C-4). The above eq 22 is an excellent correlation between IP and δ_C(C-4), for which the positive coefficient of δ_C(C-4) indicates that the IP of Ph-X increases with an increase in the value of δ_C(C-4), i.e. more deshielded. Thus, the IP of Ph-X (**XIX**) increases with electron withdrawing X substituents. The high mutual correlation between IP and σ^+ and between δ_C(C-4) and σ^+ supports the idea that δ_C(C-4) may serve as an electronic parameter for the monosubstituted benzenes.

The published data of Sakamoto and Watanabe[61] for the averaged of measured ^{13}C NMR chemical shifts ($\delta_{C\text{-ave}}$) over all the ring carbons of polycyclic aromatic hydrocarbons (PAHs) and their ionization potentials in Table 10 were used to develop the QSPR model (eq 23).

$$IP = 0.59(\pm 0.15)\delta_{C\text{-ave}} - 68.02(\pm 19.04) \quad (23)$$

where n = 14, r^2 = 0.863, s = 0.376, q^2 = 0.778, and $F_{1,12}$ = 75.591
outlier: triphenylene

The averaged ^{13}C NMR chemical shifts ($\delta_{C\text{-ave}}$) of PAHs are considered to be an index to measure the electronic structure of aromatic compounds.[61] Triphenylene was deemed to be an outlier because this compound has a higher IP value than expected, by ~3 times the standard deviation. The exact reason for this deviation is not very clear, but it may be due to faulty calculation of its IP value.

3.1.8. pK_a Values. The ionization ability of an acid has been quantified by the parameter pK_a. It can be calculated from the K_a values (also known as the protonation constant, equilibrium

Table 9. The First Ionization Potential (IP) in eV and the Experimental ^{13}C NMR Chemical Shift of the para-Carbon δ_C(C-4) (ppm rel. to TMS) for a Series of Mono-Substituted Benzenes (XIX) Used To Derive eq 22

		IP (eq 22)			
no.	X	obsd[a]	pred[b]	Δ	δ_C(C-4)[c]
1	H	9.25	9.12	0.13	128.50
2	CH_3	8.83	8.78	0.05	125.50
3	NH_2	8.00	8.01	−0.01	118.70
4	OH	8.70	8.27	0.43	121.00
5	SH	8.49	8.77	−0.28	125.40
6	Cl	9.10	8.89	0.21	126.50
7	Br	9.02	8.92	0.10	126.70
8	NO_2	9.92	9.81	0.11	134.60
9	CN	9.71	9.60	0.11	132.70
10	$NHCH_3$	7.73	7.78	−0.05	116.70
11	CHO	9.52	9.77	−0.25	134.20
12	COOH	9.73	9.71	0.02	133.70
13	COOCl	9.70	9.89	−0.19	135.30
14	$N(CH_3)_2$	7.51	7.76	−0.25	116.50
15	OCH_3	8.21	8.35	−0.14	121.70

[a] Data from refs 56−58. [b] Calculated from the C-QSAR program (ref 17). [c] Data from ref 59.

constant, or acid dissociation constant) and obtained from the equation $pK_a = -\log_{10}(K_a)$. Along with the integrity, lipophilicity, solubility, and permeability, pK_a has been considered as one of the five key physiochemical parameters to provide an early understanding toward the key properties that affect ADME characteristics.[62] Hammett was the first who quantized electronic effects in aromatic systems and explained the pK_a values of the substituted benzoic acids.[26] A measure of electronic effects can be obtained by the difference between the pK_a values for substituted and unsubstituted benzoic acids:

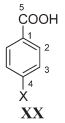
$$\log K_X - \log K_H = -(pK_a)_X + (pK_a)_H = \rho\sigma \qquad (24)$$

The above Hammett equation describes the effect of *meta*- and *para*-substituents on the acidity of substituted benzoic acids. In eq 24, K_X is the rate constant for a substituted benzoic acid, K_H is the rate constant for unsubstituted benzoic acid, σ is the substituent constant, and ρ is a "reaction constant" which depends upon the nature and conditions of the reaction. The Hammett σ constants, from the work of Hollingsworth et al.,[63] exhibited a very strong correlation with the experimental pK_a values of substituted benzoic acids:

$$pK_a = -1.005(\pm 0.011)\sigma + 4.200(\pm 0.004) \qquad (25)$$

where $n = 17$, $r^2 = 0.9985$, $s = 0.0126$, and $F_{1,15} = 8788$

The calculated ^{13}C NMR chemical shift of the C1 carbon of 4-X-benzoic acids (**XX**) was plotted against their experimental pK_a values[64] in order to develop a correlation (eq 26).[65]



XX

X = $NHCH_3$, $N(CH_3)_2$, NH_2, OCH_3, OH, F, CH_3, C_2H_5, $CH(CH_3)_2$, Cl, Br, H, CN, COOH, and NO_2

$$pK_a = -0.0891(\pm 0.013)\delta_{C1} + 15.61(\pm 1.701) \qquad (26)$$

Table 10. Average of Measured ^{13}C NMR Chemical Shifts over All the Ring Carbons (ppm rel. to TMS) of Polycyclic Aromatic Hydrocarbons $\delta_{C\text{-ave}}$ and Their IP Values Used To Derive eq 23

		IP (eq 23)			
no.	compound name	obsd[b]	pred[c]	Δ	$\delta_{C\text{-ave}}$[b]
1	naphthalene	8.13	7.89	0.24	128.28
2	anthracene	7.43	7.83	−0.40	128.17
3	phenanthrene	8.19	8.13	0.06	128.68
4	benz[α]anthracene	7.54	7.65	−0.11	127.86
5	pyrene	7.50	7.23	0.27	127.16
6[a]	triphenylene	8.15	7.03	1.12	126.82
7	dibenz[α,*h*]anthracene	7.58	7.63	−0.05	127.83
8	benzo[α]pyrene	7.23	7.20	0.03	127.11
9	9,10-dimethylanthracene	7.14	7.17	−0.03	127.06
10	7-methylbenz[α]anthracene	7.37	7.41	−0.04	127.47
11	7,12-dimethylbenz[α]anthracene	7.22	7.65	−0.43	127.87
12c	chlorobenzene	9.42	8.78	0.64	129.77
13	*o*-dichlorobenzene	9.06	9.79	−0.73	131.48
14	*m*-dichlorobenzene	9.80	9.30	0.50	130.65
15	*p*-dichlorobenzene	9.67	9.63	0.04	131.21

[a] Not included in the derivation of QSPR eq 23. [b] Data from ref 61. [c] Calculated from the C-QSAR program (ref 17).

Table 11. Experimental pK_a Values of 4-X-Benzoic Acids (XX) and σ^+ Values for Their X Substituents Used To Derive eq 27

		pK_a (eq 27)			
no.	X	obsd[a]	pred[b]	Δ	(σ^+)[b]
1	$NHCH_3$	5.04	5.17	−0.13	−1.81
2	$N(CH_3)_2$	5.03	5.10	−0.07	−1.70
3	NH_2	4.92	4.85	0.07	−1.30
4	OCH_3	4.47	4.52	−0.05	−0.78
5	OH	4.58	4.61	−0.03	−0.92
6	F	4.14	4.08	0.06	−0.07
7	CH_3	4.34	4.23	0.11	−0.31
8	C_2H_5	4.35	4.22	0.13	−0.30
9	$CH(CH_3)_2$	4.35	4.21	0.14	−0.28
10	Cl	3.99	3.97	0.02	0.11
11	Br	4.00	3.94	0.06	0.15
12	H	4.17	4.04	0.13	0.00
13	CN	3.55	3.62	−0.07	0.66
14	COOH	3.51	3.77	−0.26	0.42
15	NO_2	3.44	3.54	−0.10	0.79

[a] Data from ref 64. [b] Calculated/Data from the C-QSAR program (ref 17).

where $n = 15$, $r^2 = 0.9412$, $s = 0.129$, $q^2 = 0.928$, and $F_{1,13} = 207.339$ (Note: Statistical data and the 95% confidence intervals within parentheses were added to the original eq 26.)

From eq 26, it is clear that the C1 carbon of 4-X-benzoic acids (**XX**) is most sensitive to X substituents and its ^{13}C NMR chemical shift (δ_{C1}) could represent the electronic substituent effect. The negative coefficient of δ_{C1} in eq 26 suggests that the pK_a values of 4-X-benzoic acids (**XX**) increase with more

Table 12. Calculated Averaged ^{13}C NMR Chemical Shifts ($\delta_{C\text{-ave}}$) Over All the Ring Carbons of X-Chlorobenzene (XXI) as Well as Their Polarizabilities (α in Å^3) and Other Physicochemical Descriptors Used To Derive eqs 29, 30, and 35

		α (eq 29)			NVE (eq 30)[c]			CMR (eq 35)[c]				
no.	X	obsd[b]	pred[c]	Δ	obsd	pred	Δ	obsd	pred	Δ	$\delta_{C\text{-ave}}$[a]	(MgVol)[c]
1	H	10.48	10.74	−0.26	36	36.98	−0.98	3.18	3.26	−0.08	129.82	0.84
2	2-Cl	12.14	12.09	0.05	42	41.45	0.55	3.67	3.63	0.04	130.60	0.96
3	3-Cl	12.30	12.47	−0.17	42	42.71	−0.71	3.67	3.73	−0.06	130.82	0.96
4	4-Cl	12.35	12.66	−0.31	42	43.34	−1.34	3.67	3.78	−0.11	130.93	0.96
5	2,3-Cl_2	13.89	13.39	0.50	48	45.75	2.25	4.16	3.98	0.18	131.35	1.08
6	2,4-Cl_2	14.11	13.69	0.42	48	46.73	1.27	4.16	4.06	0.10	131.52	1.08
7	3,5-Cl_2	14.22	14.87	−0.65	48	50.62	−2.62	4.16	4.38	−0.21	132.20	1.08
8	2,3,4-Cl_3	15.75	15.26	0.49	54	51.94	2.06	4.65	4.49	0.17	132.43	1.21
9	2,3,5-Cl_3	15.92	17.12	−1.20	54	58.07	−4.07	4.65	4.99	−0.33	133.50	1.21
10	2,4,5-Cl_3	15.98	15.26	0.72	54	51.94	2.06	4.65	4.49	0.17	132.43	1.21
11	2,3,4,5-Cl_4	17.70	18.42	−0.72	60	62.37	−2.37	5.15	5.34	−0.19	134.25	1.33
12	2,3,4,5,6-Cl_5	19.48	18.33	1.15	66	62.08	3.92	5.64	5.32	0.32	134.20	1.45

[a] Calculated from the ChemBioDraw Ultra 12 program (ref 20). [b] Data from ref 67. [c] Calculated from the C-QSAR program (ref 17).

shielded C1 carbons while they decrease with more deshielded C1 carbons. The shielding effect on the C1 carbon depends on both the inductive and resonance effects of the substituent. An electron-withdrawing group (EWG) is characteristic of an −I (inductive) and resonance effects, causing a more downfield shift (more deshielded) in the C1 carbon. On the contrary, an electron-donating group (EDG) is characteristic of a +I and resonance effects, causing a more upfield shift (more shielded) in the C1 carbon. Thus, it can be suggested that the pK_a values of 4-X-benzoic acids (**XX**) will increase by the influence of a more shielded C1 carbon or electron-releasing X substituents. This finding has further been supported by the development of eq 27 (using the same pK_a values of eq 26); that is, electron-releasing X substituents will increase the pK_a values of 4-X-benzoic acids (**XX**), as suggested by the negative ρ value of the Hammett σ^+ of X substituents (see eq 27 below and their data in Table 11). There is also an excellent mutual correlation between δ_{C1} and σ^+ (r = 0.958) for this data set.

$$pK_a = -0.627(\pm 0.087)\sigma^+ + 4.035(\pm 0.074) \quad (27)$$

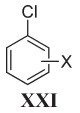
where n = 15, r^2 = 0.949, s = 0.120, q^2 = 0.931, and $F_{1,13}$ = 241.902

On comparing eqs 26 and 27, one can say that eq 26 may provide an alternative to evaluate pK_a values, especially for those substituted derivatives for which the substituent constants (σ^+ values) are unknown.

3.1.9. Polarizability (α) and NVE. The polarizability (α) of a molecule is a significant descriptor that is frequently used in QSAR and QSPR analysis. It is defined as the proportionality constant between the strength of an applied electrical field (E) and the magnitude of the induced dipole moment ($\mu_{induced}$).

$$\mu_{induced} = \alpha \times E \quad (28)$$

The molecular polarizability of X-chlorobenzene (**XXI**) obtained from density functional theory was published by Staikova et al.[66] We used their data to develop QSPR (eq 29) between polarizability (α in Å^3) and the averaged ^{13}C NMR chemical shifts over all the ring carbons of X-chlorobenzene ($\delta_{C\text{-ave}}$) calculated from the ChemBioDraw Ultra 12 program[20] (see Table 12).



XXI

$$\alpha = 1.73(\pm 0.33)\delta_{C\text{-ave}} - 214.35(\pm 44.15) \quad (29)$$

where n = 12, r^2 = 0.930, s = 0.711, q^2 = 0.888, and $F_{1,10}$ = 132.857

The number of valence electrons (NVE) is an easily calculable and a different form of polarizability parameter, which was developed by Hansch et al.[67] It can be represented as NVE = $n_\sigma + n_\pi + n_n$, where n_σ is the number of electrons in the σ-orbital, n_π is the number of electrons in π-orbitals, and n_n is the number of lone pair electrons. This parameter is simply computed by adding up the number of valence electrons in a molecule, e.g. H = 1, C = 4, Si = 4, N = 5, P = 5, O = 6, S = 6, and halogens = 7. An excellent correlation was also established between NVE and molecular polarizability, α (n = 146, r^2 = 0.987).[68] NVE was further applied successfully as a polarizability parameter in an extensive QSAR study pertaining to various chemical−biological interactions.[68−70] Using the data in Table 12, QSPR (eq 30) was developed between NVE and δ_C(C-ave).

$$NVE = 5.73(\pm 1.18)\delta_{C\text{-ave}} - 707.01(\pm 155.96) \quad (30)$$

where n = 12, r^2 = 0.921, s = 2.513, q^2 = 0.875, and $F_{1,10}$ = 116.582

The polarizability (α or NVE) of X-chlorobenzenes (**XXI**) increases with increasing the number of chlorines. A positive coefficient of $\delta_{C\text{-ave}}$ in both eqs 29 and 30 further suggests that the polarizability (α or NVE) of X-chlorobenzenes (**XXI**) increases with increasing their $\delta_{C\text{-ave}}$ values. Thus, the polarizability of this set of compounds can be measured by their $\delta_{C\text{-ave}}$ values.

3.2. Hydrophobic Descriptor

The hydrophobic parameter is one of the most important and frequently used molecular descriptors in QSAR studies due to its usefulness in the prediction of the pharmacokinetics of the drugs, and thus, it greatly influences the fate of new drug molecules. The hydrophobicity of a molecule is traditionally expressed as the

dx.doi.org/10.1021/cr100125d |Chem. Rev. 2011, 111, 2865–2899

logarithm of the partition coefficient (log P) for partitioning of the molecule between an organic and an aqueous phase, where the organic phase imitates cell membrane phospholipids in the interactions.[71] Since the pioneering work of Hansch et al.,[1,72,73] the logarithm of the partition coefficient between n-octanol and water (log P_{oct} or log P) has been widely used as a hydrophobic descriptor in QSAR analysis. It also has a critical role in each of the components of ADMET. Of course, it cannot serve as a "magic bullet" but must play an important role in the drug design and development.

While log P can be determined experimentally, either by direct methods (shake-flask, slow-stirring, filter-probe extractor, potentiometric titration methods, etc.)[74−76] or indirect methods (reversed phase thin-layer chromatography, reversed phase high-performance liquid chromatography, artificial membrane chromatography, electrokinetic chromatography, etc.),[77−79] this requires the prior synthesis of the compound. Therefore, numerous computational programs (e.g., CLOGP, KLOGP, KOWWIN, ACD/log P, SMILOGP, XLOGP, etc.) along with various QSPR models based on the calculated descriptors have been developed.[80,81]

NMR chemical shifts are quantum mechanical energies that are mainly composed of two terms, such as diamagnetic and paramagnetic, for which the diamagnetic term is directly related to the electrostatic potential at the nucleus, whereas the paramagnetic term is mainly related to the electron orbitals surrounding the nucleus and independent of the solvent−solute interactions.[9] This means that the changes occurring in chemical shifts due to solvent can be directly related to the electrostatic potential at the nucleus and may represent the solvent−solute enthalpy changes. Therefore, the difference in chemical shift values observed in two solvent systems (an aqueous and an organic solvent) may correlate to log P. By taking this assumption, a rapid and easily calculable MLR model (eq 31) was developed by Schnackenberg and Beger[82] using a diverse set of 162 compounds for the computation of log P.

$$\log P = 0.0092(\pm 0.0038)\Delta C + 0.0309(\pm 0.0013)\mathrm{MV} + 0.249(\pm 0.064)\mathrm{NHD} - 0.694(\pm 0.042)\mathrm{NHA} - 0.806(\pm 0.164) \quad (31)$$

where n = 162, r^2 = 0.880, s = 0.494, and $F_{4,157}$ = 281.87

In eq 31, ΔC is the sum of ^{13}C NMR chemical shift differences in water and methanol, MV is the molar volume, NHD is the number of hydrogen bond donors, and NHA is the number of hydrogen bond acceptors of the compound. While log P values are determined for the partitioning of a compound between water and n-octanol, methanol (MeOH-d_4) was chosen as the NMR solvent that may provide a substitute for n-octanol, as it has the most closely related polarizability and miscibility of 1-octanol.

Although computational programs are preferred for the log P calculation, they may provide misleading information for new and complex chemical/drug molecules with unknown (missing) fragments and not verified in the absence of the experimental values (either of the molecule in question or their analogues). In that situation, we believe that eq 31 may provide alertness or may prove to be an alternative tool to calculate log P with an acceptable range.

3.3. Steric Descriptors

Taft's E_S parameter[23,34] was the first successful attempt to quantify the steric features of various functional groups in QSAR/QSPR analysis, which is defined as follows:

$$E_S = \log(k_X/k_H)_{acid} \quad (32)$$

where k_X and k_H are rate constants for the acid hydrolysis of esters, XCH_2COOR and CH_3COOR, respectively. The other common steric parameters used in the development of QSAR/QSPR models are molar refractivity (MR), molar volume (MgVol), molecular weight (MW), and Verloop's sterimol parameters, such as modified length (L), minimum width ($B1$), and maximum width ($B5$).

Molar refractivity (MR) is one of the most widely used steric parameters in Hansch QSAR/QSPR analysis, which is generally considered to be a crude measure of overall bulk. It includes a polarizability component which may describe cohesion and is related to the London dispersion forces. It is represented by the following eq:[23]

$$\mathrm{MR} = \frac{4}{3}\pi N\alpha \quad (33)$$

where MR is the molar refractivity and α the polarizability of the molecule. N is Avogadro's number, and π = 3.14.

MR is also defined by the well-documented Lorentz−Lorenz equation (eq 34):[83,84]

$$\mathrm{MR} = \frac{(n^2-1)}{(n^2+2)} \times \frac{MW}{\rho} \quad (34)$$

where n is the refractive index, MW the molecular weight, and ρ the density of the compound. Thus, MR is dependent on both the volume and polarizability. MR is generally scaled by 0.1 and then utilized in the biological QSAR, where intermolecular interactions are of the primary concern. It can be used for a substituent or for the whole molecule. Using the data in Table 12, QSPR 35 was developed between CMR (calculated molar refractivity) and the calculated averaged ^{13}C NMR chemical shifts ($\delta_{C\text{-ave}}$) over all the ring carbons of X-chlorobenzene (**XXI**).

$$\mathrm{CMR} = 0.47(\pm 0.10)\delta_{C\text{-ave}} - 57.67(\pm 12.77) \quad (35)$$

where n = 12, r^2 = 0.921, s = 0.206, q^2 = 0.875, and $F_{1,10}$ = 116.582
CMR vs α; r = 0.999
CMR vs NVE; r = 1.000
CMR vs MgVol; r = 1.000

The high correlations of CMR with polarizability (α and NVE) and molar volume (MgVol) support the idea that the MR parameter is dependent on both the polarizability and volume. The above linear correlation (eq 35) suggests that the CMR of X-chlorobenzene (**XXI**) increases with increasing $\delta_{C\text{-ave}}$. This means that $\delta_{C\text{-ave}}$ represents both the electronic (polarizability) and volume terms. The polarizability term may associate with the electrostatic potential at the carbon nucleus, whereas the volume term may be due to the electron orbital (cloud) surrounding the nucleus.

3.4. Topological Descriptors

The Wiener index (W) is one of the oldest and most widely used topological indices, but it is applicable only for the acyclic graphs and not for the cyclic.[85] On the other hand, the Szeged (Sz) index is the modification of the Wiener index and applicable for both the graphs: acyclic and cyclic.[86,87] It is interesting to note that the Wiener and Szeged indices coincide with each other for

the acyclic graphs. For this reason, another index, called the Padmakar−Ivan (PI) index, was proposed.[88,89] Unlike the Sz index, the PI index is different for acyclic and cyclic graphs. Randic connectivity ($^0\chi$, $^1\chi$, $^2\chi$) indices are the integer numbers by applying structural operations involving more than one vertex at a time.[90,91] Kier and Hall valence connectivity ($^0\chi^v$, $^1\chi^v$, $^2\chi^v$) indices are the extended form, for which valence deltas δ^v are used in computing subgraph terms.[92,93]

Experimental ^{13}C NMR chemical shifts (ppm, TMS = 0) on carbinol carbon atoms of a reasonably large set of alcohols (n = 32) were used by Jaiswal and Khadikar[94] to correlate it with a large set of distance based topological indices (e.g., W, PI, $^m\chi$, and $^m\chi^v$), resulting in the development of 19 correlations with correlation coefficients (r) ranging from 0.774 to 0.889. The best result with fewer descriptors and high statistics came in the form of the following QSPR model (eq 36):

$$\delta_{C-OH} = 31.944(\pm 8.876)^0\chi - 31.871(\pm 9.360)^1\chi - 15.225(\pm 5.615)^2\chi + 27.951(\pm 7.524) \quad (36)$$

where n = 32, r = 0.852, s = 4.315, and $F_{3,28}$ = 24.627

δ_{C-OH} is the experimental ^{13}C NMR chemical shifts of carbinol carbon atoms in alcohols. The positive coefficient of $^0\chi$ suggests that the nature of atom(s) has a significant effect on the ^{13}C NMR chemical shift, while the negative coefficients of $^1\chi$ (provides information about the nature of atoms and first-order branching) and $^2\chi$ (encodes more information about branching) descriptors indicate the detrimental effects of branching on the exhibition of the ^{13}C NMR chemical shift of the present data set of alcohols. The corresponding model containing $^0\chi^v$, $^1\chi^v$, and $^2\chi^v$ gave slightly lower statistics (r = 0.844, s = 4.419, and F = 23.038). The valence connectivity indices encode electronic information in addition to the simple connectivity indices.[95] Thus, δ_{C-OH} can measure both the electronic and steric effects for this data set. Although tetraparametric (containing PI, $^0\chi$, $^1\chi$, and $^2\chi$) and pentaparametric (containing PI, W, $^0\chi$, $^1\chi$, and $^2\chi$) models showed slightly better statistics, those models containing one or more parameters have coefficients that are considerably smaller than their corresponding standard deviations.

4. QSAR RESULTS AND DISCUSSION

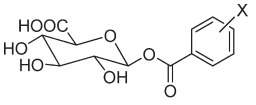
4.1. Physical (or Nonbiological) QSAR

4.1.1. Degradation. Glucuronidation of carboxylic acid drugs, such as nonsteroidal anti-inflammatory drugs (NSAIDs), has been found to be in a metabolic activation pathway. This may possibly be due to the covalent binding of resultant 1-β-O-acyl glucuronides (βGAs) to proteins.[96−98] βGAs are potentially reactive electrophiles and thought to be implicated in the adverse drug reactions (ADRs) of the parent carboxylic acid drugs.[99] Thus, the development of QSAR models for this group of compounds could be of great interest for designing new drugs. Vanderhoeven and co-workers were the first who explored correlations between the degradation rate constants (log k_d) of nine 1-β-O-acyl glucuronides of p-substituted-benzoic acids with $\delta_C(C{=}O)_{acid}$ (r^2 = 0.783) and $\delta_C(C{=}O)_{glucuronide}$ (r^2 = 0.786).[100] In their correlations, $\delta_C(C{=}O)_{acid}$ corresponds to the experimental ^{13}C NMR chemical shifts of the carbonyl carbons of benzoic acids, whereas $\delta_C(C{=}O)_{glucuronide}$ represents the experimental ^{13}C NMR chemical shifts of the carbonyl carbons of the glucuronides of the corresponding benzoic acids. The identical r^2 values in both of their correlations suggest that either $\delta_C(C{=}O)_{acid}$ or $\delta_C(C{=}O)_{glucuronide}$ might be used as descriptor for the prediction of log k_d values of 1-β-O-benzoyl glucuronides with p-substituents (BAGAs). Of course, $\delta_C(C{=}O)_{acid}$ is the better descriptor, since it is more readily available as compared to $\delta_C(C{=}O)_{glucuronide}$ of the corresponding synthesized 1-β-O-benzoyl glucuronides (BAGAs). This study was further extended by the development of QSAR eqs 37 and 38 for the prediction of log k_d values of 1-β-O-benzoyl glucuronides with o-, m-, and p-substituents, **XXII** (BAGAs).[101]

Table 13. Degradation Rate Constants (log k_d) of BAGAs (XXII) and the Measured ^{13}C NMR Chemical Shift (ppm rel. to TMS) of the Carbonyl Carbon in Their 1-β-O-Acyl Group Used To Derive eq 38

		log k_d (eq 38)			
no.	X	obsd[a]	pred[b]	Δ	$[\delta_C(C{=}O)_{glucuronide}]^a$
1	H	−0.46	−0.65	0.19	166.54
2	2-Cl	0.08	−0.08	0.16	165.19
3	3-Cl	0.10	−0.11	0.21	165.25
4	4-Cl	−0.21	−0.26	0.05	165.62
5	2-CH_3	−0.91	−0.93	0.02	167.19
6	3-CH_3	−0.60	−0.71	0.11	166.68
7	4-CH_3	−0.78	−0.68	−0.10	166.61
8	2,4-$(CH_3)_2$	−1.19	−0.89	−0.30	167.10
9	2-OCH_3	−0.84	−0.44	−0.40	166.05
10	4-OCH_3	−1.09	−0.57	−0.52	166.34
11	2-CF_3	−0.63	−0.39	−0.24	165.93
12	4-CF_3	0.25	−0.12	0.37	165.28
13	4-F	−0.42	−0.22	−0.20	165.51
14	2-Ph	−1.44	−1.31	−0.13	168.09
15	3-Ph	−0.35	−0.64	0.29	166.51
16	4-Ph	−0.46	−0.54	0.08	166.29
17	2-NHPh	−1.09	−1.33	0.24	168.14
18	2-NHPh(2′,3′-$(CH_3)_2$)	−1.36	−1.53	0.17	168.62

[a] $\delta_C(C{=}O)_{glucuronide}$: ^{13}C NMR chemical shift of the carbonyl carbon in the 1-β-O-acyl group of BAGAs measured in MeOH-d_4. Data from ref 101. [b] Calculated from the C-QSAR program (ref [17]).



XXII

$$\log k_d = -0.329(\pm 0.053)\delta_C(C{=}O)_{acid} + 54.55(\pm 8.88) \quad (37)$$

where n = 18, r^2 = 0.707, s = 0.279, and $F_{1,16}$ = 38.6

$$\log k_d = -0.42(\pm 0.13)\delta_C(C{=}O)_{glucuronide} + 69.75(\pm 21.87) \quad (38)$$

where n = 18, r^2 = 0.744, s = 0.261, q^2 = 0.682, and $F_{1,16}$ = 46.5 (Note: QSAR eq 38 is from our present work. The detail data is given in Table 13.)

Again the results from QSARs 37 and 38 support the idea that either $\delta_C(C{=}O)_{acid}$ or $\delta_C(C{=}O)_{glucuronide}$ might be used as

Table 14. Statistical Parameters for Multiple Regression Analyses Using Electronic and Steric Descriptors for log k_d Values of 18 BAGAs (XXII)[a]

QSAR no.	electronic descriptor	steric descriptor	const	n	r^2	q^2	s	F	ref
i	δ_{COOH} [0.84 (±0.15)]	$\delta_C(C{=}O)_{acid}$ [−0.269 (±0.034)]	33.46	18	0.901	0.840	0.167	68.30	101
ii	H_{PAC} [137 (±22)]	$\delta_C(C{=}O)_{acid}$ [−0.201 (±0.035)]	−39.24	18	0.921	0.890	0.150	87.50	101
iii	pK_a [−1.12 (±0.12)]	Es [0.500 (±0.047)]	4.17	16	0.907	0.776	0.157	63.50	101
iv	H_{PAC} [180 (±19)]	Es [0.128 (±0.040)]	−95.07	16	0.909	0.863	0.155	65.00	101
v	pK_a [−0.461 (±0.272)]	$\delta_C(C{=}O)_{acid}$ [−0.368 (±0.088)]	62.81	18	0.843	0.773	0.211	40.27	PW
vi	δ_{COOH} [1.74 (±0.421)]	Es [0.066 (±0.174)]	−23.10	15	0.872	0.822	0.191	40.88	PW
vii	σ [0.776 (±0.447)]	$\delta_C(C{=}O)_{acid}$ [−0.311 (±0.125)]	51.35	16	0.855	0.743	0.196	38.33	PW

[a] $\delta_C(C{=}O)_{acid}$ and δ_{COOH}: ^{13}C NMR chemical shift of the carbonyl carbon and ^1H NMR chemical shift of the carboxylic hydrogen in the parent BAs measured in DMSO-d_6, respectively. H_{PAC}: the calculated partial atomic charge. PW: present work. Adapted with permission from ref 101. Copyright 2009 American Chemical Society.

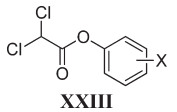
Table 15. Rate Coefficient (log k_o) for the Neutral Hydrolysis of X Phenyl Dichloroacetates (XXIII) in 20% Acetonitrile−Water at 298.2 K, Measured ^{13}C NMR Chemical Shift (ppm rel. to TMS) of the Carbonyl Carbon, and σ Constants of X Substituents Used To Derive eqs 39 and 40

			log k_o (eq 39)[b]		log k_o (eq 40)[b]			
no.	X	log k_o (obsd)[a]	pred	Δ	pred	Δ	σ^b	$\delta_C(C{=}O)^a$
1	4-NO_2	1.16	1.10	0.06	1.06	0.10	0.78	162.13
2	3-NO_2	0.93	0.67	0.26	0.95	−0.02	0.71	162.43
3	4-CN	0.94	1.00	−0.06	0.87	0.07	0.66	162.20
4	4-Cl	0.13	0.16	−0.03	0.16	−0.03	0.23	162.78
5	3-Cl	0.27	0.35	−0.07	0.39	−0.12	0.37	162.65
6	4-Br	0.10	0.27	−0.17	0.16	−0.06	0.23	162.70
7	H	−0.32	−0.19	−0.13	−0.22	−0.10	0.00	163.02
8	4-CH_3	−0.56	−0.46	−0.10	−0.50	−0.06	−0.17	163.20
9	4-OCH_3	−0.46	−0.70	0.24	−0.67	0.21	−0.27	163.37

[a] Data from ref 102. [b] Calculated/Data from the C-QSAR program (ref 17).

predictor for log k_d values of BAGAs. The log k_d values of BAGAs possessing a bulkier X substituent such as Me and Ph, are seen in the following order: *m*-isomer > *p*-isomer > *o*-isomer. The steric effect of these bulkier groups may overcome the electronic effect, resulting in a lower log k_d value for the *o*-substituted BAGAs. On the other hand, log k_d values of BAGAs possessing a Cl substituent are in the following order: *m*-isomer > *o*-isomer > *p*-isomer. This may possibly be due to the intermediate bulkiness of the Cl substituent ($E_s = -0.97$). Furthermore, 2,4-Me_2-BAGA has a lower log k_d value than that of both 2-Me-BAGA and 4-Me-BAGA. The electron-releasing *p*-Me group resulted in increasing pK_a and decreasing log k_d values of 2,4-Me_2-BAGA as compared to that of 2-Me-BAGA. On the other hand, the reactivity difference between 2,4-Me_2-BAGA and 4-Me-BAGA is mainly due to the sterically hindered *o*-Me group. The steric effect may be surpassed by the electronic effect and hence decrease the log k_d value of 2,4-Me_2-BAGA as compared to that of 4-Me-BAGA (see Table 13).[101] In order to verify the practical effectiveness of the electronic and steric descriptors for the log k_d values of BAGAs (**XXII**), various QSAR models (eqs i−vii) were developed by using the combinations of electronic (δ_{COOH}, p*K*a, σ, or H_{PAC}) and steric ($\delta_C(C{=}O)_{acid}$, or E_s) descriptors (Table 14). Equation ii (Table 14) was found to be among the best, as judged by their r^2 and q^2 values.

4.1.2. Hydrolysis. The correlation between the experimental ^{13}C NMR chemical shift of carbonyl carbons (in $CDCl_3$) and the logarithms of their rate coefficients of the neutral hydrolysis (k_o) for X-phenyl dichloroacetates (**XXIII**, $n = 9$, $r^2 = 0.934$) was published by Neuvonen and Neuvonen.[102] We used their data (Table 15) and developed eqs 39 and 40:



XXIII

$$\log k_o = -1.46(\pm 0.33)\delta_C(C{=}O) + 237.35(\pm 53.55) \quad (39)$$

where $n = 9$, $r^2 = 0.940$, $s = 0.168$, $q^2 = 0.887$, and $F_{1,7} = 109.667$

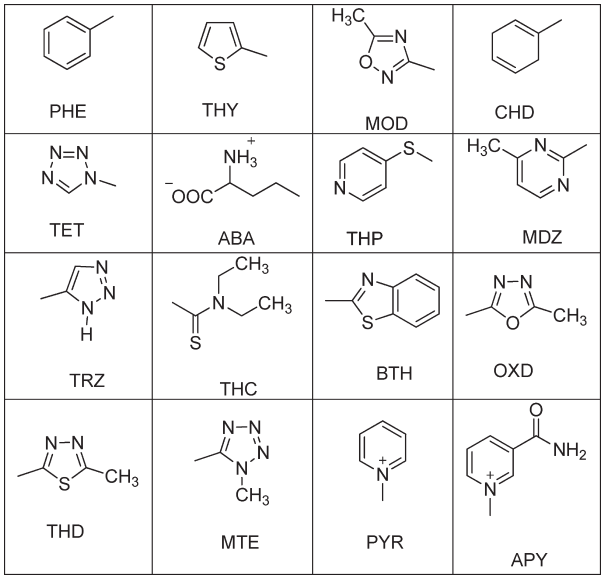
$$\log k_o = 1.65(\pm 0.25)\sigma - 0.22(\pm 0.11) \quad (40)$$

where $n = 9$, $r^2 = 0.973$, $s = 0.113$, $q^2 = 0.944$, and $F_{1,7} = 252.259$

In the above equations, k_o is the rate coefficient for the neutral hydrolysis of X-phenyl dichloroacetates (**XXIII**) in 20% acetonitrile−water at 298.2 K. The rate coefficient (k_o) of the esters (**XXIII**) increases with electron-withdrawing X substituents, as suggested by the positive ρ value of the Hammett σ in eq 40 and the data in Table 15. It is important to note that the electron-withdrawal by the X substituents results in an upfield shift of the ^{13}C NMR resonance of the carbonyl carbon, which also reflects the increased double bond character of the C=O bond. This suggests that the k_o of **XXIII** increases with an upfield shift of the ^{13}C NMR of the C=O carbon, which is supported by the negative coefficient of $\delta_C(C{=}O)$ in eq 39. There is also an excellent mutual correlation between $\delta_C(C{=}O)$ and the

dx.doi.org/10.1021/cr100125d |*Chem. Rev.* 2011, 111, 2865–2899

Table 16. Difference between the Experimental ^{13}C NMR Chemical Shifts of the C-4 and C-3 Carbons $\Delta\delta_{(4\text{-}3)}$ and The Rate Constants of Base Hydrolysis (log k) of Cephalosporin Analogues (XXV) Used To Derive eq 44



			log k (eq 44)			
no.	X	Y	obsd[a]	pred[b]	Δ	$[\Delta\delta_{(4-3)}]^a$
1	H	H	ND[d]	2.50	ND[d]	4.70
2[c]	H	$COOCH_3$	2.57	3.25	−0.68	15.60
3	$COCH_2$-PHE	H	2.23	2.47	−0.24	4.20
4	$COCH_2$-THY	H	ND[d]	2.48	ND[d]	4.30
5	$COCH_2$-MOD	H	2.32	2.46	−0.14	4.10
6	$COCH_2O$-PHE	H	ND[d]	2.47	ND[d]	4.20
7	$COCH(NH_2)$-PHE	H	2.76	2.50	0.26	4.70
8	$COCH(NH_2)$-CHD	H	2.75	2.49	0.26	4.50
9	$COCH_2$-THY	OH	ND[d]	2.74	ND[d]	8.20
10	$COCH_2$-THY	SCH_3	ND[d]	2.83	ND[d]	9.50
11[c]	$COCH_2$-THY	S-MDZ	2.28	2.85	−0.57	9.70
12[c]	$COCH_2$-MOD	S-TRZ	2.41	2.85	−0.43	9.70
13	$COCH_2$-THY	S-TRZ	2.45	2.81	−0.36	9.20
14	$COCH_2$-THY	S-THC	2.83	2.83	−0.01	9.50
15	$COCH_2$-THY	S-BTH	2.98	3.09	−0.12	13.30
16	$COCH_2$-MOD	S-OXD	3.02	3.03	−0.01	12.40
17	$COCH_2$-MOD	S-THD	3.02	3.02	−0.01	12.30
18	$COCH_2$-TET	S-THD	3.03	3.02	0.01	12.20
19	$COCH_2$-THY	S-THD	3.08	3.06	0.02	12.80
20	$COCH_2$-THY	S-MTE	3.10	3.05	0.06	12.60
21	COCH(COONa)-THY	S-MTE	3.09	3.04	0.05	12.50
22	$COCH(CH_3)$-THY	S-MTE	3.05	3.08	−0.03	13.10
23	COCH(Br)-THY	S-MTE	3.32	3.04	0.28	12.50
24	$COCH_2$-ABA	$OCOCH_3$	3.08	3.04	0.04	12.50
25	$COCH_2$-THY	$OCOCH_3$	3.17	3.22	−0.05	15.10
26	$COCH_2$-THP	$OCOCH_3$	3.20	3.22	−0.02	15.10
27	$COCH_2$−PHE	$OCOCH_3$	ND[d]	3.23	ND[d]	15.30
28	$COCH_2O$-PHE	$OCOCH_3$	ND[d]	3.21	ND[d]	15.00
29	$CONHCH_2$-THY	$OCOCH_3$	ND[d]	3.21	ND[d]	15.00
30	$COCH_2$-THY	PYR	3.72	3.75	−0.02	22.80
31	$COCH_2$-THY	APY	3.76	3.73	0.02	22.60

[a] Data from ref 104. [b] Calculated from the C-QSAR program (ref 17). [c] Not included in the derivation of QSAR eq 44. [d] Not determined.

dx.doi.org/10.1021/cr100125d |*Chem. Rev.* 2011, 111, 2865–2899

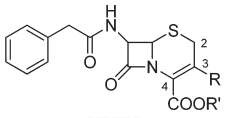
Table 17. Alkaline Hydrolysis Rate Constants ($\log k_{obsd}^{NMR}$) of Oxacephems Analogues (XXVI) Obtained by ^{1}H NMR Spectroscopy at pD 10.4 and 35 °C, and σ_I Values of X Substituents Used To Derive eq 46

No.	X	$\log k_{obsd}^{NMR}$ (eq 46)			$(\sigma_I)^{a/b}$
		obsd.[a]	pred.[b]	Δ	
1	H	-1.32	-1.16	-0.16	0.00
2	CN	-0.54	-0.51	-0.03	0.56
3	$N^+C_5H_5$	0.10	0.12	-0.02	1.09
4	OH	-0.79	-0.87	0.08	0.25
5	OCH_3	-0.69	-0.85	0.16	0.27
6	$OCOCH_3$	-0.51	-0.71	0.20	0.39
7	$OCONH_2$	-0.64	-0.62	-0.02	0.46
8	SCH_3	-0.96	-0.89	-0.07	0.23
9		-0.69	-0.54	-0.15	0.53

[a] Data from ref 105. [b] Calculated/Data from the C-QSAR program (ref 17).

Hammett σ constant of the X substituents (r = 0.981). On considering all the facts together for eqs 39 and 40, one can say that eq 39 may provide an alternative to evaluate rate coefficients, especially for the substituted derivatives for which the substituent constants (σ values) are unknown.[102]

Relationships between the chemical reactivity of 3-methylene-substituted cephalosporins (**XXIV**) and their experimental ^{13}C NMR chemical shifts were developed by Nishikawa and Tori.[103] According to their results, the values of the ^{13}C NMR chemical shifts ($\delta_{C\text{-}3}$ and δ_{COO}) of 3-methylene-substituted cephalosporins were well correlated with the logarithms of their rate constants for alkaline hydrolysis ($\log k_{obsd}$). Correlations were shown by eqs 41 and 42:



XXIV

R = H, Cl, CH_3, $CH_2COOC_2H_5$, CH_2OCOCH_3,

$$\log k_{obsd} = -0.161(\pm 0.099)\delta_{C\text{-}3} + 19.800(\pm 11.659) \quad (41)$$

where n = 7, r^2 = 0.778, s = 0.285, q^2 = 0.632, and $F_{1,5}$ = 17.523

$$\log k_{obsd} = -0.507(\pm 0.283)\delta_{COO} + 86.445(\pm 47.809) \quad (42)$$

where n = 7, r^2 = 0.809, s = 0.264, q^2 = 0.553, and $F_{1,5}$ = 21.178 (Note: Statistical data and the 95% confidence intervals within parentheses were added to the original eqs 41 and 42.)

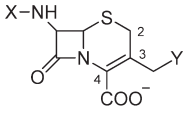
The negative coefficients associated with $\delta_{C\text{-}3}$ and δ_{COO} suggest that the rate of alkaline hydrolysis of cephalosporins (**XXIV**) increases with an upfield shift of the ^{13}C NMR chemical shift of the C-3 and COO (attached to C-4 carbon) carbons. On the other hand, the log k_{obsd} values of cephalosporins (**XXIV**) were well correlated with the ^{13}C NMR chemical shift difference between their C-4 and C-3 carbons ($\Delta\delta_{(4-3)}$), as shown in the following eq 43:

$$\log k_{obsd} = 0.085(\pm 0.071)\Delta\delta_{(4\text{-}3)} - 0.340(\pm 1.060) \quad (43)$$

where n = 5, r^2 = 0.827, s = 0.302, q^2 = 0.656, and $F_{1,3}$ = 14.341 (Note: Statistical data and the 95% confidence intervals within parentheses were added to the original eq 43; cephalosporins (**XXIV**) with R = CH_3, $CH_2COOC_2H_5$, CH_2OCOCH_3, CH_2(1-methyl-1*H*-tetrazol-5-yl)thio, $CH_2N^+C_5H_5$, and R′ = Na were used in the development of eq 43 by the authors[103].)

The above equation suggests that the polarization of the C_3═C_4 double bond may be an important factor for determining the β-lactam reactivity of 3-methylene-substituted cephalosporins. Although the predictive ability of eq 43 has been demonstrated by its high correlation coefficient (r^2 = 0.827), the data set is very small (n = 5).

In the parallel work, a relationship between the logarithm of the alkaline hydrolysis rate constant (k) and the experimental ^{13}C NMR chemical shift difference ($\Delta\delta_{(4-3)} = \delta_{C\text{-}4} - \delta_{C\text{-}3}$) for a reasonably large data set of cephalosporins (**XXV**) was graphically demonstrated by Coene et al.[104] The published data of rate constant (k) was converted into their logarithm values ($\log k$), and QSAR eq 44 was developed (see Table 16 for the data and for the structures of the substituents).



XXV

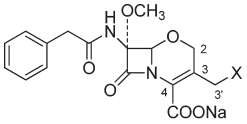
$$\log k = 0.07(\pm 0.02)\Delta\delta_{(4\text{-}3)} + 2.18(\pm 0.19) \quad (44)$$

where n = 20, r^2 = 0.840, s = 0.158, q^2 = 0.797, and $F_{1,18}$ = 94.500

outliers: X = H, Y = $COOCH_3$; X = $COCH_2$-THY, Y = S-MDZ; X = $COCH_2$-MOD, Y = S-TRZ

According to eq 44, it can be suggested that the rate of the alkaline hydrolysis of cephalosporins (**XXV**) depends directly on the C_3═C_4 double bond polarity, as reflected by $\Delta\delta_{(4-3)}$. This supports the prediction coming from the QSAR 43. Three compounds (X = H, Y = $COOCH_3$; X = $COCH_2$-THY, Y = S-MDZ; X = $COCH_2$-MOD, Y = S-TRZ) were deemed to be outliers because they predicted higher log k values than observed, by 4.3, 3.6, and 2.7 times, respectively, compared to the standard deviation. The exact reason for this deviation is not very clear, but the first outlier may suggest that the substitution on X is preferred.

This work was further extended by Narisada and coworkers[105] in order to clarify the effects of 3′-substituents upon the chemical reactivity of the β-lactam ring of oxacephems (**XXVI**) and developed the linear correlation in eq 45.



XXVI

$$\log k_{obsd}^{NMR} = 0.0655(\pm 0.022)\Delta\delta_{(4\text{-}3)} - 1.04(\pm 0.167) \quad (45)$$

where n = 9, r^2 = 0.876, s = 0.142, q^2 = 0.742, and $F_{1,7}$ = 49.452 (Note: Statistical data and the 95% confidence intervals within parentheses were added to the original eq 45.)

dx.doi.org/10.1021/cr100125d |*Chem. Rev.* 2011, 111, 2865–2899

Table 18. Calculated ^{13}C NMR Chemical Shift Difference ($\Delta\delta_{C1}$), Rate Constants of Oxidation (log k) for 4-X-Benzaldehydes (XXVII), and σ^+ Values of X Substituents Used To Derive eqs 47 and 48

			log k (eq 47)[b]		log k (eq 48)[b]			
no.	X	log k (obsd)[a]	pred	Δ	pred	Δ	$(\Delta\delta_{C1})^c$	$(\sigma^+)^b$
1	NO_2	1.41	1.33	0.08	1.40	0.01	6.10	0.79
2	CN	1.51	1.60	−0.09	1.56	−0.05	4.30	0.66
3	CF_3	1.67	1.75	−0.08	1.62	0.05	3.30	0.61
4	$COOCH_3$	1.76	1.60	0.16	1.78	−0.02	4.30	0.49
5	Br	2.23	2.40	−0.17	2.20	0.03	−1.00	0.15
6	H	2.28	2.25	0.03	2.39	−0.11	0.00	0.00
7	Cl	2.31	2.53	−0.22	2.26	0.05	−1.90	0.11
8	F	2.58	2.91	−0.33	2.48	0.10	−4.40	−0.07
9	CH_3	2.71	2.70	0.01	2.79	−0.08	−3.00	−0.31
10	SCH_3	2.99	2.79	0.20	3.15	−0.16	−3.60	−0.60
11	$NHCOCH_3$	3.27	2.91	0.36	3.15	0.12	−4.40	−0.60
12	OCH_3	3.44	3.41	0.03	3.38	0.06	−7.70	−0.78

[a] Data from ref 106. [b] Calculated/Data from the C-QSAR program (ref 17). [c] Calculated from the ChemBioDraw Ultra 12 program (ref 20).

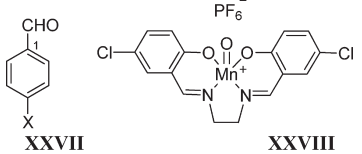
k_{obsd}^{NMR} is the alkaline hydrolysis rate constants for oxacephems (**XXVI**) obtained by ^{1}H NMR spectroscopy at pD 10.4 and 35.0 °C, whereas $\Delta\delta_{(4-3)}$ is the experimental ^{13}C NMR chemical shift differences between the C-4 and C-3 carbons of oxacephems (**XXVI**). There is an excellent mutual correlation between log k_{obsd}^{NMR} and σ_I of X substituents (r = 0.915). Thus, the effects of the 3′-substituents on the chemical reactivity of the β-lactam ring of oxacephems (**XXVI**) can be explained in terms of its inductive effect, as shown by eq 46 (Table 17).[105]

$$\log k_{obsd}^{NMR} = 1.17(\pm 0.36)\sigma_I - 1.16(\pm 0.18) \quad (46)$$

where $n = 9$, $r^2 = 0.896$, $s = 0.131$, $q^2 = 0.840$, and $F_{1,7} = 60.308$

The above equation suggests that the electron-withdrawing 3′-substituent increases the chemical reactivity of the β-lactam ring of oxacephems (**XXVI**), which correspond to the degree of polarization of the lactam moiety toward the enamine and expressed by $\Delta\delta_{(4-3)}$ values. On comparison between eqs 45 and 46, one can now suggest that the chemical reactivity of the β-lactam ring of the oxacephems (**XXVI**) may be estimated by $\Delta\delta_{(4-3)}$ values without the measurement of the pseudo-first-order rate constant (k_{obsd}^{NMR}) of the alkaline hydrolysis, especially for those derivatives with unknown substituent constants (σ_I values).

4.1.3. Oxidation. The kinetics for the oxidation of 4-X-benzaldehydes (**XXVII**) to 4-X-benzoic acids by oxo(salen) manganese(v) complex (**XXVIII**) at 298 K were reported by Bansal et al.[106] The published data of rate constants (k) for the oxidation of 4-X-benzaldehydes (**XXVII**) was converted into their logarithms (log k), and QSAR eqs 47 and 48 were developed (Tables 18).



XXVII **XXVIII**

$$\log k = -0.15(\pm 0.03)\Delta\delta_{C1} + 2.25(\pm 0.13) \quad (47)$$

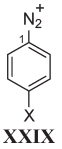
where $n = 12$, $r^2 = 0.919$, $s = 0.202$, $q^2 = 0.889$, and $F_{1,10} = 113.457$

$$\log k = -1.26(\pm 0.11)\sigma^+ + 2.39(\pm 0.06) \quad (48)$$

where $n = 12$, $r^2 = 0.984$, $s = 0.090$, $q^2 = 0.976$, and $F_{1,10} = 615$

Rate constants (k) for the oxidation of 4-X-benzaldehydes (**XXVII**) are in mol m^{-1} s^{-1}, whereas $\Delta\delta_{C1}$ is the difference between the calculated ^{13}C NMR chemical shifts of the C1 carbon of 4-X-benzaldehydes and that of the parent in ppm. The simple linear correlation (QSAR 47) suggests that the rate of the oxidation of 4-X-benzaldehydes may be determined by the $\Delta\delta_{C1}$ values without the measurement of the rate constant (k) for the oxidation and it increases with decreasing $\Delta\delta_{C1}$. There is an excellent correlation between $\Delta\delta_{C1}$ and σ^+ (r = 0.941), the Brown variant of the Hammett electronic parameter of X substituents. Thus, the effect of X substituents on the oxidation rate of 4-X-benzaldehydes can also be explained in terms of σ^+ (eq 48). The negative ρ^+ value of the σ^+ in eq 48 suggests that the rate of oxidation for 4-X-benzaldehydes increases by the electron-donating X substituents and that the same has been represented alternatively by the $\Delta\delta_{C1}$ term in eq 47 with a negative coefficient. Thus, an upfield shift in the ^{13}C NMR of the C1 carbon due to an electron-donating X substituent increases the oxidation rate of 4-X-benzaldehyde.

4.1.4. Reduction. Relative rate constants (k_{4X}/k_H) for the reduction of 4-X-benzenediazonium ions (**XXIX**) by citrato-copper(I) complex (Sandmeyer hydroxylation) were published by Hanson and co-workers.[107] The published data of the relative rate constants (k_{4X}/k_H) was converted into their logarithms (log k_{4X}/k_H) and used in the development of QSAR eqs 49 and 50 (Tables 19).



XXIX

$$\log k_{4X}/k_H = 0.09(\pm 0.03)\Delta\delta_{C1} + 0.22(\pm 0.17) \quad (49)$$

where $n = 11$, $r^2 = 0.857$, $s = 0.251$, $q^2 = 0.785$, and $F_{1,9} = 53.937$

$$\log k_{4X}/k_H = 1.19(\pm 0.18)\sigma^+ - 0.02(\pm 0.09) \quad (50)$$

where $n = 11$, $r^2 = 0.962$, $s = 0.129$, $q^2 = 0.943$, and $F_{1,9} = 227.842$

$\Delta\delta_{C1}$ is the difference between the calculated ^{13}C NMR chemical shifts of the C1 carbon of 4-X-benzenediazonium ions and that of the parent, whereas k_{4X} and k_H are the rate constants

Table 19. Calculated ^{13}C NMR Chemical Shift Difference ($\Delta\delta_{C1}$), Relative Rate Constants of Reduction (log k_{4X}/k_H) for 4-X-Benzenediazonium Ions (XXIX), and σ^+ Values of X Substituents Used To Derive eqs 49 and 50

			log k_{4X}/k_H (eq 49)[b]		log k_{4X}/k_H (eq 50)[b]			
no.	X	log k_{4X}/k_H (obsd)[a]	pred	Δ	pred	Δ	$(\Delta\delta_{C1})^c$	$(\sigma^+)^b$
1	OCH_3	−1.11	−0.88	−0.23	−0.95	−0.16	−11.60	−0.78
2	OC_6H_5	−0.60	−0.80	0.20	−0.62	0.02	−10.80	−0.50
3	CH_3	−0.47	−0.07	−0.40	−0.39	−0.08	−3.00	−0.31
4	H	0.00	0.22	−0.22	−0.02	0.02	0.00	0.00
5	F	0.14	−0.20	0.34	−0.11	0.25	−4.40	−0.07
6	Cl	0.30	0.04	0.26	0.11	0.19	−1.90	0.11
7	$COCH_3$	0.44	0.63	−0.19	0.57	−0.13	4.40	0.50
8	$COOC_2H_5$	0.52	0.62	−0.10	0.55	−0.03	4.30	0.48
9	CF_3	0.63	0.53	0.10	0.70	−0.07	3.30	0.61
10	CN	0.75	0.62	0.13	0.76	−0.01	4.30	0.66
11	NO_2	0.91	0.79	0.12	0.92	−0.01	6.10	0.79

[a] Data from ref 107. [b] Calculated/Data from the C-QSAR program (ref 17). [c] Calculated from the ChemBioDraw Ultra 12 program (ref 20).

Table 20. Calculated/Measured ^{13}C NMR Chemical Shifts Difference ($\Delta\delta_{(\beta-\alpha)}$), Experimental Redox Potential (E_p), and Biological (MIC; mol L^{-1}) Parameters for β-Nitro-styrene Analogues (XXX) Used To Derive eqs 51 and 52

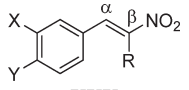
					log 1/MIC (eq 51)[b]		log 1/MIC (eq 52)[b]			
no.	X	Y	R	log 1/MIC (obsd)[a]	pred	Δ	pred	Δ	$(E_p)^a$	$\Delta\delta_{(\beta-\alpha)}$
1[f]	H	H	H	2.77	2.66	0.11	2.11	0.66	−0.46	−1.40[c]
2[f]	H	H	CH_3	3.41	3.57	−0.16	4.18	−0.77	−0.55	14.20[c]
3	OH	OH	H	2.85	3.07	−0.22	3.06	−0.21	−0.50	5.80[d]
4	OH	OH	CH_3	3.48	3.84	−0.36	3.63	−0.15	−0.57	10.10[d]
5	OCH_3	OCH_3	H	2.61	2.82	−0.21	2.84	−0.23	−0.47	4.10[a]
6	OCH_3	OCH_3	CH_3	3.84	3.68	0.16	3.86	−0.02	−0.56	11.80[a]
7	OH	OCH_3	H	2.88	2.88	0.00	2.86	0.02	−0.48	4.30[a]
8	OH	OCH_3	CH_3	4.12	3.74	0.38	3.82	0.30	−0.56	11.50[a]
9	OCH_3	OH	H	2.88	3.03	−0.15	3.01	−0.13	−0.49	5.40[a]
10	OCH_3	OH	CH_3	3.82	3.79	0.03	3.68	0.14	−0.57	10.50[a]
11	-OCH_2O-		H	3.18	2.84	0.34	2.76	0.42	−0.47	3.50[a]
12	-OCH_2O-		CH_3	3.81	3.75	0.06	3.96	−0.15	−0.56	12.60[e]

[a] Data from ref 111. [b] Calculated from the C-QSAR program (ref 17). [c] Calculated from the ChemBioDraw Ultra 12 program (ref 20). [d] Data from ref 112. [e] Data from ref 113. [f] Not included in the derivation of QSAR eq 52.

for the reduction of 4-X-benzenediazonium ions and unsubstituted benzenediazonium ion, respectively. Rhe relative rate constant (k_{4X}/k_H) for the reduction of 4-X-benzenediazonium ions increases with electron-withdrawing X groups, as suggested by the positive coefficient of $\Delta\delta_{C1}$ (eq 49) and σ^+ (eq 50). A high mutual correlation between $\Delta\delta_{C1}$ and σ^+ (r = 0.961) supports the idea that the relative rate constants for the reduction of 4-X-benzenediazonium ions (**XXIX**) can alternatively be estimated only by $\Delta\delta_{C1}$ values, especially for those derivatives with unknown substituent constants (σ^+ values).

4.2. Biological QSAR

4.2.1. Antibacterial Activity. The β-lactam antibiotics, such as cephalosporins, oxacephalosporins, penicillins, etc., inhibit biosynthesis of bacterial cell walls in bacteria by acylating and thereby inactivating the active center of the targeted transpeptidases, which may play an important role in the construction of the three-dimensional network of the cell walls as well as to induce the enzymatic self-lysis of the cell walls by interferences with the murein metabolism.[108−110] The *in vitro* antibacterial activity (log 1/MIC) of β-nitrostyrene derivatives (**XXX**) against Gram-positive bacteria (*Enterococcus faecalis* ATCC 29212) was found to be well correlated with their experimental redox potential (E_p), as demonstrated graphically by Milhazes and co-workers.[111] In turn, no correlation was found between antibacterial activity and lipophilicity. We used their data (Table 20) and developed the following QSAR 58.



XXX

$$\log 1/\mathrm{MIC} = -9.89(\pm 3.47)E_p - 1.84(\pm 1.81) \quad (51)$$

where n = 12, r^2 = 0.802, s = 0.238, q^2 = 0.710, and $F_{1,10}$ = 40.505

MIC represents the minimum inhibitory concentration in molar unit. The negative coefficient associated with E_p suggests that the antibacterial activity of β-nitrostyrenes (**XXX**) increases with decreasing their E_p values. It can now be suggested that the mode of action of β-nitrostyrenes (**XXX**) may closely be related to their electrophilicity, thus being a consequence of their ability

Table 21. Calculated ^{13}C NMR Chemical Shift (δ_{C1}) and Biological (log 1/IC_{50}; mol L^{-1}) and Physicochemical Parameters for Cinnamic Acid Analogues (XXXI) Used To Derive eqs 53 and 54[a]

			log 1/IC_{50} (eq 53)[c]		log 1/IC_{50} (eq 54)[c]				
no.	X	log 1/IC_{50} (obsd)[b]	pred	Δ	pred	Δ	(C log P)[c]	(σ^+)[c]	(δ_{C1})[d]
1	H	5.24	5.32	−0.08	5.21	0.03	6.80	0.00	135.20
2[e]	NO_2	5.19	4.47	0.72	4.47	0.72	6.54	0.79	141.30
3	CHO	4.40	4.39	0.01	4.20	0.20	6.15	0.47	141.00
4[e]	$COOCH_3$	6.00	4.88	1.12	4.78	1.22	6.77	0.49	139.50
5	CF_3	5.24	5.09	0.15	5.17	0.07	7.68	0.61	138.50
6	Cl	5.26	5.50	−0.24	5.65	−0.39	7.51	0.11	133.30
7	Br	5.49	5.48	0.01	5.58	−0.09	7.66	0.15	134.20
8	NH_2	5.26	5.21	0.05	5.11	0.15	5.57	−1.30	125.20
9	$CH(CN)_2$	4.37	ND	ND	4.76	−0.39	6.09	NA	134.50
10	CH_3	5.85	5.81	0.04	5.70	0.15	7.30	−0.31	132.20
11	OCH_3	5.89	5.94	−0.05	5.91	−0.02	6.72	−0.78	127.50
12	C_2H_5	5.92	5.87	0.05	5.76	0.16	7.83	−0.30	132.40
13	$CH(CH_3)_2$	5.92	5.78	0.14	5.71	0.21	8.22	−0.28	132.40
14	$C(CH_3)_3$	5.52	5.59	−0.07	5.61	−0.09	8.62	−0.26	132.10
15	OCH_2CH_3	6.07	6.22	−0.15	6.21	−0.14	7.25	−0.81	126.80
16	$O(CH_2)_2CH_3$	6.70	6.32	0.38	6.29	0.41	7.78	−0.83	126.80
17	$O(CH_2)_3CH_3$	5.96	6.21	−0.25	6.23	−0.27	8.30	−0.81	126.80

[a] ND = not determined, NA = not available. [b] Data from ref 119. [c] Calculated/Data from the C-QSAR program (ref 17). [d] Calculated from the ChemBioDraw Ultra 12 program (ref 20). [e] Not included in the derivation of QSAR eqs 53 and 54.

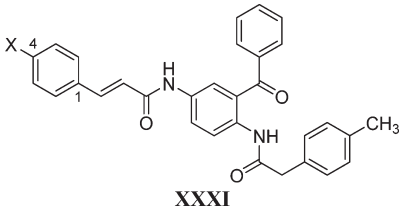
to accept electrons, which enables the reduction of the nitro group.[111] It has been noted that the antibacterial activities (Table 20) of the β-methyl-β-nitrostyrene derivatives are higher than those of their corresponding β-nitrostyrenes. This may possibly be due to the lower E_p values of the β-methyl-β-nitrostyrene derivatives as compared to their corresponding β-nitrostyrenes. Hence, the reduction potential seems to be highly sensitive to a methyl substitution in the β-position. Since there is a high mutual correlation (r = 0.920) between redox potentials (E_p)[111] of β-nitrostyrenes (**XXX**) and their $\Delta\delta_{(\beta-\alpha)}$ (calculated/experimental ^{13}C NMR chemical shift differences between C-β and C-α carbons),[20,111−113] we developed QSAR eq 52 from the same activity data in Table 20.

$$\log 1/\mathrm{MIC} = 0.13(\pm 0.05)\Delta\delta_{(\beta-\alpha)} + 2.29(\pm 0.43) \quad (52)$$

where n = 10, r^2 = 0.826, s = 0.235, q^2 = 0.709, and $F_{1,8}$ = 37.977
outliers: X = Y = R = H and X = Y = H, R = CH_3

The above equation suggests that the antibacterial activity of β-nitrostyrenes (**XXX**) increases with increasing $\Delta\delta_{(\beta-\alpha)}$ values; that is, it depends directly on the $C_\alpha{=}C_\beta$ double bond polarity. On comparison between QSARs 51 and 52, one can suggest that the antibacterial activity of β-nitrostyrenes (**XXX**) against Gram-positive bacteria (*Enterococcus faecalis* ATCC 29212) may be estimated by $\Delta\delta_{(\beta-\alpha)}$ values without the measurement of the redox potential (E_p) for this series of compounds. Two compounds (X = Y = R = H and X = Y = H, R = CH_3) were deemed to be outliers on the basis of their deviation between observed and calculated activities from the equation (Obsd − Pred > 2s). The exact reasons for these deviations are not very clear, but they may be due to the simultaneous use of the calculated/experimental $\Delta\delta_{(\beta-\alpha)}$ values. It must be noted that all the compounds except these two outliers have the experimental $\Delta\delta_{(\beta-\alpha)}$ values. Thus, the calculated $\Delta\delta_{(\beta-\alpha)}$ values may not be correct.

4.2.2. Antimalarial Activity. Malaria, caused by the protozoal species *Plasmodium*, continues to be one of the most devastating infectious tropical diseases, with approximately 243 million infections and 863,000 deaths globally in 2009 alone.[114,115] These occur particularly among children and women, primarily in Africa. The most lethal form of malaria is caused by *Plasmodium falciparum*, and its prevalence and rapid spread, coupled with the emergence of global resistance to the commonly used antimalarials, suggest an urgent need for the development of new antimalarial drugs in order to combat the record numbers of malarial infections.[116−118] A series of cinnamic acid derivatives (**XXXI**) was synthesized and evaluated for their inhibitory activity against intraerythrocytic forms of *P. falciparum* strain Dd2 using a semiautomated microdilution assay by Wiesner and co-workers.[119] We converted their inhibition data (IC_{50}; μM) into activity (log 1/IC_{50}; mol/L) and used it in the development of the following QSAR models 53 and 54 (see data in Table 21):



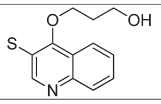
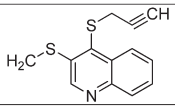
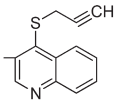
XXXI

$$\log 1/IC_{50} = 4.96(\pm 2.20)C\log P - 0.32(\pm 0.15)C\log P^2 - 0.86(\pm 0.23)\sigma^+ - 13.61(\pm 7.82) \quad (53)$$

where n = 14, r^2 = 0.909, s = 0.187, q^2 = 0.826, and $F_{3,10}$ = 33.296

dx.doi.org/10.1021/cr100125d |*Chem. Rev.* 2011, 111, 2865–2899

Table 22. Experimental ^{13}C NMR Chemical Shifts ($\delta_{C\text{-}3}$ and $\delta_{C\text{-}4}$) and Biological (log $1/ID_{50}$; mol L^{-1}) Parameters of Quinoline Analogues (XXXIV) Used To Derive eqs 56 and 57

No.	X	Y	log $1/ID_{50}$ (obsd.)[a]	log $1/ID_{50}$ (eq 56)[b]		log $1/ID_{50}$ (eq 57)[b]		$(\delta_{C\text{-}3})$[a]	$(\delta_{C\text{-}4})$[a]
				pred.	Δ	pred.	Δ		
1		$CH_2C{\equiv}CH$	5.08	4.89	0.19	5.00	0.08	136.45	136.60
2	$SCH_2C{\equiv}CH$		4.73	4.82	-0.09	4.80	-0.07	135.09	141.64
3	$SeCH_2C{\equiv}CH$	CH_3	5.21	5.05	0.16	5.05	0.16	139.52	135.26
4	$SCH_2C{\equiv}CH$	CH_3	4.63	5.02	-0.39	5.01	-0.38	139.01	136.31
5	$OCH_2C{\equiv}CH$	CH_3	4.04	4.14	-0.10	4.11	-0.07	121.88	159.67
6	OCH_3	$CH_2C{\equiv}CH$	3.99	3.92	0.07	3.94	0.05	117.61	163.93
7	$SeCH_2C{\equiv}CH$	C_2H_5	5.05	4.99	0.06	4.98	0.07	138.40	137.04
8	$SCH_2C{\equiv}CH$		4.95	4.85	0.10	4.79	0.16	135.73	141.93

[a] Data from ref 123. [b] Calculated from the C-QSAR program (ref 17).

optimum $C\log P = 7.75(7.48-8.38)$
outliers: X = NO_2 and $COOCH_3$

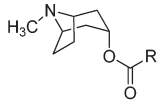
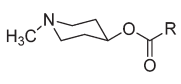
$$\log 1/IC_{50} = 4.15(\pm 2.85)C\log P - 0.27(\pm 0.20)C\log P^2 - 0.10(\pm 0.03)\delta_{C1} + 2.25(\pm 10.11) \quad (54)$$

where n = 15, r^2 = 0.861, s = 0.259, q^2 = 0.710, and $F_{3,11}$ = 22.712
optimum $C\log P = 7.82(7.41-10.12)$
outliers: X = NO_2 and $COOCH_3$

δ_{C1} is the calculated ^{13}C NMR chemical shift (ppm) of the C1 carbon, σ^+ is the Brown variant of the Hammett electronic parameter for X substituents, and $C\log P$ is the calculated partition coefficient in *n*-octanol/water and is a measure of hydrophobicity for the whole molecule. Parabolic correlations (eqs 53 and 54) in terms of $C\log P$ suggest that the antimalarial activities of cinnamic acid derivatives (**XXXI**) against intraerythrocytic forms of the *P. falciparum* strain Dd2 first increase with an increase in their hydrophobicity up to an optimum $C\log P$ of ~7.80 and then decrease. QSARs 53 and 54 are almost parallel, and their identical optimum $C\log P$ values (7.75 and 7.82) suggest that the δ_{C1} values for this data set can be used as an alternative descriptor for that of σ^+, which is supported further by the strong mutual correlation between δ_{C1} and σ^+ (r = 0.964). The advantage of using δ_{C1} as an alternative descriptor for σ^+ is that the QSAR 54 has predicted the activity of compound **XXXI** (X = $CH(CN)_2$), while the QSAR 53 is unable to predict the activity of this analogues due to the unavailability of the σ^+ value for their $CH(CN)_2$ substituent. The negative ρ^+ value of the σ^+ in eq 53 suggests that the antimalarial activity of compounds (**XXXI**) further increases with the electron-donating X substituents and that the same has been represented alternatively by the δ_{C1} term in eq 54 with a negative coefficient. Thus, an upfield chemical shift of the C1 carbon in ^{13}C NMR (electron-donating X substituent effect) indicates the increase in the antimalarial activity.

Two same data points (X = NO_2 and $COOCH_3$) in both the QSARs were considered to be outliers and excluded from the analysis since their deviations were greater than twice the standard deviation of the regression line. The unusual behavior of the nitro analogue could be attributed to the formation of the active nitro anion radical and its subsequent reduction to a nitrosobenzene and/or phenylhydroxylamine.[25,120] The reason for the second outlier (X = $COOCH_3$) is not very clear, but it may be associated with the experimental error, or the used parameters may not be the best.

4.2.3. Antimuscarinic Activity. A series of tropinyl esters (**XXXII**) and piperidinyl esters (**XXXIII**) was synthesized and evaluated for their inhibitory activities against the endothelial muscarinic receptors of rat (M_3) and rabbit (M_2) aorta by the research group of Xu et al.[121] A QSAR analysis was then performed for which M_2 and M_3 antagonistic activities (log $1/EC_{50}$) of esters **XXXII** and **XXXIII** (including atropine) were well correlated with their two electronic parameters (Taft's polar substituent constant σ^* for the group R in the acyl side chain R-CO- and the experimental ^{13}C NMR chemical shift difference of the carbonyl carbon $\Delta\delta$). 80% and 90% variances, respectively, of the M_2- and M_3- inhibitory activities of these esters were accounted for their two electronic parameters (σ^* and $\Delta\delta$) of the acyl side chain. The QSAR model for the M_3- inhibitory activities has been shown in eq 55:[121]

XXXII **XXXIII**

R = CH_2Ph, $CH(CH_3)Ph$, $CH(Ph)_2$, $C(CH_3)(Ph)_2$, $C(OH)(Ph)_2$, $CH(C_6H_{11})Ph$, $C(CH_3)(C_6H_{11})Ph$, $C(OH)(C_6H_{11})Ph$, CH_3

$$\log 1/EC_{50(M3)} = 2.47(\pm 0.36)\sigma^* + 0.78(\pm 0.08)\Delta\delta + 4.98(\pm 0.23) \quad (55)$$

where n = 19, r^2 = 0.900, s = 0.480, q^2 = 0.850, and $F_{2,16}$ = 69.910.

log $1/EC_{50(M3)}$ is the M_3 antagonistic activities of esters **XXXII** and **XXXIII** (including atropine), where $EC_{50(M3)}$ is the molar concentration required to inhibit 50% of the (ACh)-induced relaxation of phenylepherine-induced precontracted endothelium-intact rat aortic rings. $\Delta\delta$ is the difference between

dx.doi.org/10.1021/cr100125d |*Chem. Rev.* 2011, 111, 2865–2899

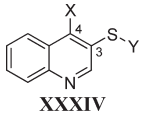
the experimental ^{13}C NMR chemical shifts of the carbonyl carbon in the ester (δ_i) and that of the methyl ester (δ_{Me}) of tropine (171.543 ppm) or *N*-methylpiperidine-4-ol (170.943 ppm). It must be noted that the Taft's polar substituent constant σ^* is a direct measure of the electron-withdrawing capacity of the R group while $\Delta\delta$ is an indirect measure of the electronic effect of the R group. Since $\Delta\delta$ does not provide the exact level of the mechanistic insight as σ^*, the variations in $\Delta\delta$ might be attributed to a complex mixture of both the steric and the electronic effects.[122]

In addition to the conventional QSAR, a comparative molecular field analysis (CoMFA) was also performed by the same research group using the same data set of esters.[121] The advantage of this analysis is that it allows us to visualize the electrostatic and steric regions in three-dimensional space. The statistics of the developed CoMFA models for the esters are as follows:

CoMFA model for the M_2 antagonistic activities (log $1/EC_{50(M2)}$):
$n = 19$, $r_{cv}^2 = 0.820$, $s_{cv} = 0.870$, principal components = 3.
Relative contribution: steric = 0.69% and electrostatic =0.31%.
CoMFA model for the M_3 antagonistic activities (log $1/EC_{50(M3)}$):
$n = 19$, $r_{cv}^2 = 0.720$, $s_{cv} = 0.840$, principal components = 2.
Relative contribution: steric = 0.69 and electrostatic = 0.31.

r_{cv}^2 and s_{cv} are respectively the cross-validated r^2 and the cross-validated standard error of the models. The CoMFA results for the M_2 and M_3 antagonistic activities of esters suggest the contributions of both the steric and electrostatic effects. In contrast, the conventional QSAR models suggest only the contribution of electrostatic parameters (σ^* and $\Delta\delta$). A possible explanation for this is the high correlations between $\Delta\delta$ and log MV ($r = 0.739$) and between $\Delta\delta$ and $N - 3$ ($r = 0.827$), and thus, the former ($\Delta\delta$) is reflected into the combined steric and electrostatic components of CoMFA where MV is the molecular volume (size) and *N* is the number of hyperconjugable αC−H atoms.[121,122]

4.2.4. Antiproliferative Activity. Boryczka et al.[123] synthesized a series of quinoline derivatives (**XXXIV**) and evaluated for their *in vitro* antiproliferative activities against T47D (breast cancer) cells, and they also developed various QSAR models using the experimental ^{13}C NMR chemical shift of a selected carbon atom as an electronic descriptor(s). The best QSARs from their results were the following: ($ID_{50} = -0.99\delta_{C\text{-}3} + 141.01$; $n = 8$, $r^2 = 0.960$) and ($ID_{50} = 0.74\delta_{C\text{-}4} - 97.46$; $n = 8$, $r^2 = 0.970$). Since we do not prefer to use ID_{50} as dependent variable in the QSAR model development, the ID_{50} (μg/mL) values of compounds (**XXXIV**) were transformed into log $1/ID_{50}$ (mol/L) and the following QSAR models 56 and 57 were developed (see data in Table 22):



XXXIV

$$\log 1/ID_{50} = 0.051(\pm 0.022)\delta_{C\text{-}3} - 2.10(\pm 3.00) \quad (56)$$

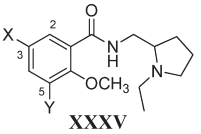
where $n = 8$, $r^2 = 0.838$, $s = 0.204$, $q^2 = 0.733$, and $F_{1,6} = 31.037$.

$$\log 1/ID_{50} = -0.039(\pm 0.016)\delta_{C\text{-}4} + 10.27(\pm 2.25) \quad (57)$$

where $n = 8$, $r^2 = 0.860$, $s = 0.190$, $q^2 = 0.779$, and $F_{1,6} = 36.857$.

$\delta_{C\text{-}3}$ and $\delta_{C\text{-}4}$ are the experimental ^{13}C NMR chemical shifts (ppm) of the C-3 and C-4 carbons, respectively, for quinoline derivatives (**XXXIV**). It is well established that the ^{13}C NMR chemical shifts of the quinoline carbons are sensitive to electron density, and they shift upfield in an electron-rich environment.[124] In the present case, signals for C-3 of the quinoline for the more active derivatives are shifted downfield in comparison to that of the corresponding less active compounds and vice versa for the signals of the C-4 carbon. Since there is an excellent mutual correlation between $\delta_{C\text{-}3}$ and $\delta_{C\text{-}4}$ ($r = 0.993$), $\delta_{C\text{-}3}$ and $\delta_{C\text{-}4}$ explain almost similar variance, 84% and 86%, respectively, in the data set of QSARs 56 and 57. But results from both the QSARs are not identical. A positive coefficient (+0.051) associated with $\delta_{C\text{-}3}$ in eq 56 suggests that the antiproliferative activity of quinolines (**XXXIV**) increases with downfield chemical shift in the C-3 carbon. On the contrary, a negative coefficient (−0.039) associated with $\delta_{C\text{-}4}$ in eq 57 suggests that the upfield chemical shift in the C-4 carbon increases the antiproliferative activity of quinolines. On combining both results, one can say that the upfield and downfield chemical shifts, respectively, of the C-4 and C-3 carbons will increase the antiproliferative activity of quinolines (**XXXIV**) against T47D cells.

4.2.5. Antipsychotic/Antidopaminergic Activity. To assess the influence of structural factors on dopamine D_2 receptor binding, the *in vitro* dopamine D_2 receptor affinity (pK_i) data of different series of benzamides were subjected to QSAR analysis. From the dopamine D_2 receptor affinity (pK_i) data of a series of orthopramides (**XXXV**), the QSAR model (eq 58) was developed.[125,126]



XXXV

X = H, Cl, Br, I, NH_2SO_2, CH_3, SCH_3, C_2H_5 or *n*-Pr;
Y = H, OCH_3, Cl, Br or OH

$$pK_i = 0.254(\pm 0.043)\delta_{C\text{-}2} + 0.030(\pm 0.009)\delta_{C\text{-}3} + 0.083(\pm 0.011)\delta_{C\text{-}5} + 0.026(\pm 0.007)\sum MR - 46.100(\pm 8.000) \quad (58)$$

where $n = 18$, $r^2 = 0.899$, $s = 0.308$, $q^2 = 0.818$, and $F_{4,13} = 29.000$.

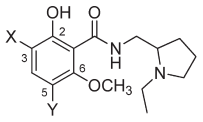
$\delta_{C\text{-}2}$, $\delta_{C\text{-}3}$, and $\delta_{C\text{-}5}$ are the experimental ^{13}C NMR chemical shifts of the C-2, C-3, and C-5 carbons, respectively. ΣMR is the sum of the molar refractivity of substituents X and Y. A positive coefficient associated with all the descriptors of QSAR 58 demonstrates that the dopamine D_2 receptor affinity of compounds (**XXXV**) increases by increasing the downfield shift of the C-2, C-3, and C-5 carbons (electronic influence of the ring substituents) as well as the sum of the molar refractivity of the X and Y substituents (steric/polarizability effect). Of course, the ^{13}C NMR chemical shifts of the C-2 and C-5 carbons in the aromatic ring are the most crucial structural properties for the affinity of orthopramides (**XXXV**) to the dopamine D_2 receptor. It must be noted that the chemical shift of the C-5 carbon is mainly influenced by the Y substituents, for which the Y substituents, such as OCH_3 and OH, increase the downfield shift and enhance the binding affinity, while the other Y substituents, such as H, Cl, and Br, decrease the downfield shift and diminish the binding affinity.

Table 23. Calculated ^{13}C NMR Chemical Shifts ($\delta_{C\text{-}4b}$ and $\delta_{C\text{-}7}$), *C* log *P*, and Biological (log 1/MIC; mol L^{-1}) Parameters of Polyhydroxyxanthones (XXXVIII) Used To Derive eqs 62 and 63

			log 1/MIC (eq 62)[b]		log 1/MIC (eq 63)[b]				
no.	X	log 1/MIC (obsd)[a]	pred	Δ	pred	Δ	($\delta_{C\text{-}4b}$)[c]	($\delta_{C\text{-}7}$)[c]	(*C* log *P*)[b]
1	1,3-(OH)$_2$	3.12	3.19	−0.07	3.25	−0.13	155.60	124.30	3.06
2[d]	1,3,5-(OH)$_3$	3.12	2.61	0.51	3.01	0.11	145.40	122.90	2.43
3	1,3,6-(OH)$_3$	3.12	3.12	0.00	2.92	0.20	157.00	114.10	2.43
4	1,3,7-(OH)$_3$	3.30	3.18	0.12	3.31	−0.01	148.20	153.90	2.43
5	1,3,8-(OH)$_3$	3.12	3.07	0.05	3.21	−0.09	157.00	110.50	3.33
6	1,3,5,6-(OH)$_4$	2.52	2.53	−0.01	2.70	−0.18	146.40	113.10	1.84
7	1,3,6,7-(OH)$_4$	3.12	3.12	0.00	3.01	0.11	149.60	143.90	1.84
8	1,3,6,8-(OH)$_4$	3.00	2.96	0.04	2.83	0.17	158.40	96.40	2.67
9	1,3,5,6,8-(OH)$_5$	2.52	2.49	0.03	2.61	−0.09	149.40	97.80	2.01
10	1,3,6,7,8-(OH)$_5$	2.82	3.00	−0.18	2.94	−0.12	151.00	128.90	2.08

[a] Data from ref 128. [b] Calculated from the C-QSAR program (ref 17). [c] Calculated from the ChemBioDraw Ultra 12 program (ref 20). [d] Not included in the derivation of QSAR eq 62.

In another attempt, a QSAR model (eq 59) was developed from the dopamine D_2 receptor affinity (pK_i) data of a series of salicylamides (**XXXVI**) by the same research group.[125,126]



XXXVI

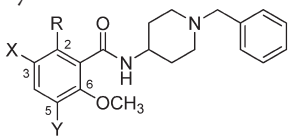
X = H, F, Cl, Br, I, OCH_3, CH_3, C_2H_5, *n*-Pr, *n*-Bu or NO_2
Y = H, F, Cl, Br, I, OCH_3, CH_3, C_2H_5, *n*-Pr, OH, NH_2 or NO_2

$$pK_i = 2.10(\pm 0.36)\pi_X - 0.754(\pm 0.217)\pi_X^2 - 0.060(\pm 0.017)\delta_{C\text{-}6} + 10.60(\pm 2.30) \quad (59)$$

where $n = 48$, $r^2 = 0.668$, $s = 0.543$, $q^2 = 0.627$, and $F_{3,44} = 28.81$
optimum $\pi_X = 1.39$ (1.10−2.30) (Note: The optimum value of π_X was added to the original eq 59)

This is a parabolic correlation in terms of π_X (hydrophobic parameter of X substituents) followed by an electronic parameter $\delta_{C\text{-}6}$ (experimental ^{13}C NMR chemical shifts of C-6 carbon). In this model, 55.2% of the variance in the data is explained by the hydrophobic descriptor π_X while the electronic descriptor $\delta_{C\text{-}6}$ accounts for only 11.6% of the variance in the data. This strong dependence on mostly the hydrophobic term (π_X) shows how the hydrophobic X substituents increase the binding affinity of salicylamides (**XXXVI**) to the dopamine D_2 receptor. On the other hand, it is very difficult to explain the electronic effect because the chemical shift of the C-6 carbon is influenced by the both X and Y substituents. Of course, the negative coefficient associated with $\delta_{C\text{-}6}$ suggests that the affinity of compounds (**XXXVI**) decreases by increasing the downfield shift of the C-6 carbon.

A QSAR model (eq 60) was also developed from the dopamine D_2 receptor affinity (pK_i) data of a series of substituted benzamides (**XXXVII**) by De Paulis et al.[125]



XXXVII

R = H or OH; X = H, Cl, Br, C_2H_5 or *n*-Pr; Y = H, Cl, Br, OCH_3 or C_2H_5

$$pK_i = 0.030(\pm 0.006)\delta_{C\text{-}2} - 0.029(\pm 0.008)\delta_{C\text{-}3} - 0.226(\pm 0.025)\delta_{C\text{-}6} + 35.80(\pm 3.80) \quad (60)$$

where $n = 18$, $r^2 = 0.871$, $s = 0.291$, and $F_{3,14} = 31.50$.

It is interesting to note that the above QSAR suggests the electronic contributions of only the ^{13}C NMR chemical shifts (experimental) of the C-2, C-3, and C-6 carbons of the aromatic ring. The positive coefficient of $\delta_{C\text{-}2}$ suggests that the more downfield shift at the C-2 carbon is favored for the increased activity. The chemical shift of the C-2 carbon is mainly influenced by the R substituent; OH increases the chemical shift and enhances the activity, while H decreases the chemical shift and diminishes the affinity. On the other hand, the negative coefficient associated with $\delta_{C\text{-}3}$ and $\delta_{C\text{-}6}$ suggests that the affinity of compounds (**XXXVII**) decreases by increasing the downfield shift of the C-3 and C-6 carbons. The chemical shift of the C-3 carbon, in particular, is mainly influenced by the X substituent; substituents such as C_2H_5 or *n*-Pr increase the chemical shift and diminish the affinity, while substituents such as H, Cl, and Br decrease the chemical shift and enhance the activity.

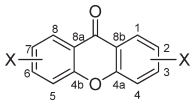
From the combined data of dopamine D_2 receptor affinity (pK_i) of two series of benzamides (**XXXV** and **XXXVI**), the QSAR model (eq 61) was developed.[126]

$$pK_i = 0.569(\pm 0.087)\log k_w + 0.377(\pm 0.086)\kappa_Y - 0.270(\pm 0.081)\delta_{C\text{-}6} + 0.25(\pm 0.09)I_{C\text{-}2} \quad (61)$$

where $n = 66$, $r = 0.834$, $s = 0.621$, and $F_{4,61} = 34.90$

log k_w is the overall molecular hydrophobicity obtained from HPLC at pH = 7.4, whereas κ_Y is the Livingstone's donor−acceptor parameter of Y substituents. $I_{C\text{-}2}$ is an indicator variable having value 0 for those compounds unsubstituted at the C-2 position and 1 for those compounds with the OH group at the C-2 position. The positive coefficient of the indicator variable suggests that the presence of the OH group at the C-2 position is favorable for the activity. The activity of benzamides (**XXXV** and **XXXVI**) increases with increasing their hydrophobicity, as evident by the positive coefficient of log k_w. Similarly, the increased value of κ_Y is also favored. Again the negative coefficient associated with $\delta_{C\text{-}6}$ suggests that the affinity of the benzamides diminishes by increasing the chemical shift of the C-6 carbon.

4.2.6. Antituberculous Activity. Polyhydroxyxanthones are known to possess antituberculotic activity against *Mycobacterium tuberculosis*.[127] A QSAR model between the MIC values of polyhydroxyxanthones (**XXXVIII**) and their ^{13}C NMR chemical shifts ($\delta_{C\text{-}4b}$ and $\delta_{C\text{-}7}$) with high statistics (n = 9, r = 0.969, s = 0.28, and F = 46.1) was published by Hambloch et al.[128] for which MIC represents the minimal inhibitory concentration values of polyhydroxyxanthones (**XXXVIII**) against *Mycobacterium tuberculosis* in micromoles per milliliter, while $\delta_{C\text{-}4b}$ and $\delta_{C\text{-}7}$ are the calculated ^{13}C NMR chemical shifts of C-4b and C-7 carbons in parts per million, respectively. Since we do not prefer to use MIC as the dependent variable in the QSAR model development, the MIC (μmol/mL) values of compounds (**XXXVIII**) were transformed into log 1/MIC (mol/L) and the following QSAR model 62 was developed (see data in Table 23):



XXXVIII

$$\log 1/\mathrm{MIC} = 0.054(\pm 0.02)\delta_{C\text{-}4b} + 0.014(\pm 0.005)\delta_{C\text{-}7} - 6.97(\pm 3.38) \quad (62)$$

where n = 9, r^2 = 0.916, s = 0.094, q^2 = 0.821, and $F_{2,6}$ = 32.714
outlier: X = 1,3,5-$(OH)_3$
C log P vs $\delta_{C\text{-}4b}$; r = 0.765

The positive coefficient associated with $\delta_{C\text{-}4b}$ and $\delta_{C\text{-}7}$ suggests an increase in the activity of compounds (**XXXVIII**) with their both downfield shifted C-atoms. An upfield shift of the C-4b carbon is mainly due to the presence of a hydroxyl group in position 5, and to a lesser extent from a hydroxyl group in position 7. On the other hand, the strong downfield shift is observed due to the presence of a hydroxyl group in position 7. Compounds with hydroxyl groups in both positions 5 and 6 (X = 1,3,5,6-$(OH)_4$ and X = 1,3,5,6, 8-$(OH)_5$) have upfield shifted C-4b and C-7 carbons and thus displayed poorest activity (log 1/MIC = 2.52 mol/L) in the data set. Thus, the C-7 carbon must be substituted, while the C-5 and C-6 carbons should not be. One data point (X = 1,3,5-$(OH)_3$) was deemed to be an outlier and excluded from the analysis, since their deviation was more than five times the standard deviation from the regression line. The reason for this outlier is not very clear. Since there is a good mutual correlation between the hydrophobic parameter of the compounds and their ^{13}C NMR chemical shifts of the C-4b carbon, QSAR 63 was developed to see the influence of the hydrophobicity of the compounds (see data in Table 23):

$$\log 1/\mathrm{MIC} = 0.358(\pm 0.252)C\log P + 0.010(\pm 0.007)\delta_{C\text{-}7} + 0.925(\pm 1.117) \quad (63)$$

where n = 10, r^2 = 0.735, s = 0.157, q^2 = 0.502, and $F_{2,7}$ = 9.708.

Positive coefficients of C log P and $\delta_{C\text{-}7}$ suggest that the activity of compounds (**XXXVIII**) increases with increasing their hydrophobicity and downfield shifted C-7 atom. Although QSAR 63 has no outlier, it displays only fair statistics that are not much improved by dropping one data point. The best statistics are coming from QSAR 62.

QSAR studies were also carried out by the research group of Schaper et al.[129] between the antituberculous activities of a series of 61 substituted xanthones (**XXXVIII**, where X = CH_3, CN, COOH, $COOCH_3$, $CONH_2$, NO_2, and NH_2) and their experimental ^{13}C NMR chemical shifts, hydrophobicity, and molar refractivities of the substituents. In addition to these parameters, the test concentrations of the compounds were also used as input descriptor because of the varying solubility, leading to 127 data points obtained from 61 compounds. The Adaptive Least Squares (ALS) technique based on an MLR with stepwise adaptation was used to develop a number of QSAR models. The best model for the biological activity (two activity classes) against *M. tuberculosis* was established with the following statistics: n = 61, n_1 = 127, n_{mis} = 9, R_S = 0.765, $R_{S(LOO)}$ = 0.489, and E = 0.842. For which n, n_1, and n_{mis} are, respectively, the number of compounds, measurements, and misclassifications. R_S is the Spearman rank correlation coefficient between observed and calculated activity classes, while $R_{S(LOO)}$ is the cross-validated R_S obtained from the leave-one-out procedure. E represents the error function.

In addition to the conventional QSAR models, the Artificial Neural Networks (ANN) model was also performed on the same data set (**XXXVIII**; where X = CH_3, CN, COOH, $COOCH_3$, $CONH_2$, NO_2, and NH_2) in order to compare the two methods (ALS vs ANN). Using a three-layer back-propagation neural network, the final network was obtained with 11 input variables (experimental ^{13}C NMR chemical shifts, hydrophobicity, molar refractivities of the substituents, and applied concentration values) and 2 hidden layer neurons. The statistics of that ANN-QSAR were n = 61, n_1 = 127, n_{mis} = 4, and R_S = 0.897. These results of the ANN calculation were further validated by the leave-10%-out method. Although a direct comparison of these two QSAR (ALS and ANN) models is very difficult, the predictive power of these models can be compared by their n_{mis} and R_S values. However, both models are able to recognize the most important structural features determining the antituberculous activity of the substituted xanthones using ^{13}C NMR chemical shifts, molar refractivities, and log P values as physicochemical descriptors in addition to the applied concentration (log C).[129] In ANN-QSAR, the QSAR information is being stored in the weight matrices representing connection strengths between neurons, whereas a single coefficient per variable is obtained in the case of ALS analyses, thus facilitating the interpretation of the ALS model.[130]

4.2.7. Binding to Aromatase Enzyme. Aromatase is an important enzyme in the evolution and development of estrogen-dependent tumors because it catalyzes the conversion of testosterone to estradiol. Thus, the inhibition of aromatase enzyme is therapeutically significant to control breast cancer.[131,132] Five quantitative spectroscopic data−activity relationships (QSDARs) based on simulated ^{13}C NMR spectral data were developed for the binding activities of 50 steroids to aromatase enzyme and compared with the conventional QSAR model as well as CoMFA.[133] Out of these five QSDAR models, three of them were based on comparative spectral analysis (CoSA), and the other two models were based on comparative structurally assigned spectral analysis (CoSASA). For the development of CoSA models, the ^{13}C NMR peak area of a certain spectral range was normalized to an integer; for example, a single chemical shift frequency in the 1.0 ppm spectral bin had an area of 100, and two chemical shifts under similar conditions had an area of 200, and so on. The two CoSASA models were developed either by the use

of ^{13}C NMR chemical shifts from the steroidal backbone at five selected positions 3, 6, 7, 9, and 12 or by the use of the 5 most correlated principal components (PCs) built from all the assigned ^{13}C NMR chemical shifts in the steroidal backbone. The statistical parameters of all the QSDAR/QSAR models for the binding activities of 50 steroids to aromatase enzyme and the number/type of components used in their development are listed in Table 24.[133]

All five QSDAR models showed better r^2 and q^2 values than that of the classical QSAR model obtained from five structural parameters (see Table 24). It may be possible only by considering that the ^{13}C NMR spectral data used in the development of QSDAR models, based on quantum mechanical principles, is more reflective to binding than the QSAR model's calculated structural parameters. On the other hand, the CoMFA model has better r^2 and q^2 values than that of the CoSA (except ANN-CoSA) and CoSASA models. A possible explanation for this situation is that the ^{13}C NMR spectra based on substructure predictions are only a two-dimensional representation of a three-dimensional molecule. Thus, important 3D information might be lost during the conversion. The Artificial Neural Network CoSA (ANN CoSA) model is able to reproduce an r^2 value of 1.00, but it is not surprising because all the 87 bins of ^{13}C NMR spectral data were used in the development of this model.[133]

Comparative structural connectivity spectra analyses (CoSCoSA) were also carried out for the binding activity of the same data set of 50 steroids to the aromatase enzyme by Beger and Wilkes.[134] This technique is also referred to as 3D-QSDAR modeling. In this method, the 3D-connectivity matrix of molecules was developed by combining the ^{13}C NMR spectral and structural information for modeling of their biological activity. All the four 2.0 ppm resolution CoSCoSA models for the binding activity of 50 steroids to the aromatase enzyme were developed using the combinations of 2D COSY and 2D long-range distance spectra as modeling parameters. A CoSCoSA model based on $^{13}C-^{13}C$ COSY spectra and nine PCs gave an r^2 and a cross-validated r^2 (q^2) of 0.89 and 0.89, respectively. Similarly, the CoSCoSA model based on a $^{13}C-^{13}C$ distance greater than 6.0 Å and seven PCs gave an r^2 and a q^2 of 0.72 and 0.72, respectively. It is interesting that the CoSCoSA model based on the combined $^{13}C-^{13}C$ COSY and $^{13}C-^{13}C$ distance connectivity and 10 PCs provided an excellent r^2 and q^2 of 0.92 and 0.86, respectively. However, the statistics were reduced for the CoSCoSA model based on the combined $^{13}C-^{13}C$ COSY and $^{13}C-^{13}C$ distance connectivity and eight PCs when the COSY and distance data were combined before PC extraction. This model had an r^2 and a q^2 of 0.87 and 0.83, respectively. It must be noted that all the CoSCoSA models of this study have comparative statistics than that of the CoMFA model (see Table 24). The advantage of the CoSCoSA modeling includes the ease of its development, since this technique does not require molecular docking.[134]

4.2.8. Binding to Aryl Hydrocarbon Receptor (AhR). QSDARs for the binding affinities of polychlorinated dibenzofurans (PCDFs), dibenzodioxins (PCDDs), and biphenyls (PCBs) to the aryl hydrocarbon receptor (AhR) were developed by Beger and Wilkes[135] based on MLR analysis of the simulated ^{13}C NMR data. A CoSA model with 1.0 ppm resolution for 26 PCDFs based on chemical shifts in 5 bins provided an r^2 and q^2 of 0.93 and 0.90, respectively. The exact similar CoSA model for 12 PCBs yielded an r^2 and q^2 of 0.87 and 0.45, respectively. A 2.0 ppm resolution CoSA model for 14 PCDDs also based on the chemical shifts in 5 bins presented an r^2 and q^2 of 0.91 and 0.81, respectively. A 1.0 ppm resolution CoSA model for the combined data set of 52 compounds was finally performed based on chemical shifts in 12 bins that gave an r^2 of 0.85 and q^2 of 0.71. A canonical variance analysis of the 1.0 ppm resolution CoSA model for all 52 compounds when they were splitted into 27 strong and 25 weak AhR binders was obtained as 98% correct.[135]

Table 24. Statistical Parameters (r^2 and q^2) of Different QSDAR/QSAR Models for the Binding Activities of 50 Steroids to Aromatase Enzyme and Number/Type of Components Used in Their Development[a]

no.	QSDAR/QSAR model	r^2	q^2	components
1	CoSASA	0.75	0.66	5 atoms
2	CoSASA	0.75	0.67	5 PCs
3	CoSA	0.82	0.77	5 bins
4	CoSA	0.78	0.71	5 PCs
5	ANN-CoSA	1.00	0.75	all 87 bins
6	QSAR	0.73	0.66	5 parameters
7	CoMFA	0.94	0.72	5 PCs

[a] Adapted with permission from ref 133. Copyright 2001 American Chemical Society.

CoSCoSA (3D-QSDAR) models were also developed for the binding affinities of PCDFs, PCDDs, and PCBs to the AhR using forward multiple linear regression analysis of the predicted ^{13}C NMR structure-connectivity spectral bins.[136] In this method, two-dimensional $^{13}C-^{13}C$ COSY and 2D slices from the distance dimension of the 3D-connectivity matrix were used to produce a relationship among the 2D spectral patterns for PCDFs, PCDDs, and PCBs binding to the AhR. A CoSCoSA model for 26 PCDFs presented an r^2 and an average leave-four-out cross-validated r^2 ($q_4{}^2$) of 0.93 and 0.89, respectively. Similarly, CoSCoSA models for 14 PCDDs and 12 PCBs, respectively, produced r^2 of 0.90 and 0.91, and average leave-two-out cross-validated r^2 ($q_2{}^2$) of 0.79 and 0.80. Another CoSCoSA model was developed for the whole data set of 52 compounds, which gave an r^2 of 0.85 and an average $q_4{}^2$ of 0.52.[136]

Spectroscopic data−activity relationships (SDARs) for the binding affinities of PCDFs and PCDDs to the AhR were developed by Shade et al.[137] using discriminant function analysis of the simulated ^{13}C NMR data. In this modeling, raw population values in the spectral bins (100, 200, etc.) were Fisher weighted prior to pattern recognition. SDAR models with two classifications were developed from the experimental AhR binding affinities of 14 PCDDs, 26 PCDFs, and 40 PCDDs & PCDFs. All these three models (derived from the first four PCs) showed a LOO cross-validation accuracy of 92.8%, 88.5%, and 87.5%, while they used 95%, 88.5%, and 82.4%, respectively, of the total variance in the data. The best SDAR model for the combined data set (14 PCDDs and 26 PCDFs) based on four classifications and the first eight PCs explained 91.7% of the total variance in the data and showed a LOO cross-validation accuracy of 85%.[137]

4.2.9. Binding to Estrogen Receptor (ER). Two SDAR models based on experimental/calculated ^{13}C NMR and electron ionization mass spectra (EI MS) were developed by Beger et al.[138] for the relative binding affinities (RBA) of 108 compounds to the estrogen receptor (ER). The RBA of a molecule to the estrogen receptor is the ratio of the molar concentrations of 17β-estradiol and the test compound required to decrease the

Table 25. Statistical Parameters of Different QSDAR Models and Number of Principal Components (PCs) Used in Their Development[a]

no.	QSDAR model	no. of PCs	r^2	q^2	SE	S_{press}	F
1	EEVA	3	0.91	0.42	0.40	0.99	104
2	EVA	2	0.47	0.22	0.94	1.14	14.5
3	CoSA (^{13}C)	6	0.96	0.69	0.29	0.76	105
4	CoSA (^{1}H)	3	0.75	0.35	0.64	1.05	33.5

[a] Adapted with permission from ref 142. Copyright 2003 American Chemical Society.

receptor-bound compound by 50% then multiplied by 100. Thus, 17β-estradiol had a log RBA of 2.0. The first SDAR model based on ^{13}C NMR and 22 PCs yielded an r^2 and q^2 of 0.90 and 0.75, respectively. The second SDAR model based on ^{13}C NMR, EI MS, and 21 PCs gave an r^2 of 0.80 and a q^2 of 0.82. These two SDAR models were then successfully used to divide the 108 modeled compounds into three classifications: 73 weak, 15 medium, and 20 strong relative binding classifications.[138]

Another SDAR model based on ^{13}C NMR spectral data and 8 PCs was also developed by Beger et al.[139] for 30 estrogenic chemicals of known RBA to the alpha (ERα) and beta (ERβ) estrogen receptors. The SDAR model showed the total variance of 81.1% and a cross-validation of 90%. The significance of the model was 91.9% at one discriminant function with 28 degrees of freedom. On the other hand, QSAR models for both the α and β estrogen receptor bindings of a similar set of compounds based on 4 PCs had a 95% correlation between CoMFA calculated fields and the corresponding RBA of both the α and β estrogen receptor.[139,140]

Two CoSCoSA (3D-QSDAR) models were also developed for a large data set of 130 diverse organic compounds with known RBA to the estrogen receptor.[141] One CoSCoSA MLR model using 16 2D bins selected from the ^{13}C−^{13}C COSY spectral data gave r^2, q^2, and $q_{13}{}^2$ of 0.827, 0.780, and 0.780, respectively, for which $q_{13}{}^2$ is an averaged leave-13-out cross-validated r^2. A second CoSCoSA model based on the MLR analysis of 15 selected 2D ^{13}C−^{13}C COSY bins plus one additional distance−related variable gave r^2, q^2, and $q_{13}{}^2$ of 0.833, 0.790, and 0.780, respectively. From these two CoSCoSA models, it was concluded that the addition of more distance related variables should improve the predictive accuracy of the CoSCoSA models.[141]

A comparative QSAR study was also performed by Asikainen et al.[142] with respect to the RBA for a diverse set of 36 estrogens to ER using three "spectroscopic" QSAR methods: eigenvalue (EVA), electronic eigenvalue (EEVA), and comparative spectra analysis (CoSA). The statistical parameters of different QSDAR models and the number of principal components (PCs) used in their development are shown in Table 25.

The CoSA model with ^{13}C NMR chemical shifts is a predictive model that has the following statistics for their external predictability: r^2_{ex} = 0.54, q^2 = 0.56, S_{press} = 0.90, and pr-r^2 = 0.49. The EEVA model is in the borderline, whereas the performances of the EVA and CoSA models with ^{1}H NMR chemical shifts are substandard. On the basis of this comparative QSAR study, the authors suggested that the CoSA method with ^{13}C NMR chemical shifts is an alternative to the conventional 3D-QSAR methods for rationalizing and predicting the estrogenic activities of molecules.[142]

4.2.10. Binding to Corticosteroid Binding Globulin. Beger et al.[13,143] developed 10 QSDAR (2 CoSA, 2 CoSASA, and 6 CoSCoSA) models, based on the simulated ^{13}C NMR data, for the binding activities of 30 steroids to corticosteroid binding globulin. A comparative QSAR study was then carried out among these 10 QSDAR models and four previously developed QSAR (one each from CoMFA, HE-state/E-state, E-state, and SOM) models.[13] The statistical parameters of all of these QSDAR/QSAR models and the number/type of components used in their development are shown in Table 26.[13,143−146]

All the ten QSDAR models (Nos. 1−10) in Table 26 were developed with one less compound, because the authors did not used aldosterone due to its two structural conformations.[13,143] The best CoSCoSA (COSY + distance) model based on 8 PCs gave an r^2 of 0.96 and a q^2 of 0.92, which are better than those of the HE-state/E-state QSAR model.[13]

4.2.11. Binding to P450 LM2. Cytochromes P450 are ubiquitous heme-containing enzymes with thiolate from a protein cysteine as the fifth ligand to iron. In the human liver, broad-spectrum P450s oxidize drugs, pro-drugs, and other xenobiotics. P450 LM2 is a hepatic phenobarbital-induced P450 from rabbit, which is also known as CYP2B4. It was the first mammalian P450 that solubilized from microsomal membranes.[147−149] Petzold et al.[150] evaluated the binding affinities (K_s) of 12 tertiary alkyl amines (can be regarded as the derivatives of benzphetamine) to soluble cytochrome P450 LM2 and liposomal cytochrome P450 LM2. The structures of those 12 benzphetamine derivatives were as follows: $(C_6H_5CH_2)_2NCH_3$, $(C_6H_5CH_2CH_2)_2NCH_3$, $(C_6H_5CHCH_3)_2NCH_3$, $(2\text{-Cl-}C_6H_4CH_2)_2NCH_3$, $(4\text{-Cl-}C_6H_4CH_2)_2NCH_3$, $C_6H_5CH_2CH_2N(CH_3)CH_2C_6H_5$, $2\text{-Cl-}C_6H_4CH_2N(CH_3)CH_2C_6H_5$, $C_6H_5CH_2C(CH_3)_2N(CH_3)CH_2C_6H_5$, $C_6H_5CH_2CH(CH_3)N(CH_3)CH_2C_6H_5$, $C_6H_5CH_2CH(CH_3)N(CH_3)CH_2C_6H_4\text{-4-Cl}$, $2\text{-Cl-}C_6H_4CH_2N(CH_3)CH_2C_6H_4\text{-4-Cl}$, and $4\text{-Cl-}C_6H_4CH_2N(CH_3)CH_2C_6H_5$. The ^{13}C NMR chemical shifts of the *N*-methyl (δ_{CH3}) and the quaternary phenyl (δ_{Cqu}) carbons of the benzphetamine derivatives were used to correlate their binding affinities (K_s) to the soluble cytochrome P450 LM2 as shown by eqs 64 and 65, respectively.

For N-methyl carbon:

$$\ln K_s = 2.70(\pm 0.54)\delta_{CH3} - 118(\pm 21) \qquad (64)$$

where n = 12, r = 0.90, and $F_{1,10}$ = 42.632.

For quaternary phenyl carbon:

$$\ln K_s = 1.46(\pm 0.20)\delta_{Cqu} - 199(\pm 25) \qquad (65)$$

where n = 12, r = 0.94, and $F_{1,10}$ = 76.207.

These two identical QSARs, eqs 64 and 65, suggest that the binding affinity of benzphetamine derivatives to the soluble cytochrome P450 LM2 increases by downfield chemical shifts of the *N*-methyl (δ_{CH3}) and the quaternary phenyl (δ_{Cqu}) carbons. A decreased electron density of the quaternary phenyl carbon atom of benzphetamine derivatives with an upfield shift (δ_{Cqu}) lowered the electron density of the nitrogen via inductive effects; that is, it diminished the basicity, which results in a lowered solubility in aqueous solution.

4.2.12. Binding to Receptor Proteins. A CoSA model was developed by Bursi et al.[151] using simulated ^{13}C NMR data of a set of 45 progestagens. The binding affinities of that set of progestagens to receptor proteins in MCF-7 cells were taken from the literature and used as the biological activities.[152] The

Table 26. Statistical Parameters of Different QSDAR/QSAR Models for the Binding Activities of Steroids to Corticosteroid Binding Globulin and Number/Type of Components Used in Their Development[a]

no.	model	n	r^2	q^2	components	ref
1	CoSCoSA (COSY)	30	0.84/0.93	0.74/0.88	3 PCs/8 PCs	13
2	CoSCoSA (distance)	30	0.66/0.89	0.38/0.79	3 PCs/8 PCs	13
3	CoSCoSA (COSY + distance)	30	0.84/0.96	0.74/0.92	3 PCs/8 PCs	13
4	CoSCoSA (3D)[b]	30	0.78/0.92	0.68/0.81	3 PCs/8 PCs	13
5	CoSCoSA-PD-ANN (3D)[b]	30	0.96	0.78[c]	593:198:1	13
6	CoSCoSA-PD-ANN (3D)[b]	30	0.96	0.73[d]	593:198:1	13
7	CoSASA	30	0.80	0.73	3 atoms	143
8	CoSASA	30	0.68	0.60	3 PCs	143
9	CoSA (1.0 ppm)	30	0.80	0.78	3 bins	143
10	CoSA (1.0 ppm)	30	0.71	0.65	3 PCs	143
11	CoMFA	31		0.72[e]/0.69[f]	(3 PCs)[e]/(2 PCs)[f]	144
12	HE-state/E-state	31	0.98[e]/0.96[f]	0.80[e]/0.76[f]	(3 PCs)[e]/(5 PCs)[f]	145
13	E-state	31	0.96[e]/00.96[f]	0.79[e]/0.67[f]	(3 PCs)[e]/(4 PCs)[f]	145
14	SOM	31	0.85		3 PCs	146

[a] Adapted with permission from ref 13. Copyright 2002 American Chemical Society. [b] 3D is the combination of COSY and distance data before PC extraction. [c] A leave-three-out cross-validation r^2 (q_3^2). [d] A leave-ten-out cross-validation r^2 (q_{10}^2). [e] 1.0 Å models. [f] 2.0 Å models.

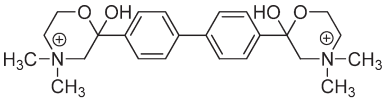
CoSA model based on simulated ^{13}C NMR chemical shifts in five bins yielded an r^2 and q^2 of 0.987 and 0.395, respectively. On the other hand, the CoMFA (rigid) model with four components gave an r^2 of 0.865 and a q^2 of 0.395, while the CoMFA (MIMIC) model with three components gave an r^2 of 0.841 and a q^2 of 0.391. These results suggest that the predictions from both the alignments of CoMFA models are comparable to that of the CoSA model.[151]

4.2.13. Biological Activity of Angiotensin II Analogues. The seven angiotensin II analogues Asp-Arg-Val-Tyr-X-His-Pro-Phe, in which only residue X in position 5 was varied, were synthesized and evaluated for their biological activities by Fauchère and Lauterwein.[153] The biological activity (BA), which was evaluated as a percent of that of the natural harmone, takes into account the potency to increase blood pressure in the pressure assay and the binding affinity in smooth muscle preparations. In this study, the authors showed that the chemical shift of the α-carbon in the 5-residue (δC ppm from glycine-C$_\alpha$) is the single most important parameter that accurately predicts the biological activity of angiotensin II (see eq 66).[153]

$$\log \mathrm{BA} = 0.095(\pm 0.030)\delta\mathrm{C} + 0.291 \qquad (66)$$

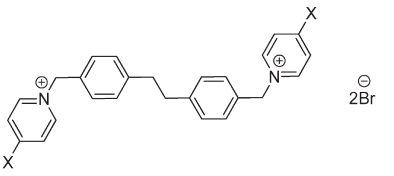
where $n = 7$, $r = 0.96$, $s = 0.20$, and $F_{1,5} = 65.70$.

4.2.14. Inhibition to Choline Kinase (ChoK). Choline kinase (ChoK) is an important enzyme responsible for the conversion of choline into phosphorylcholine (*P*Cho). The increase in ChoK activity results in increased levels of *P*Cho, a putative novel second messenger involved in the proliferation. Hemicholinium-3 (**XXXIX**) is one of the most potent ChoK inhibitors that blocks DNA synthesis stimulation by growth factors without any effect on the serum or insulin.[154,155]



XXXIX

In order to improve the potency of HC-3 (**XXXIX**), a series of compounds (**XXXX**) was synthesized and evaluated for their activity to inhibit ChoK under *ex vivo* conditions. The obtained result was responsible for the development of an excellent correlation (eq 67) between the *ex vivo* inhibitory potency of compounds (**XXXX**) and the experimental ^{13}C NMR chemical shifts (in CD_3OD) of their methylene group bearing the positively charged nitrogen (δ_{CH2N+}).[156]



XXXX

X = $N(CH_3)_2$, NH_2, CH_2OH, CH_3, COOH, CN, $N(CH_2CH{=}CH_2)_2$, Pyrrolidin-1-yl, Piperidin-1-yl, and Azepan-1-yl

$$\log 1/(\mathrm{IC}_{50})_{ex\ vivo} = -0.24(\pm 0.04)\delta_{CH2N+} + 19.35(\pm 1.20) \qquad (67)$$

where $n = 9$, $r = 0.978$, $s = 0.097$, and $F_{1,7} = 152.091$.

One compound X = CN was not considered in eq 67 due to its unclear activity. From eq 67, it is apparent that the ^{13}C NMR chemical shifts (δ_{CH2N+}) is a suitable descriptor for the QSAR modeling of compounds (**XXXX**) for their *ex vivo* inhibitory activities to ChoK. The negative coefficient associated with δ_{CH2N+} suggests that the *ex vivo* inhibitory activity of compounds (**XXXX**) increases with decreasing δ_{CH2N+} values. Since the value of δ_{CH2N+} increases with electron withdrawing X substituents, an electron releasing X substituent will be favored for its increased activity.

4.2.15. Inhibition to the Proliferation of Cancer Cell Lines. The inhibitory activity of compounds (**XXXX**) on the proliferation against HT-29 cell lines was also investigated by Campos et al.,[156] which showed a great correlation (eq 68) with the *ex vivo* inhibitory potency of these compounds against ChoK.

$$\log 1/(\mathrm{IC}_{50})_{antiprolif} = 1.48(\pm 0.17)\log 1/(\mathrm{IC}_{50})_{ex\ vivo} + 0.25(\pm 0.05)\sum f - 1.35 \qquad (68)$$

where $n = 7$, $r = 0.987$, and $s = 0.123$

Three compounds (X = CN, CH_2OH, and COOH) were not used in eq 68 due to their unclear activities. Σf is the sum of the Rekker hydrophobic fragmental constant for X substituents.

An excellent correlation (eq 69) between the antiproliferative activity and the electronic and the hydrophobic descriptors for the compounds (**XXXX**) was then developed by the same authors.[156]

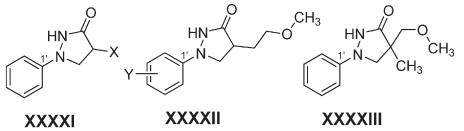
$$\log 1/(IC_{50})_{antiprolif} = -3.90(\pm 0.22)\delta^*_{CH2N+} + 0.33(\pm 0.04)\sum f + 29.68 \quad (69)$$

where n = 8, r = 0.996, s = 0.098, $F_{2,5}$ = 310, and δ^*_{CH2N+} = $0.1\delta_{CH2N+}$.

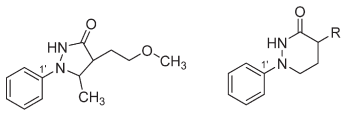
Two compounds (X = CH_2OH and COOH) were not used in eq 69 due to their unclear activities. In eq 69, the experimental descriptor δ_{CH2N+} was scaled by 0.1 (δ^*_{CH2N+}) to make it equiscalar to the hydrophobic term (Σf). Now one can easily ascertained that the electronic term is more important than the hydrophobic term in eq 69. The negative coefficient associated with the electronic parameter δ^*_{CH2N+} suggests that the electron releasing X substituent increases the activity. On the other hand, the more hydrophobic X substituent increases the activity, as evident by the positive coefficient of Σf. Thus, the more hydrophobic and electron releasing X group will be favored for the enhanced activity.

4.2.16. Inhibition to Lipoxygenase. Lipoxygenases (LOs) are a group of non-heme iron containing dioxygenases, which are classified according to their positional specificity of arachidonate oxygenation into 5-, 8-, 9-, 11-, 12-, and 15-LOs.[157] Of those, 5-LO is a key enzyme for the biosynthesis of leukotrienes (LT), powerful eicosanoid mediators involved in inflammation, in cell−cell communication, and in a number of other important pathological and physiological conditions.[158−160] It plays a critical role in Gastro Esophageal Reflux Disease (GERD), in tumor formation, and in cancer metastasis.[161] High expression of 5-LO was found in various cancer cell lines, for example, prostate, lung, colorectal, and others.[162−164] 5-LO is also expressed in brain cells and may participate in neuropathologic processes.[165] Inhibitors of the 5-LO, therefore, have a therapeutic potential in a variety of diseases.

Hlasta and co-workers[166] synthesized a series of 1-phenyl-3-pyrazolidinones (**XXXXI**−**XXXXIV**) and 1-phenyltetrahydro-3(2*H*)-pyridazinones (**XXXXV**), and they evaluated for their inhibitory activities to 5-LO. The biological data was then used in the development of a QSAR model (eq 70).[166]



XXXXI **XXXXII** **XXXXIII**



XXXXIV **XXXXV**

X = H, CH_2OCH_3, $CH_2OCH_2CH_3$, $CH_2OCH_2CH_2CH_3$, $CH_2OCH_2CH_2OCH_3$, $CH_2CH_2OCH_3$, $CH_2CH_2OCH_2CH_3$, CH_2-2-Thienyl, SCH_3, SCH_2CH_3, SCH_2Ph, S-2-Pyridinyl;
Y = 3,4-Cl_2, 3-CF_3, 4-Cl, 4-CH_3, 4-OCH_3, 4-OH;
R = H, $CH_2CH_2OCH_3$

$$\log 1/IC_{50} = 0.90(\pm 0.19)\log P + 0.15(\pm 0.06)CNMR1' - 24.19(\pm 8.67) \quad (70)$$

where n = 21, r^2 = 0.79, s = 0.49, $F_{2,18}$ = 33.60
outlier: **XXXXI** (X = SCH_2Ph)

The above QSAR model (eq 70) is a correlation between an increased inhibitor potency (log $1/IC_{50}$) and an increased compound hydrophobicity (log P) with an N-phenyl electronic effect as measured by the experimental ^{13}C NMR chemical shift parameter CNMR1′. Since the value of CNMR1′ increases with an electron withdrawing Y substituent, a strong electron withdrawing Y substituent will be favored for the enhanced activity of compound **XXXXII**. One compound, **XXXXI** (X = SCH_2Ph), was deemed to be an outlier and excluded from the analysis, since its deviation was greater than twice the standard deviation of the regression line. A possible reason for this outlier is not very clear, although its bulk and/or geometry due to the presence of a phenyl group may reduce the coplanarity with the X group.[167] The second possibility may include a significantly larger log P value of 3.63 for this outlier, which is 1.1 log units greater than the next lower value within the data set. Examination of the log P calculation for this compound showed several approximations were used for substituent constants.[166]

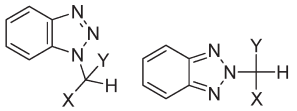
4.2.17. Intrinsic Sympathomimetic Activity (ISA). The β_2-adrenoceptor intrinsic activity of *N-tert*-butylphenylethanolamines (**XXXXVI**), in Table 27, was evaluated by IJzerman et al.[168] The obtained results were considered as the direct measures of intrinsic sympathomimetic activity (ISA). From those data, significant correlations were established and demonstrated that the ISA governs only electronic effects. The use of the experimental ^{13}C NMR chemical shift of the aromatic C atoms as a QSAR descriptor was proved to be a valuable tool in that analyses. The QSAR model (eq 71) was among the best one.

$$\log 1/ISA = -0.062(\pm 0.028)\Delta C_2 + 0.20(\pm 0.007)[\Delta C_1 + \Delta C_5] + 0.097(\pm 0.038)\Delta C_6 + 0.369(\pm 0.085) \quad (71)$$

where n = 11, r = 0.9175, s = 0.0997, and $F_{3,7}$ = 16.103.

ISA is the intrinsic sympathomimetic activity of compounds (**XXXXVI**) relative to salbutamol. $\Delta C_1 = \delta(C_{1\ compound}) - \delta(C_{1\ benzene})$ (in ppm) = $\delta(C_{1\ compound}) - 128.5$, as derived from ^{13}C NMR data. The value of ΔC_2 is directly influenced by the R_2 substituent and has a higher value for R_2 = Cl (the only substituent at the R_2 position in the present data set) as compared to an unsubstituted one. Thus, the unsubstituted R_2 position (R_2 = H) will be favored for the activity.

4.2.18. Plant Growth Regulators (Auxin-like Activity). In an effort to quantitatively rationalize the interactions of a large set of 1- and 2-substituted benzotriazole isomers (**XXXXVII**) active as plant growth regulators, various QSAR models were developed by Sparatore et al.[169] The best QSAR model using the experimental ^{13}C NMR data was eq 72.



(a) (b)

XXXXVII

X = H, Y = H, CN, CH_2OH, $(CH_2)_2OH$, $(CH_2)_4OH$, COOH, CH_2COOH, and $(CH_2)_3COOH$;
X = CH_3, Y = CH_3, $(CH_2)_3CH_3$, $CONH_2$, COOH, and $CH_2COOC_2H_5$;
X = $CH(CH_3)_2$, Y = $CONH_2$; X = $(CH_2)_3CH_3$, Y = $COOC_2H_5$;
X/Y = -$(CH_2)_4$- and -$(CH_2)_5$-

$$\log 1/C = -0.030(\pm 0.009)^{13}C\text{-}\delta_{a,b} + 0.20(\pm 0.06)\log D + 0.79(\pm 0.23)I + 4.94(\pm 0.52) \quad (72)$$

dx.doi.org/10.1021/cr100125d |*Chem. Rev.* 2011, 111, 2865–2899

where $n = 34$, $r = 0.921$, $s = 0.246$, and $F_{3,30} = 56.22$.

^{13}C-$\delta_{a,b}$ are the experimental ^{13}C NMR chemical shifts of the CH(X)(Y) carbons for 1-substituted benzotriazoles (δ_a) and the same for its 2-substituted derivatives (δ_b). log D is the logarithm of the octanol/buffer distribution coefficient, for which the distribution coefficient (D) of the compounds was determined by the classical "shake-flask" method, using *n*-octanol as the hydrophobic phase and phosphate buffer at pH 5.6 as the hydrophilic phase. The indicator variable (I) was used to explain the structural features of the compounds associated with the $-(CH_2)_nCOR'$ moiety, and its positive coefficient suggests that the activity of compounds increases with the presence of the $-(CH_2)_nCOR'$ moiety in the side chain. On the basis of this QSAR model, the authors concluded that the electronic and hydrophobic components are playing a major role in promoting the auxin-like activity of the benzotriazoles under consideration.[169]

4.2.19. Toxicity. From the cytotoxicity data of 4-X-phenols (where X = electron releasing groups) to CCRF-CEM cells, a QSAR (eq 73) dependent on the hydrophobic and electronic character was published by Selassie et al.[24]

$$\log 1/ID_{50} = 0.28(\pm 0.10)C\log P - 1.52(\pm 0.39)\sigma^{+} + 2.69(\pm 0.40) \quad (73)$$

where $n = 10$, $r^2 = 0.933$, $s = 0.176$, $q^2 = 0.807$, and $F_{2,7} = 48.739$

ID_{50} is the molar concentration of 4-X-phenol causing 50% reduction in the fluorescence as compared to the controls. The coefficient with the σ^+ (-1.52) is a measure of the susceptibility of cytotoxicity to electronic effects. This may account for a close interaction between the radical intermediate and the electron transport chain. The high mutual correlation ($r = 0.977$) between σ^+ and $\delta_{C(C-OH)}$ for this data set suggests the possibility that the calculated ^{13}C NMR chemical shift of the phenolic (C1) carbon ($\delta_{C(C-OH)}$) can be used as an alternative descriptor for the Hammett electronic parameter of the X substituent (σ^+). We used the data of Selassie et al.[24] (Table 28) and developed the following QSAR 74:

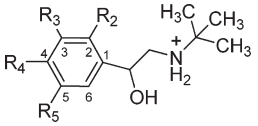
$$\log 1/ID_{50} = 0.24(\pm 0.08)C\log P - 0.18(\pm 0.04)\delta_{C(C-OH)} + 30.66(\pm 6.13) \quad (74)$$

where $n = 10$, $r^2 = 0.947$, $s = 0.157$, $q^2 = 0.895$, and $F_{2,7} = 62.538$

The value of $\delta_{C(C-OH)}$ decreases (toward upfield shift) with increasing electron releasing capacity of the X substituent, and thus, the cytotoxic activity of 4-X-phenols to CCRF-CEM cells increases. The increased hydrophobicity of the compounds further enhances the activity, as evident by the positive coefficient of the hydrophobic parameter.

4.2.20. Toxin Equivalence Factors. Buzatu and co-workers[170] developed two QSDAR models that correlate the simulated ^{13}C NMR data of 29 doxin (or doxin-like) molecules with their toxic equivalence factors (TEFs). The first MLR CoSA model based on seven spectral bins gave an r^2 of 0.88, a q^2 of 0.78, a standard deviation (s) of 0.26, and the F-value of 20.20. The second MLR CoSA model based on 10 spectral bins presented an r^2 of 0.95, a q^2 of 0.88, an s value of 0.25, and the F-value of 36.80. A third, artificial neural network model was also constructed from all the 55 multiply populated ^{13}C NMR bins by using a feed forward, back-propagating, and three-layer neural network. The ANN model gave an r^2 of 0.99, a q^2 of 0.82, and a q_3^2 of 0.81. On the basis of these results, the authors suggested

Table 27. Structure of *N-tert*-Butylphenylethanolamines (XXXXVI) Used in the Development of eq 71



XXXXVI

no.	R_2	R_3	R_4	R_5
1	H	OH	OH	H
2	H	OH	H	OH
3	H	H	OH	H
4	H	NH_2	OH	H
5	H	$NHCH_3$	OH	H
6	H	CH_2OH	OH	H
7	H	$CONH_2$	OH	H
8	H	Cl	NH_2	Cl
9	Cl	H	H	H
10	H	Cl	OH	Cl
11	H	H	H	H

that the QSDAR models are better than the traditional QSAR models because the ^{13}C NMR spectral data used in QSDAR model is more informative for the biochemical properties of each molecule than are the calculated electrostatic potentials and the molecular alignment used in the development of classical QSAR models.[170]

Another attempt of Wilkes and co-workers[171] was to improve the above QSDAR models for TEFs of doxin (or doxin-like) molecules in order to predict TEFs of two molecules. An improved QSDAR model using ^{13}C NMR chemical shifts was developed that predicted TEFs of 0.037 and 0.004, respectively, for 1,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) and 1,2,3,4,7-pentachlorodibenzo-*p*-dioxin (PeCDD), among the 390 congeners for which TEFs are assumed to be zero. The second QSDAR model of relative potency (REP) predicted TEFs of 0.115 and 0.020, respectively, for TCDD and PeCDD. The prediction of TEFs for these two molecules was then confirmed by a luciferase gene expression assay based on mouse liver cells that gave the corresponding TEFs of 0.027 and 0.013. On the basis of this close agreement for the TEFs of TCDD and PeCDD obtained from both the QSDAR models and the gene-expression assay, it was suggested that the QSDAR prediction followed by an *in vitro* assay could be very useful to select a few potential candidates among hundreds for further investigation.[171]

5. AN OVERVIEW

In the field of QSAR/QSPR, the ^{13}C NMR chemical shifts can mainly be regarded as useful electronic descriptors for both the aromatic and aliphatic systems. This could be demonstrated by means of some excellent correlations with commonly used Hammett electronic parameters (σ, σ^+, and σ^-) and its extended form (σ_I, σ_R, σ_F, and σ^*), as shown in Tables 1−3 and eqs 6−11. This could further be supported by excellent correlations with the F and R values of Swain and Lupton (see eqs 14 and 15). The advantage of the ^{13}C NMR chemical shifts is their

Table 28. Calculated ^{13}C NMR Chemical Shift ($\delta_{C(C-OH)}$) and Biological (log 1/ID$_{50}$; mol L^{-1}) and Physicochemical Parameters of 4-X-Phenols Used To Derive eqs 74

		log 1/ID$_{50}$ (eq 74)				
no.	X	obsd[a]	pred[b]	Δ	(C log P)[b]	($\delta_{C(C-OH)}$)[c]
1	NH_2	4.61	4.57	0.04	0.25	148.50
2	OC_6H_{13}	5.34	5.23	0.11	4.22	150.10
3	CH_3	3.82	3.75	0.07	1.97	155.50
4	$C(CH_3)_3$	3.86	4.08	−0.22	3.30	155.40
5	H	3.27	3.10	0.17	1.48	158.50
6	OCH_3	4.41	4.48	−0.07	1.57	150.80
7	C_3H_7	3.72	3.95	−0.23	3.03	155.80
8	$C(CH_3)_2C_6H_4$-4′-OH	4.18	4.06	0.12	3.67	156.00
9	C_2H_5	3.79	3.84	−0.05	2.50	155.70
10	C_8H_{17}	4.66	4.59	0.07	5.68	155.70

[a] Data from ref 24. [b] Calculated from the C-QSAR program (ref 17). [c] Calculated from the ChemBioDraw Ultra 12 program (ref 20).

ability to explain the electronic substituent effects on all the carbon atoms of the aromatic/aliphatic systems. Strong correlations in Table 6 support the idea that the substituent effects both on the atomic charges and on the chemical shifts reflect electronic substituent effects. High correlations with electronic parameters such as BDE, E_{HOMO}, E_{LUMO}, and $E_{HOMO} - E_{LUMO}$ can be seen in eqs 16−19. Other electronic parameters, such as electronegativity, ionization potential, pK_a, polarizability, and NVE also gave the best correlations with the ^{13}C NMR chemical shifts (see eqs 21−23, 26, 29, and 30).

The hydrophobic parameter (log P) is one of the most frequently used molecular descriptors in QSAR modeling that plays an important role in the drug design and development. This is mainly because of its usefulness to predict the pharmacokinetics of the drug molecules. The idea that the difference in chemical shifts observed in an aqueous and organic solvent may correlate to log P in the QSPR model (eq 31) with high correlation was developed for the computation of log P.

The steric effects are of overwhelming importance in ligand−receptor interactions as well as in transport phenomena in cellular systems. Thus, the importance of steric descriptors is obvious in the QSAR analysis. The ^{13}C NMR chemical shifts can be regarded as steric descriptors but cannot be preferred against the traditional steric parameters. A correlation between the ^{13}C NMR chemical shifts and calculated molar refractivity (CMR) can be seen in the QSPR model (eq 35).

The other important class of frequently used descriptors in QSAR modeling is the topological descriptors. They are truly structural descriptors, since they are based only on the two-dimensional representation of chemical structure. The ^{13}C NMR chemical shifts can be used in place of topological descriptors and may provide an easy explanation as compared to the original descriptors. This could be demonstrated by good correlations with the most widely used topological descriptors, such as Randic connectivity ($^0\chi$, $^1\chi$, $^2\chi$) and Kier and Hall valence connectivity ($^0\chi^v$, $^1\chi^v$, $^2\chi^v$) indices (see eq 36 for $^0\chi$, $^1\chi$, $^2\chi$ and statistics in the text for $^0\chi^v$, $^1\chi^v$, $^2\chi^v$).

Excellent to good correlations of ^{13}C NMR chemical shifts with a large set of routinely utilized molecular descriptors established the idea that the ^{13}C NMR chemical shifts not only can be regarded as the electronic descriptors but may also be the steric and topological descriptors in QSAR/QSPR modeling. The main concern is about its use in a different form; that is, we can say that it is not a single parameter but the collection of parameters. Thus, caution must be exercised in the use of an appropriate form of the ^{13}C NMR chemical shifts as QSAR/QSPR descriptor for the best result.

An analysis of a total number of 74 QSPR (21 from eqs 6−36 and 46 and the correlations in Tables 1−3 and 6) reveals the most important form of the ^{13}C NMR chemical shifts used as QSPR descriptor is the chemical shift of a C-atom common to all compounds. Out of 74 QSPR, 67 (eqs 6−11, 14−16, 18, 19, 22, 26, and 36 and the correlations in Tables 1−3 and 6) contain a correlation between physical characteristics and the ^{13}C NMR chemical shift of a C-atom common to all compounds. The second important form of the ^{13}C NMR chemical shifts is the averaged chemical shift of the whole molecule, which is used in 4 QSPR (eqs 23, 29, 30, and 35). QSPR 21 is an interesting example where two descriptors of the ^{13}C NMR chemical shift of two different C-atoms common to all compounds are used. The other forms of ^{13}C NMR chemical shifts used as QSPR descriptors are as follows: the difference of chemical shifts between two carbon atoms (eq 17) and the sum of carbon chemical shift differences in two solvents (eq 31).

The most important form of the ^{13}C NMR chemical shifts used as descriptor in the physical (or nonbiological) QSAR is the chemical shift of a C-atom common to all compounds (eqs 37, 38, 39, 41, and 42). Equations 37 and 38 suggest that the increased degradation rate constants of either benzoic acids or its corresponding glucuronides are directly associated with their ^{13}C NMR chemical shifts δ_C(C═O) and fall relatively in the higher field region. The rate coefficient for the neutral hydrolysis of X-phenyl dichloroacetates (**XXIII**) increases with upfield shift of their carbonyl carbons, as evident by the negative coefficient of δ_C(C═O) in eq 39. Equations 41 and 42 are for the alkaline hydrolysis of cephalosporins (**XXIV**), where the rate constants are well correlated with the ^{13}C NMR chemical shifts $\delta_{C\text{-}3}$ and δ_{COO}, respectively. The negative coefficients associated with $\delta_{C\text{-}3}$ and δ_{COO} suggest that the rate of alkaline hydrolysis of cephalosporins (**XXIV**) increases with the higher field chemical shifts of $\delta_{C\text{-}3}$ and δ_{COO}. The other forms of ^{13}C NMR chemical shifts used as QSAR descriptors are the difference of chemical shifts between two carbon atoms (eqs 43−45) and the difference of chemical shift of a C-atom common to all compounds from that of the parent (eqs 47 and 49). In eq 47, the rate constant for the oxidation of 4-X-benzaldehydes to 4-X-benzoic acids is well correlated with $\Delta\delta_{C1}$. The negative coefficient of $\Delta\delta_{C1}$ suggests that the lower chemical shifts of the C-1 carbon of 4-X-benzaldehydes in comparison to that of unsubstituted benzaldehyde are associated with the increased oxidation rate constants of 4-X-benzaldehydes. On the other hand, it is not true for the reduction of 4-X-benzenediazonium ions (**XXIX**) by citratocopper(I) complex (Sandmeyer hydroxylation), as shown in eq 49, where the higher chemical shift of the C-1 carbon of 4-X-benzenediazonium ions in comparison to that of unsubstituted benzenediazonium ion is associated with the increased relative rate constants for the reduction. But it is not surprising because oxidation and reduction are opposite reactions to each other.

The application of the ^{13}C NMR chemical shifts as descriptors for biological QSAR modeling is utmost important. An analysis of 20 biological 2D-QSAR (classical QSAR/QSDAR) models reveals one of the most important forms of the ^{13}C NMR

Table 29. Comparison between Physical (or Nonbiological) QSAR Models Developed by Using Traditional and/or ^{13}C NMR Chemical Shift Descriptors

		QSAR models using traditional descriptor(s)						QSAR models using ^{13}C NMR chemical shifts as descriptor(s)					
compd	end point	QSAR no.	descriptor(s)	n	r^2	q^2	s	QSAR no.	descriptor(s)	n	r^2	q^2	s
XXII	degradation (log k_d)	iv[b]	H_{PAC} + Es	16	0.909	0.863	0.155	ii[a,b]	$H_{PAC} + \delta_C(C{=}O)_{acid}$	18	0.921	0.890	0.150
		iii[b]	pK_a + Es	16	0.907	0.776	0.157	v[a,b]	p$K_a + \delta_C(C{=}O)_{acid}$	18	0.843	0.773	0.211
XXIII	hydrolysis (log k_o)	40	σ	9	0.973	0.944	0.113	39	$\delta_C(C{=}O)$	9	0.940	0.887	0.168
XXVI	hydrolysis (log k_{obsd}^{NMR})	46	σ_I	9	0.896	0.840	0.131	45	$\Delta\delta_{(4-3)}$	9	0.876	0.742	0.142
XXVII	oxidation (log k)	48	σ^+	12	0.984	0.976	0.090	47	$\Delta\delta_{C1}$	12	0.919	0.889	0.202
XXIX	reduction (log k_{4X}/k_H)	50	σ^+	11	0.962	0.943	0.129	49	$\Delta\delta_{C1}$	11	0.857	0.785	0.251

[a] QSAR contains traditional as well as ^{13}C NMR chemical shift descriptors. [b] See QSARs ii-v in Table 14.

chemical shifts used as QSAR descriptor is the chemical shift of a C-atom common to all compounds. Out of 20 QSAR, 17 (eqs 54, 56−67, 69, 70, 72, and 74) contain a correlation between the biological activities and the ^{13}C NMR chemical shift of a C-atom common to all compounds along with some physicochemical parameters in few cases. QSAR 54 is a parabolic correlation in terms of *C* log *P* with an optimum *C* log *P* of ∼7.80 followed by an electronic parameter as represented by the ^{13}C NMR chemical shift of C1 carbon (δ_{C1}) of cinnamic acid derivatives (**XXXI**) for their antimalarial activities against intraerythrocytic forms of *P. falciparum* strain Dd2. Equations 56 and 57 represent the excellent correlations between antiproliferative activities of quinolines (**XXXIV**) and their ^{13}C NMR chemical shifts of C-3 ($\delta_{C\text{-}3}$) and C-4 ($\delta_{C\text{-}4}$) carbons, respectively. On combining the results of both QSARs 56 and 57, one can say that the upfield and downfield chemical shifts, respectively, of C-4 and C-3 carbons will increase the antiproliferative activity of quinolines (**XXXIV**) against T47D cells. QSARs 58−61 are for the dopamine D_2 receptor affinity of benzamides (**XXXV**−**XXXVII**), for which eq 59 is an interesting parabolic correlation for benzamides (**XXXVI**) in terms of π_X (optimum π_X = 1.39) followed by ^{13}C NMR chemical shifts of C-6. QSAR 62 describes the correlation between the antituberculotic activities of polyhydroxyxanthones (**XXXVIII**) and their ^{13}C NMR chemical shifts of C-4b and C-7 carbons, which suggests an increase in activity of polyhydroxyxanthones (**XXXVIII**) with both downfield shifted C-atoms. QSAR 63 is the modified form of QSAR 62 for which the $\delta_{C\text{-}4b}$ is replaced by the hydrophobic parameter (Clog *P*). This QSAR suggests that the activity of compounds (**XXXVIII**) increases with increasing their hydrophobicity and downfield shifted C-7 atom. The identical QSARs 64 and 65 for the binding affinities of benzphetamine derivatives to the soluble cytochrome P450 LM2 suggest that the binding affinity increases by downfield chemical shifts of the *N*-methyl and the quaternary phenyl carbons. The chemical shift of α-carbon in 5-residue (δC ppm from glycine-C_α) was found to be the single most important parameters that accurately predict the biological activity of angiotensin II analogues Asp-Arg-Val-Tyr-X-His-Pro-Phe as represented by eq 66. It is apparent from eq 67 that the *ex vivo* inhibitory activity of compounds (**XXXX**) to ChoK increases with decreasing δ_{CH2N+} values. Similar results were obtained for the antiproliferative activity of these compounds (**XXXX**) against the HT-29 cells in addition to an expected hydrophobic term as evident by eq 69. QSAR model (eq 70) suggests that the inhibitor potency of compounds (**XXXXI**−**XXXXV**) to 5-LO increases with an increased compound hydrophobicity and an increased N-phenyl electronic effect (CNMR1′). The QSAR 72 suggests that the electronic and hydrophobic components are playing a major role in promoting the auxin-like activity of benzotriazoles (**XXXXVII**). Similar observation obtained from QSAR 74 for the cytotoxic activity of 4-X-phenols to CCRF-CEM cells. The second important form of ^{13}C NMR chemical shifts used as QSAR descriptors is the difference of chemical shift of a C-atom common to all compounds from that of the parent (eqs 55 and 71). From these two QSARs, eq 55 is an interesting example for which M_3 antagonistic activities of esters **XXXII** and **XXXIII** (including atropine) were correlated with their two electronic parameters (Taft's σ^* and $\Delta\delta$) of the side chain. In addition to the 2D QSAR, a CoMFA model was also performed for the comparison. CoMFA model suggests the contributions of both the steric and electrostatic effects in contrast to the only electrostatic parameters (σ^* and $\Delta\delta$) as suggested by the 2D QSAR. The other form of ^{13}C NMR chemical shifts used as QSAR descriptor is the difference of chemical shifts between two carbon atoms (eq 52). For which the antibacterial activity of β-nitrostyrene derivatives (**XXX**) against the gram-positive bacteria (*Enterococcus faecalis* ATCC 29212) is well correlated with the ^{13}C NMR chemical shift differences between C-β and C-α carbons ($\Delta\delta_{(\beta-\alpha)}$) and increases with increasing $\Delta\delta_{(\beta-\alpha)}$ values.

In addition to the classical QSDAR models (classical QSAR using ^{13}C NMR as descriptors), the following other QSDAR models were also discussed: SDAR (see sections 4.2.8 and 4.2.9), CoSA (see sections 4.2.7−4.2.10, 4.2.12, and 4.2.20), CoSASA (see sections 4.2.7 and 4.2.10), and CoSCoSA (see sections 4.2.7− 4.2.10). In these QSDAR models, the ^{13}C NMR spectra were broken into bins and principal components (PCs). The bin intensities or PCs were then used to produce the QSDAR models. The spectral bins are the number of chemical shift peaks with a specific ppm range, while the PCs are the contributions from each of the 1.0 ppm spectral bins.

Since our primary goal is to study the potentials of the proposed use of ^{13}C NMR chemical shifts as descriptors in the classical QSAR/QSPR analyses, a comparative QSAR/QSPR study between the models obtained from the traditional descriptor(s) and the ^{13}C NMR chemical shift descriptor(s) was carried out. The comparative QSPR study can be seen in the correlations of ^{13}C NMR chemical shift descriptors with different electronic, steric, hydrophobic, and topological descriptors (see 22 QSPR from eqs 6−36 and 53 correlations in Tables 1−3 and 6). A comparative QSAR study for the physical QSARs was done by comparing their QSAR models developed by using traditional and/or ^{13}C NMR chemical shift descriptors (see Table 29). Similarly, the comparative QSAR study for the biological QSARs

Table 30. Comparison between Biological QSAR Models Developed by Using Traditional and/or ^{13}C NMR Chemical Shift Descriptors

		QSAR models using traditional descriptor(s)						QSAR models using ^{13}C NMR chemical shifts as descriptor(s)					
compd	end points	QSAR No.	descriptor(s)	n	r^2	q^2	s	QSAR No.	descriptor(s)	n	r^2	q^2	s
XXX	antibacterial activity (log 1/MIC)	51	E_p	12	0.802	0.710	0.238	52	$\Delta\delta_{(\beta-\alpha)}$	10	0.826	0.709	0.235
XXXI	antimalarial activity (log 1/IC_{50})	53	$C\log P + \sigma^+$	14	0.909	0.826	0.187	54[a]	$C\log P + \delta_{C1}$	15	0.861	0.710	0.259
XXXVIII	antituberculous activity (log 1/MIC)	63[a]	$C\log P + \delta_{C\text{-}7}$	10	0.735	0.502	0.157	62	$\delta_{C\text{-}4b} + \delta_{C\text{-}7}$	9	0.916	0.821	0.094
4-X-phenols	cytotoxicity to CCRF-CEM cells (log 1/ID_{50})	73	$C\log P + \sigma^+$	10	0.933	0.807	0.176	74[a]	$C\log P + \delta_{C(C-OH)}$	10	0.947	0.895	0.157

[a] QSAR contains hydrophobic (Clog P) as well as ^{13}C NMR chemical shift descriptors

is summarized in Table 30. From the above comparative QSAR/QSPR studies, it can be concluded that the ^{13}C NMR chemical shifts have the potential to be used as electronic parameters in classical QSAR/QSPR studies. They can also be used as steric and topological descriptors in some specific cases, but they cannot be used as the hydrophobic descriptor because we did not obtain any direct high mutual correlation between the ^{13}C NMR chemical shifts and the hydrophobic parameter of the compounds. It is further supported by several biological QSAR models (eqs 54, 59, 61, 63, 69, 70, 72, and 74) where the hydrophobic parameter was used along with the ^{13}C NMR chemical shifts descriptor. Of course, the ^{13}C NMR chemical shifts descriptor can be used along with four different physicochemical parameters to calculate the hydrophobic parameter (see eq 31).

A comparative QSAR/QSDAR study using various methods was also demonstrated that includes the classical QSAR, CoMFA, E-state, HE-state/E-state, and SOM models as well as different types of QSDAR models. A comparative QSDAR/QSAR study among five QSDAR (2 CoSA, 1 ANN-CoSA, and 2 CoSASA), one 2D-QSAR (based on five structural parameters), and one CoMFA models for the binding activities of 50 steroids to aromatase enzyme was carried out (see statistical data in Table 24). All five QSDAR models showed better statistics than that of the classical QSAR model of five structural parameters. However, the CoMFA model has the better statistics compared to those of the CoSA (except ANN-CoSA) and CoSASA models. The ANN-CoSA model reproduces an r^2 value of 1.00, but it is not surprising because all the 87 bins of the ^{13}C NMR data were used in the development of that model. Another comparative QSDAR/QSAR study among 10 QSDAR (2 CoSA, 2 CoSASA, and 6 CoSCoSA) and 4 QSAR (one each from CoMFA, HE-state/E-state, E-state, and SOM) models for the binding activities of 30 (or 31) steroids to corticosteroid binding globulin was carried out (see statistical data in Table 26). The best CoSCoSA (COSY + distance) model based on 8 PCs gave an r^2 of 0.96 and q^2 of 0.92, which are better than that of the HE-state/E-state QSAR model. A comparative QSDAR study was also performed by using three different QSDAR methods (e.g., EVA, EEVA, and CoSA) for a diverse set of 36 estrogens with respect to their RBA to ER (see statistical data in Table 25). The CoSA model with the ^{13}C NMR chemical shifts gave the best statistics.

6. CONCLUSION

The ^{13}C NMR chemical shifts can be regarded as the electronic descriptors in classical QSPR/QSAR modeling and demonstrated by various excellent correlations (QSPR) between the ^{13}C NMR chemical shifts of compounds and a large set of routinely utilized electronic descriptors (σ, σ^+, σ^-, σ_I, σ_R, σ_F, σ^*, F, R, q, BDE, E_{HOMO}, E_{LUMO}, $E_{HOMO} - E_{LUMO}$, χ, IP, pK_a, α, and NVE) as well as the comparative QSAR studies (see Tables 29 and 30). Although the ^{13}C NMR chemical shifts cannot be considered as a good hydrophobic descriptor, it can be used in the computation of the hydrophobic parameter (see eq 31). Of course, the ^{13}C NMR chemical shifts are not preferred steric and/or topological descriptors; it can be used in some cases, as evident by the strong correlations (QSPR) between the ^{13}C NMR chemical shifts of compounds and their steric/topological descriptors (CMR, $^0\chi$, $^1\chi$, $^2\chi$, $^0\chi^v$, $^1\chi^v$, and $^2\chi^v$) as well as the comparative QSAR studies (see Table 29).

The above results indicate that the ^{13}C NMR chemical shifts are sufficiently rich in chemical information and able to encode the structural features of the molecules contributing significantly to their biological activity, chemical reactivity, or physical characteristics, thus highlighting the promising potential of the ^{13}C NMR chemical shifts as descriptor in classical QSAR/QSPR modeling studies. The main concern is about their use in different formats. This is the reason; we assigned it as the collection of parameters and not the single descriptor. Thus, caution must be exercised in QSAR/QSPR modeling for the selection of ^{13}C NMR chemical shifts as model descriptor in an appropriate format for optimum results. The QSAR/QSPR models using ^{13}C NMR chemical shifts as input parameter is usually referred to as QSDAR.

The present review covers different types of QSDAR models, with special emphasis on the classical QSDAR (classical QSAR/QSPR models using ^{13}C NMR chemical shifts as descriptor). We discussed the 2D-QSAR results under two sections: Physical (or Nonbiological) QSAR and Biological QSAR. In the Biological QSAR section, the other QSDAR (SDAR, CoSA, CoSASA, and CoSCoSA) models are also accommodated. An analysis of the physical and biological 2D-QSAR (classical QSAR/QSDAR) models reveals one of the most important formats of the ^{13}C NMR chemical shifts used as QSAR descriptor is the chemical shift of a C-atom common to all compounds. The other QSDAR (SDAR, CoSA, CoSASA, and CoSCoSA) models utilize either the spectral bins intensities or the principal components of the ^{13}C NMR data in their development. Comparative QSAR study among the QSDAR and CoMFA models suggests that CoSCoSA (3D-QSDAR) models are always competitive to the CoMFA. It suggests the idea that the CoMFA model may be replaced by the CoSCoSA model. The advantage of the CoSCoSA model is its easy development, since this technique does not need to use molecular docking. On the basis of the close agreement between the biological activity, chemical reactivity, or physical

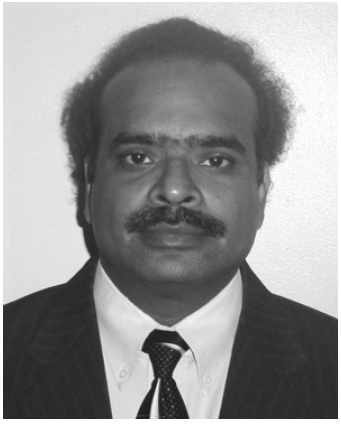
properties of compounds and their ^{13}C NMR chemical shifts as represented by different kinds of QSDAR models, the ^{13}C NMR chemical shifts can be considered as significant descriptors in QSAR/QSPR (QSDAR) modeling in order to establish the chemical−biological interactions, which may provide strategies that might aid in the drug design and development processes.

AUTHOR INFORMATION

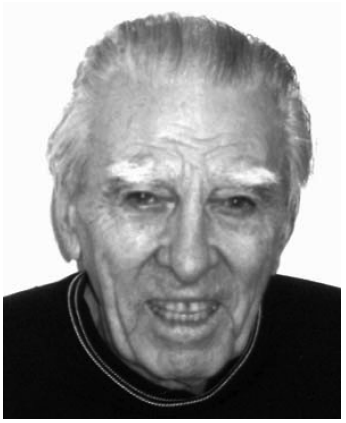
Corresponding Author

*Phone: (909)607-4249. Fax: (909)607-7726. E-mail: rverma@pomona.edu.

BIOGRAPHIES



Rajeshwar P. Verma received his M.Sc. (1988) and Ph.D. (1992) degrees in chemistry from Magadh University, Bodh-Gaya (India). After working one year at the same university as a postdoctoral fellow with Professor K. S. Sinha, he joined Indian Institute of Technology Roorkee as a research associate and worked with Professor S. M. Sondhi (1993−1997). He also worked as a Lecturer of Chemistry at Gurukula Kangri University, Hardwar (1994−1995). He won a Research Associateship Award in December 1994 from the Council of Scientific & Industrial Research, New Delhi (India). In 1997, he moved to Pomona College to join the renowned QSAR research group of Professor Corwin Hansch and Cynthia Selassie, working as a postdoctoral research associate. At present, he is also working as a Lecturer in Chemistry at Pomona College and a senior scientist (consultant) at BioByte Corp., Claremont. Dr. Verma has 66 publications in high impact journals, such as *Chemical Reviews*, *Journal of Medicinal Chemistry*, *Chemical Research in Toxicology*, *Molecular Pharmaceutics*, *Synthesis*, *ChemMedChem*, *Journal of Computer-Aided Molecular Design*, *Journal of Pharmaceutical Sciences*, etc., as well as six chapters in treatises such as *Combating the Threat of Pandemic Influenza: Drug Discovery Approaches*, the seventh edition of *Berger's Medicinal Chemistry: Drug Discovery and Development*, etc. He has also made numerous presentations at regional and national scientific meetings. Dr. Verma's current research interests include the following: chemistry of isothiocyanates, synthesis of biologically important heterocyclic and phenolic compounds, quantitative structure−activity relationships (QSAR) and its application to computer-assisted molecular design (CAMD), multidrug resistance, and radical-based toxicity of phenolic compounds.



Corwin Hansch received his undergraduate education at the University of Illinois and his Ph.D. degree in organic chemistry from New York University in 1944. After working with the Du Pont Co., first on the Manhattan Project and then in Wilmington, DE, he joined the Pomona College faculty in 1946. He has remained at Pomona except for two sabbaticals: one at the Federal Institute of Technology in Zurich, Switzerland, with Professor Prelog, and the other at the University of Munich with Professor Huisgen. The Pomona group published the first paper on the QSAR approach, relating chemical structure with biological activity in 1962. Since then, QSAR has received widespread attention. He is an honorary fellow of the Royal Society of Chemistry and received the ACS Award for Computers in Chemical and Pharmaceutical Research for 1999.

ABBREVIATIONS

ADME	absorption, distribution, metabolism, and excretion
3D-QSAR	three dimensional QSAR
3D-QSDAR	three dimensional QSDAR
ADRs	adverse drug reactions
AhR	aryl hydrocarbon receptor
ANN	artificial neural networks
BDE	bond dissociation energy
ChoK	choline kinase
CoMFA	comparative molecular field analysis
CoSA	comparative spectral analysis
CoSASA	comparative structurally assigned spectral analysis
CoSCoSA	comparative structural connectivity spectra analysis
DSP	dual-substituent parameter
EEVA	electronic eigenvalue analysis
E_{HOMO}	energy of the highest occupied molecular orbital
EI MS	electron ionization mass spectra
E_{LUMO}	energy of the lowest unoccupied molecular orbital
ER	estrogen receptor
EVA	eigenvalue analysis
GERD	gastro esophageal reflux disease
HC-3	Hemicholinium-3
IP	ionization potential
ISA	intrinsic sympathomimetic activity
MIC	minimum inhibitory concentration
NHA	number of hydrogen bond acceptors
NHD	number of hydrogen bond donors
NMR	nuclear magnetic resonance

NSAIDs	nonsteroidal anti-inflammatory drugs
NVE	number of valence electrons
PAHs	polycyclic aromatic hydrocarbons
PCBs	polychlorinated biphenyls
PCDDs	polychlorinated dibenzodioxins
PCDFs	polychlorinated dibenzofurans
PCs	principal components
QSAR	quantitative structure−activity relationship
QSDAR	quantitative spectrometric data−activity relationship
QSPR	quantitative structure−property relationship
RBA	relative binding affinities
SDAR	spectroscopic data−activity relationship

REFERENCES

(1) Hansch, C.; Maloney, P. P.; Fujita, T.; Muir, R. M. *Nature* **1962**, *194*, 178.
(2) Verma, R. P.; Hansch, C. *Chem. Rev.* **2009**, *109*, 213.
(3) Gedeck, P.; Lewis, R. A. *Curr. Opin. Drug Discovery Dev.* **2008**, *11*, 569.
(4) Witkowski, S.; Maciejewska, D.; Wawer, I. *J. Chem. Soc., Perkin Trans. 2* **2000**, *7*, 1471.
(5) Neuvonen, H.; Neuvonen, K. *J. Chem. Soc., Perkin Trans. 2* **1999**, *7*, 1497.
(6) Staunton, D.; Owen, J.; Campbell, I. D. *Acc. Chem. Res.* **2003**, *36*, 207.
(7) Wüthrich, K. *Nat. Struct. Biol.* **2001**, *8*, 923.
(8) Meiler, J.; Will, M. *J. Am. Chem. Soc.* **2002**, *124*, 1868.
(9) Emsley, J. W.; Feeney, J.; Sutcliffe, L. H. *High Resolution Nuclear Magnetic Resonance*; Pergamon Press Ltd.: Oxford, 1965; Vol. I, pp 1−287.
(10) Wishart, D. S.; Sykes, B. D. *Methods Enzymol.* **1994**, *239*, 363.
(11) Kvasnicka, V. *J. Math. Chem.* **1991**, *6*, 63.
(12) Beger, R. D. *Drug Discovery Today* **2006**, *11*, 429.
(13) Beger, R. D.; Buzatu, D. A.; Wilkes, J. G.; Lay, J. O., Jr. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1123.
(14) Miller, D. W.; Beger, R.; Lay, J. O., Jr.; Wilkens, J. G.; Freeman, J. P. . WO 2001057495 A2 20010809, 2001.
(15) Beger, R. D.; Wilkes, J. G. . US 2003229456 A1 20031211, 2003.
(16) Miller, D. W.; Beger, R.; Lay, J. O., Jr.; Wilkens, J. G.; Freeman, J. P. . US 6898533 B1 20050524, 2005.
(17) C-QSAR Program; BioByte Corp.: 201W. 4th St., Suit 204, Claremont, CA 91711, USA. www.biobyte.com
(18) Hansch, C.; Hoekman, D.; Leo, A.; Weininger, D.; Selassie, C. D. *Chem. Rev.* **2002**, *102*, 783.
(19) Verma, R. P.; Hansch, C. *Nat. Protoc.*, doi: 10.1038/nprot.2007.125,http://www.natureprotocols.com/2007/03/05/development_of_qsar_models_usi_1.php.
(20) ChemBioDraw Ultra 12, CambridgeSoft. www.cambridgesoft.com
(21) Cramer, R. D., III; Bunce, J. D.; Patterson, D. E.; Frank, I. E. *Quant. Struct.−Act. Relat.* **1988**, *7*, 18.
(22) Bennett, C. A.; Franklin, N. L. *Statistical Analysis in Chemistry and the Chemical Industry*; John Wiley & Sons: New York, 1967; pp 708−709; 5th printing.
(23) Selassie, C. D.; Verma, R. P. *Burger's Medicinal Chemistry, Drug Discovery and Development: History of Quantitative Structure−Activity Relationships*, 7th ed.; Abraham, D. J., Rotella, D. P., Eds.; John Wiley & Sons: New York, 2010; Vol. 1, pp 1−95.
(24) Selassie, C. D.; Kapur, S.; Verma, R. P.; Rosario, M. *J. Med. Chem.* **2005**, *48*, 7234.
(25) Selassie, C. D.; Garg, R.; Kapur, S.; Kurup, A.; Verma, R. P.; Mekapati, S. B.; Hansch, C. *Chem. Rev.* **2002**, *102*, 2585.
(26) Hammett, L. P. *J. Am. Chem. Soc.* **1937**, *59*, 96.
(27) Kirby, A. J. In *Comprehensive Chemical Kinetics*; Bamford, C. H., Tipper, C. F. H., Eds.; Elsevier Publishing Co.: Amsterdam, 1980; Vol. 10, Chapter 2, p 161.
(28) Hansch, C.; Leo, A. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*; American Chemical Society: Washington, DC, 1995.
(29) Okamoto, Y.; Brown, H. C. *J. Org. Chem.* **1957**, *22*, 485.
(30) Yoshioka, M.; Hamamoto, K.; Kubota, T. *Bull. Chem. Soc. Jpn.* **1962**, *35*, 1723.
(31) Wells, P. R.; Ehrenson, S.; Taft, R. W. *Prog. Phys. Org. Chem.* **1968**, *6*, 147.
(32) Ehrenson, S.; Brownlee, R. T. C.; Taft, R. W. *Prog. Phys. Org. Chem.* **1973**, *10*, 1.
(33) Hansch, C.; Leo, A.; Taft, R. W. *Chem. Rev.* **1991**, *91*, 165.
(34) Taft, R. W. In *Steric Effects in Organic Chemistry*; Newman, M. S., Ed.; Wiley: New York, 1956; p 556.
(35) Craik, D. J.; Brownlee, R. T. C. *Prog. Phys. Org. Chem.* **1983**, *14*, 1.
(36) Neuvonen, H.; Neuvonen, K.; Koch, A.; Kleinpeter, E.; Pasanen, P. *J. Org. Chem.* **2002**, *67*, 6995.
(37) Neuvonen, H.; Neuvonen, K.; Pasanen, P. *J. Org. Chem.* **2004**, *69*, 3800.
(38) Neuvonen, H.; Neuvonen, K.; Fülöp, F. *J. Org. Chem.* **2006**, *71*, 3141.
(39) Reis, A. K. C. A.; Rittner, R. *Spectrochim. Acta, Part A* **2007**, *66*, 681.
(40) Thirunarayanan, G. *Acta Cienc. Indica* **2005**, *XXXI*, 299.
(41) Swain, C. G.; Lupton, E. C., Jr. *J. Am. Chem. Soc.* **1968**, *90*, 4328.
(42) Hansch, C.; Leo, A.; Unger, S. H.; Kim, K. H.; Nikaitani, D.; Lien, E. J. *J. Med. Chem.* **1973**, *16*, 1207.
(43) Neuvonen, K.; Fülöp, F.; Neuvonen, H.; Koch, A.; Kleinpeter, E.; Pihlaja, K. *J. Org. Chem.* **2003**, *68*, 2151.
(44) Lin, S.-T.; Lee, C.-C.; Liang, D. W. *Tetrahedron* **2000**, *56*, 9619.
(45) Hehre, W. J.; Taft, R. W.; Topsom, R. D. *Prog. Phys. Org. Chem.* **1976**, *12*, 159.
(46) Selassie, C. D.; Shusterman, A. J.; Kapur, S.; Verma, R. P.; Zhang, L.; Hansch, H. *J. Chem. Soc., Perkin Trans. 2* **1999**, *12*, 2729.
(47) Fukui, K.; Yonezawa, T.; Nagata, C.; Shingu, H. *J. Chem. Phys.* **1954**, *22*, 1433.
(48) Karelson, M.; Lobanov, V. S.; Katritzky, A. R. *Chem. Rev.* **1996**, *96*, 1027.
(49) Ebenso, E. E.; Arslan, T.; Kandemirli, F.; Caner, N.; Love, I. *Int. J. Quantum Chem.* **2010**, *110*, 1003.
(50) Pearson, R. G. *J. Org. Chem.* **1989**, *54*, 1423.
(51) Ishihara, M.; Fujisawa, S. *Dent. Mater. J.* **2009**, *28*, 113.
(52) Yoshii, E. *J. Biomed. Mater. Res.* **1997**, *37*, 517.
(53) Hatada, K.; Kitayama, T.; Nishiura, T.; Shibuya, W. *Curr. Org. Chem.* **2002**, *6*, 121.
(54) You, Z.; Brezzell, M. D.; Das, S. K.; Hooberman, B. H.; Sinsheimer, J. E. *Mutat. Res.* **1994**, *320*, 45.
(55) Zhang, L.; Gao, H.; Hansch, C.; Selassie, C. D. *J. Chem. Soc., Perkin Trans. 2* **1998**, *11*, 2553.
(56) Kimura, K.; Katsumata, S.; Achiba, Y.; Yamazaki, T.; Iwata, S. *Handbook of HeI Photoelectron Spectra of Fundamental Organic Molecules*; Halsted Press: New York, 1981.
(57) Weast, R. C., Ed. *CRC Handbook of Chemistry and Physics*, 55th ed.; Cleveland, OH, 1973.
(58) Baker, A. D.; May, D. P.; Turner, D. W. *J. Chem. Soc. B* **1968**, *1*, 22.
(59) Johnson, L. F.; Jankowski, W. J. *Carbon-13 NMR spectra. A collection of assigned, coded and indexed spectra*; Wiley: New York, 1972.
(60) Jørgensen, K. A.; Dalgaard, E. *Chem. Phys. Lett.* **1985**, *114*, 192.
(61) Sakamoto, Y.; Watanabe, S. *Bull. Chem. Soc. Jpn.* **1986**, *59*, 3033.
(62) Wang, H.; Ulander, J. *Expert Opin. Drug Metab. Toxicol.* **2006**, *2*, 139.
(63) Hollingsworth, G. A.; Seybold, P. G.; Hadad, C. M. *Int. J. Quantum Chem.* **2002**, *90*, 1396.

(64) Brown, H. C.; McDaniel, D. H.; Hafliger, O. *Determination of Organic Structures by Physical Methods*; Braude, E. A., Nachod, F. C., Eds.; Academic Press: New York, 1955; p 588.
(65) Wang, H. *J. Chem. Educ.* **2005**, *82*, 1340.
(66) Staikova, M.; Wania, F.; Donaldson, D. J. *Atmos. Environ.* **2004**, *38*, 213.
(67) Hansch, C.; Steinmetz, W. E.; Leo, A. J.; Mekapati, S. B.; Kurup, A.; Hoekman, D. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 120.
(68) Verma, R. P.; Hansch, C. *Bioorg. Med. Chem.* **2005**, *13*, 2355.
(69) Hansch, C.; Kurup, A. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1647.
(70) Verma, R. P.; Kurup, A.; Hansch, C. *Bioorg. Med. Chem.* **2005**, *13*, 237.
(71) Abraham, M. H.; Takács-Novák, K.; Mitchell, R. C. *J. Pharm. Sci.* **1997**, *86*, 310.
(72) Hansch, C.; Fujita, T. *J. Am. Chem. Soc.* **1964**, *86*, 1616.
(73) Leo, A.; Hansch, C.; Elkins, D. *Chem. Rev.* **1971**, *71*, 525.
(74) Leo, A. J. *Methods Enzymol.* **1991**, *202*, 544.
(75) Sijm, D. T. H. M.; Sinnige, T. L. *Chemosphere* **1995**, *31*, 4427.
(76) Tomlinson, E. *J. Pharm. Sci.* **1982**, *71*, 602.
(77) Smith, E. C. B.; Westall, R. G. *Biochim. Biophys. Acta* **1950**, *4*, 427.
(78) Hartmann, T.; Schmitt, J. *Drug Discovery Today: Technol.* **2004**, *1*, 431.
(79) Wong, K.-S.; Kenseth, J.; Strasburg, R. *J. Pharm. Sci.* **2004**, *93*, 916.
(80) Mannhold, R.; van de Waterbeemd, H. *J. Comput. Aided Mol. Des.* **2001**, *15*, 337.
(81) Erős, D.; Kövesdi, I.; Őrfi, L.; Takács-Novák, K.; Acsády, G.; Kéri, G. *Curr. Med. Chem.* **2002**, *9*, 1819.
(82) Schnackenberg, L. K.; Beger, R. D. *J. Chem. Inf. Model.* **2005**, *45*, 360.
(83) Lorentz, H. A. *The Theory of Electrons*, 2nd ed.; Dover: New York, 1952.
(84) Partington, J. R. *An Advanced Treatise on Physical Chemistry*; Longmans, Green & Co. Ltd: New York, 1953; Vol. IV: Physico-Chemical Properties.
(85) Wiener, H. *J. Am. Chem. Soc.* **1947**, *69*, 17.
(86) Gutman, I. *Graph Theory Notes New York* **1994**, *20*, 9.
(87) Khadikar, P. V.; Deshpandey, N. V.; Kale, P. P.; Dobrymin, A.; Gutman, I.; Domotor, G. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 547.
(88) Khadikar, P. V. *Nat. Acad. Sci. Lett.* **2000**, *23*, 113.
(89) Khadikar, P. V.; Karmarkar, S.; Agrawal, V. K. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 934.
(90) Randić, M. *J. Am. Chem. Soc.* **1975**, *97*, 6609.
(91) Balaban, A. T. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 23.
(92) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
(93) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure-Activity Analysis*; Wiley: New York, 1986.
(94) Jaiswal, M.; Khadikar, P. *Bioorg. Med. Chem.* **2004**, *12*, 1793.
(95) Hall, L. H.; Kier, L. B. *J. Mol. Graphics Modell.* **2001**, *20*, 4.
(96) Skonberg, C.; Olsen, J.; Madsen, K. G.; Hansen, S. H.; Grillo, M. P. *Expert Opin. Drug Metab. Toxicol.* **2008**, *4*, 425.
(97) Kalgutkar, A. S.; Gardner, I.; Obach, R. S.; Shaffer, C. L.; Callegari, E.; Henne, K. R.; Mutlib, A. E.; Dalvie, D. K.; Lee, J. S.; Nakai, Y.; O'Donnell, J. P.; Boer, J.; Harriman, S. P. *Curr. Drug Metab.* **2005**, *6*, 161.
(98) Yang, X.-X.; Hu, Z.-P.; Boelsterli; Zhou, S.-F. *Curr. Pharm. Anal.* **2006**, *2*, 259.
(99) Kuehl, G. E.; Lampe, J. W.; Potter, J. D.; Bigler, J. *Drug Metab. Dispos.* **2005**, *33*, 1027.
(100) Vanderhoeven, S. J.; Troke, J.; Tranter, G. E.; Wilson, I. D.; Nicholson, J. K.; Lindon, J. C. *Xenobiotica* **2004**, *34*, 889.
(101) Yoshioka, T.; Baba, A. *Chem. Res. Toxicol.* **2009**, *22*, 1559.
(102) Neuvonen, H.; Neuvonen, K. *J. Chem. Soc., Perkin Trans. 2* **1999**, *7*, 1497.
(103) Nishikawa, J.; Tori, K. *J. Med. Chem.* **1984**, *27*, 1657.
(104) Coene, B.; Schanck, A.; Dereppe, J.-M.; Meerssche, M. V. *J. Med. Chem.* **1984**, *27*, 694.
(105) Narisada, M.; Nishikawa, J.; Watanabe, F.; Terui, Y. *J. Med. Chem.* **1987**, *30*, 514.
(106) Bansal, V.; Sharma, P. K.; Banerji, K. K. *J. Chem. Res. (S)* **1999**, *8*, 480.
(107) Hanson, P.; Jones, J. R.; Taylor, A. B.; Walton, P. H.; Timms, A. W. *J. Chem. Soc., Perkin Trans 2* **2002**, *6*, 1135.
(108) Tomasz, A. In *Handbook of Experimental Pharmacology: Antibiotics Containing the Beta-lactam Structure*; Demain, A. L., Solomon, N. A., Ed.; Springer-Verlag: West Berlin, 1983; Part I, Chapter 2, p 67.
(109) Kelly, J. A.; Moews, P. C.; Knox, J. R.; Frére, J.-M.; Ghuysen, J.-M. *Science (Washington, D. C.)* **1982**, *218*, 479.
(110) Tomasz, A. *Annu. Rev. Microbiol.* **1979**, *33*, 113.
(111) Milhazes, N.; Calheiros, R.; Marques, M. P. M.; Garrido, J.; Cordeiro, M. N. D. S.; Rodrigues, C.; Quinteira, S.; Novais, C.; Peixe, L.; Borges, F. *Bioorg. Med. Chem.* **2006**, *14*, 4078.
(112) Calheiros, R.; Milhazes, N.; Borges, F.; Marques, M. P. M. *J. Mol. Struct.* **2004**, *692*, 91.
(113) Milhazes, N.; Borges, F.; Calheiros, R.; Marques, M. P. M. *Analyst* **2004**, *129*, 1106.
(114) World Health Organization. World malaria report. 2009 (http://www.who.int/malaria/publications/atoz/9789241563901/en/index.html).
(115) Gamo, F.-J.; Sanz, L. M.; Vidal, J.; de Cozar, C.; Alvarez, E.; Lavandera, J.-L.; Vanderwall, D. E.; Green, D. V. S.; Kumar, V.; Hasan, S.; Brown, J. R.; Peishoff, C. E.; Cardon, L. R.; Garcia-Bustos, J. F. *Nature* **2010**, *465*, 305.
(116) Ashley, E.; McGready, R.; Proux, S.; Nosten, F. *Travel Med. Infect. Dis.* **2006**, *4*, 159.
(117) White, N. J. *J. Clin. Invest.* **2004**, *113*, 1084.
(118) Kumar, A.; Katiyar, S. B.; Agarwal, A.; Chauhan, P. M. S. *Curr. Med. Chem.* **2003**, *10*, 1137.
(119) Wiesner, J.; Mitsch, A.; Wissner, P.; Jomaa, H.; Schlitzer, M. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 423.
(120) Mason, R. P.; Holtzman, J. L. *Biochemistry* **1975**, *14*, 1626.
(121) Xu, R.; Sim, M.-K.; Go, M.-L. *Chem. Pharm. Bull.* **1998**, *46*, 231.
(122) Wernly-Chung, G. N.; Mayer, J. M.; Tsantili-Kakoulidou, A.; Testa, B. *Int. J. Pharm.* **1990**, *63*, 129.
(123) Boryczka, S.; Wietrzyk, J.; Nasulewicz, A.; Pełczyńska, M.; Opolski, A. *Pharmazie* **2002**, *57*, 733.
(124) Zuika, I. V.; Popelis, Y. Y.; Sekacis, I. P.; Bruvers, Z. P.; Tsirule, M. A. *Khim. Geterotsikl. Soedin.* **1979**, *12*, 1665.
(125) De Paulis, T.; El Tayar, N.; Carrupt, P.-A.; Testa, B.; Van de Waterbeemd, H. *Helv. Chim. Acta* **1991**, *74*, 241.
(126) El Tayar, N.; Van de Waterbeemd, H.; De Paulis, T.; Carrupt, P. A.; Testa, B. In *QSAR: Rational Approaches to the Design of Bioactive Compounds: ^{13}C Chemical Shifts and QSAR of Substituted Benzamides*; Silipo, C., Vittoria, A., Ed.; Elsevier Science Publishers B. V.: Amsterdam, 1991; pp 163−166.
(127) Ghosal, S.; Biswas, K.; Chaudhuri, R. K. *J. Pharm. Sci.* **1978**, *67*, 721.
(128) Hambloch, H.; Frahm, A. W.; Wiedemann, B. *Eur. J. Med. Chem.—Chem. Ther.* **1985**, *20*, 71.
(129) Schaper, K.-J.; Pickert, M.; Frahm, A. W. *Arch. Pharm. Pharm. Med. Chem.* **1999**, *332*, 91.
(130) Wiese, M.; Schaper, K.-J. *SAR QSAR Environ. Res.* **1993**, *1*, 137.
(131) Chen, S.; Besman, M. J.; Shively, J. E.; Yanagibashi, K.; Hall, P. F. *Drug Metab. Rev.* **1989**, *20*, 511.
(132) Brodie, A. M. *J. Steroid Biochem. Mol. Biol.* **1994**, *49*, 281.
(133) Beger, R. D.; Buzatu, D. A.; Wilkes, J. G.; Lay, J. O., Jr. *J. Chem. Inf. Comp. Sci.* **2001**, *41*, 1360.
(134) Beger, R. D.; Wilkes, J. G. *J. Mol. Recognit.* **2002**, *15*, 154.
(135) Beger, R. D.; Wilkes, J. G. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1322.
(136) Beger, R. D.; Buzatu, D. A.; Wilkes, J. G. *J. Comput. Aided Mol. Des.* **2002**, *16*, 727.

(137) Shade, L.; Beger, R. D.; Wilkes, J. G. *Environ. Toxicol. Chem.* **2003**, *22*, 501.
(138) Beger, R. D.; Freeman, J. P.; Lay, J. O., Jr.; Wilkes, J. G.; Miller, D. W. *Toxicol. Appl. Pharmacol.* **2000**, *169*, 17.
(139) Beger, R. D.; Freeman, J. P.; Lay, J. O., Jr.; Wilkes, J. G.; Miller, D. W. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 219.
(140) Tong, W.; Perkins, R.; Xing, L.; Welsh, W. J.; Sheehan, D. M. *Endocrinology* **1997**, *138*, 4022.
(141) Beger, R. D.; Holm, K. J.; Buzatu, D. A.; Wilkes, J. G. *Internet Electron. J. Mol. Des.* **2003**, *2*, 435http://www.biochempress.com.
(142) Asikainen, A.; Ruuskanen, J.; Tuppurainen, K. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1974.
(143) Beger, R. D.; Wilkes, J. G. *J. Comput. Aided Mol. Des.* **2001**, *15*, 659.
(144) Good, A. C.; So, S.-S.; Richards, W. G. *J. Med. Chem.* **1993**, *36*, 433.
(145) Kellogg, G. E.; Kier, L. B.; Gaillard, P.; Hall, L. H. *J. Comput. Aided Mol. Des.* **1996**, *10*, 513.
(146) Polanski, J. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 553.
(147) Guengerich, F. P. In *Cytochrome P450 Structure, Mechanism, and Biochemistry*, 3rd ed.; Ortiz de Montellano, P. R., Ed.; Kluwer: New York, 2005; pp 377−530.
(148) Lu, A. Y. H.; Coon, M. J. *J. Biol. Chem.* **1968**, *243*, 1331.
(149) Sheng, X.; Zhang, H.; Im, S.-C.; Horner, J. H.; Waskell, L.; Hollenberg, P. F.; Newcomb, M. *J. Am. Chem. Soc.* **2009**, *131*, 2971.
(150) Petzold, D. R.; Rein, H.; Schwarz, D.; Sommer, M.; Ruckpaul, K. *Biochim. Biophys. Acta* **1985**, *829*, 253.
(151) Bursi, R.; Dao, T.; van Wijk, T.; de Gooyer, M.; Kellenbach, E.; Verwer, P. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 861.
(152) Berking, E. W.; van Meel, F.; Turpijn, E. W.; van der Vies, J. *J. Steroid Biochem.* **1983**, *19*, 1563.
(153) Fauchère, J.-L.; Lauterwein, J. *Quant. Struct.-Act. Relat.* **1985**, *4*, 11.
(154) Pelech, S. L.; Vance, D. E. *Biochim. Biophys. Acta* **1984**, *779*, 217.
(155) Cuadrado, A.; Carnero, A.; Dolfi, F.; Jiménez, B.; Lacal, J. C. *Oncogene* **1993**, *8*, 2959.
(156) Campos, J.; del Carmen Núñez, M.; Rodríguez, V.; Gallo, M. A.; Espinosa, A. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 767.
(157) Rapoport, S. M.; Schewe, T.; Wiesner, R.; Halangk, W.; Ludwig, P.; Janicke-Hohne, M.; Tannert, C.; Hiebsch, C.; Klatt, D. *Eur. J. Biochem.* **1979**, *96*, 545.
(158) Samuelsson, B. *Science* **1983**, *220*, 568.
(159) Funk, C. D. *Science* **2001**, *294*, 1871.
(160) Helgadottir, A.; Manolescu, A.; Thorleifsson, G.; Gretarsdottir, S.; Jonsdottir, H.; Thorsteinsdottir, U.; Samani, N. J.; Gudmundsson, G.; Grant, S. F. A.; Thorgeirsson, G.; Sveinbjörnsdottir, S.; Valdimarsson, E. M.; Matthiasson, S. E.; Johannsson, H.; Gudmundsdottir, O.; Gurney, M. E.; Sainz, J.; Thorhallsdottir, M.; Andresdottir, M.; Frigge, M. L.; Topol, E. J.; Kong, A.; Gudnason, V.; Hakonarson, H.; Gulcher, J. R.; Stefansson, K. *Nat. Genet.* **2004**, *36*, 233.
(161) Nie, D.; Che, M.; Grignon, D.; Tang, K.; Honn, K. V. *Cancer Metastasis Rev.* **2001**, *20*, 195.
(162) Anderson, K. M.; Seed, T.; Vos, M.; Mulshine, J.; Meng, J.; Alrefai, W.; Ou, D.; Harris, J. E. *Prostate* **1998**, *37*, 161.
(163) Avis, I. M.; Jett, M.; Boyle, T.; Vos, M. D.; Moody, T.; Treston, A. M.; Martinez, A.; Mulshine, J. L. *J. Clin. Invest.* **1996**, *97*, 806.
(164) Soumaoro, L. T.; Iida, S.; Uetake, H.; Ishiguro, M.; Takagi, Y.; Higuchi, T.; Yasuno, M.; Enomoto, M.; Sugihara, K. *World J. Gastroenterol.* **2006**, *12*, 6355.
(165) Zhang, L.; Zhang, W.-P.; Hu, H.; Wang, M.-L.; Sheng, W.-W.; Yao, H.-T.; Ding, W.; Chen, Z.; Wei, E.-Q. *Neuropathology* **2006**, *26*, 99.
(166) Hlasta, D. J.; Casey, F. B.; Ferguson, E. W.; Gangell, S. J.; Heimann, M. R.; Jaeger, E. P.; Kullnig, R. K.; Gorgon, R. J. *J. Med. Chem.* **1991**, *34*, 1560.
(167) Verma, R. P.; Kapur, S.; Barberena, O.; Shusterman, A.; Hansch, C. H.; Selassie, C. D. *Chem. Res. Toxicol.* **2003**, *16*, 276.
(168) IJzerman, A. P.; Bultsma, T.; Timmerman, H. *J. Med. Chem.* **1986**, *29*, 549.
(169) Sparatore, F.; Caliendo, G.; La Rotonda, M. I.; Novellino, E.; Silipo, C.; Vittoria, A. *Quant. Struct.-Act. Relat.* **1988**, *7*, 178.
(170) Buzatu, D. A.; Beger, R. D.; Wilkes, J. G.; Lay, J. O., Jr. *Environ. Toxicol. Chem.* **2004**, *23*, 24.
(171) Wilkes, J. G.; Hass, B. S.; Buzatu, D. A.; Pence, L. M.; Archer, J. C.; Beger, R. D.; Schnackenberg, L. K.; Halbert, M. K.; Jennings, L.; Kodell, R. L. *Toxicol. Sci.* **2008**, *102*, 187.