International Journal of Engineering and Applied Physics (IJEAP)
Vol. 3, No. 3, September 2023, pp. 843~857
ISSN: 2737-8071

Substituent analysis for accurate prediction of the acidity constants of phosphonic acids

Rachid BOUADDI[1]*, Abdelilah FATNI[1], Lahcen EL MAIMOUNI[2] , Abdellatif EL HAMMADI[1]

[1] Modeling, Materials and Physical Chemistry of the Environment Team, Laboratory of Analysis, Modeling, Engineering, Natural Substances, and Environment, Polydisciplinary Faculty of Taroudant, Ibn Zohr University, Agadir, Morocco
[2] Applied Chemistry, Geo-mining, and Modeling Research Team,
Polydisciplinary Faculty of Ouarzazate, Ibn Zohr University, Morocco

Article Info

Article history:

Received Jun 18, 2023
Revised Jun 25, 2023
Accepted July 15, 2023

Keywords:

Substituent effects
Linear regression
Acidity constants
Phosphonic acids
Statistical analysis

ABSTRACT

This study aims to develop simple and reliable models to predict the pKa of phosphonic acids using empirical substituent effects descriptors. The effects of substituents on the acidity and basicity of organic acids have been examined using various approaches, such as the Taft-Topsom model. This study used linear regression methods to construct equations that predict the acidity constants of phosphonic acids. The validity of the linear regressions was confirmed through statistical analysis of parameters and residuals. However, it is essential to note that the models are limited to substituents with tabulated Taft parameters. Despite these limitations, the developed models provide a simple and effective method for estimating the acidity constants of phosphonic acids with a confidence interval greater than 95%. This study aims to analyze the effect of substituents on the dissociation of phosphonic acid and propose prediction models for new phosphonic derivatives

Corresponding Author:

Rachid BOUADDI,
LAMISNE Lab,
Faculté Polydisciplinaire Taroudant, Université Ibn Zohr.
Email: rach.bouaddi@gmail.com

1. Introduction

This work is a follow-up of previous works [1]on determining pKa phosphonic acids by ab initio methods. The central idea of this work is to examine the substituent effects in the case of a phosphonic function where the central atom is a phosphorus atom, an element of the third period of the periodic table. The aim is to examine the role played by the central atom in transferring substituent effects to the deprotonation center and to propose alternative models for predicting pKa1 and pKa2 constants of phosphonic acids.

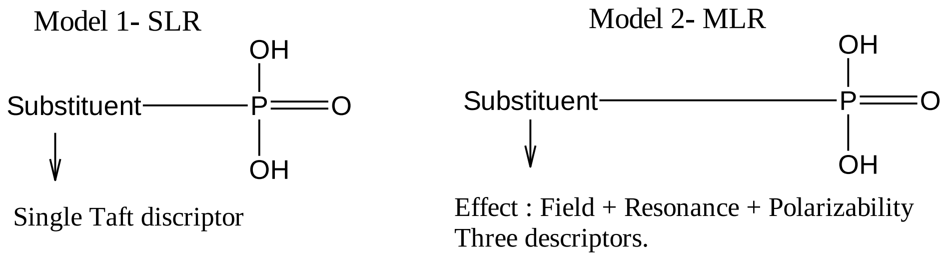

Figure 1: Description of the two models for predicting the effect of substituents on phosphonic acids

The substituent effect can thus be defined as a perturbation of a given property (acidity constant, rate or equilibrium constant, electronic structure ...) induced by a substituent, compared to the unsubstituted compound.

The first attempt to quantify the effects of substituents is due to the work of Hammett[2], and basicity of organic acids such as the Taft-Topsom model[3].

The first quantitative approach to describing the substituent effect was introduced by Louis Plack Hammett, who linked the rates and equilibria of many reactions involving aromatic meta and para phenyl compounds. This method has been successfully applied to a large number of reactions and substituents[4].

Indeed, Hammett's equation failed to explain the structural effect of ortho-benzene and aliphatic derivatives. Subsequently, Taft formulated an equation [5]based on the same motifs as Hammett's equation[6] but using data on acid and base-catalyzed hydrolysis of esters.

Before discussing methods for separating these effects, it is necessary to define inductive and mesomeric effects, as well as other terms that will be used later in this manuscript.

- Electronegativity: is a concept introduced in 1932 by Pauling [7]: It measures the power of an atom or a group of atoms to attract electrons from the rest of the molecular entity considered.
- Inductive effect: the definition retained by IUPAC is an experimentally observable effect of charge transmission along a chain of atoms by electrostatic induction[8].

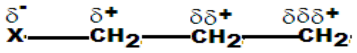

Figure 2: The inductive effect propagates along a carbon chain but rapidly weakens as one moves away from the source of the effect

The field effect corresponds to transmission in space: this effect, indicated by the letter F, is often considered predominant compared to the inductive effect as soon as more than two carbon atoms separate the reaction center from the substituent. The Kirkwood and Westheimer equation can estimate this intramolecular coulombic interaction between the center of interest and the induced dipole [9].

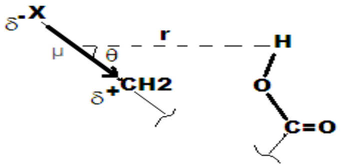

Figure 3: Illustration of the spatial transmission (field effect) between substituent X and carboxylic acid reaction center

- Polarisability is a specific molecular property [10]that describes the tendency of a charge distribution to be distorted from its equilibrium arrangement. In addition to the inductive and field effects, the substituent polarizability effect affects the acidity and basicity of gas phase systems.

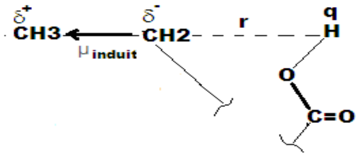

Figure 4: Illustration of the polarizability effect between the CH3 substituent and the carboxylic acid proton

- The mesomeric effect [11]is the effect on reaction constants and ionization equilibria attributed to a substituent and caused by the overlap of the p or π orbitals of the substituent with the p or π orbitals of the rest of the molecular entity. This effect can also be called the resonance effect or the conjugation effect.

Taft's approach consists in separating the inductive and mesomeric effects according to the following equation, known as the DSP [12] (dual substituent parameters) equation:

$$log\frac{k}{k_0} = \rho_I\sigma_I + \rho_R\sigma_R$$

- Taft, therefore, assumed the additivity of the inductive and mesomeric components[13]. Subsequently, based on physicochemical and spectroscopic properties, he evaluated the values of these two parameters.
- Taft and Topsom proposed a four-parameter partition to improve his approach further, introducing a polarization term and splitting the inductive term into a majority, so-called field contribution, denoted by F, and an electronegativity-related contribution, denoted by χ. The equation is written as follows:

$$\log\frac{\mathrm{k}}{\mathrm{k}_0} = \rho_\alpha\sigma_{\sigma_\mathrm{R}} + \rho_\mathrm{F}\sigma_\mathrm{F} + \rho_\mathrm{R}\sigma_\mathrm{R} + \rho_\mathrm{X}\sigma_\mathrm{X}$$

- This model has been successfully applied to many reactions [11], [14]–[16].

2. Proposed Method and test

2.1. Model's basic concept

We propose to present the model used to analyze the effect of substituents on the deprotonation of phosphonic acids in the aqueous phase treated in this work. This model is based on the Gibbs free energy corresponding to the isodesmic proton exchange reaction between a substituted phosphonic acid $\mathrm{HAH_n^p}$ and a conjugated base $\mathrm{HAH_{n-1}^{p-1}}$ of the acid $\mathrm{HAH_n^p}$ as reference:

$$\mathrm{XAH_n^p + HAH_{n-1}^{p-1} \xleftrightarrow{\Delta G_{r(aq)}} XAH_{n-1(aq)}^{p-1} + HAH_n^p}$$

Where p and p-1 represent the overall charge of the acid species and its conjugate base respectively for n=2, p=0 (first dissociation, r=1) and n=1, p=1 (second dissociation, r=2). ΔGr(aq) is the change in Gibbs energy of the reaction (1) related to the equilibrium constants KX and KH by the following relations:

$$\log\frac{\mathrm{K_X}}{\mathrm{K_H}} = -\frac{\Delta\mathrm{G_r}}{2.3\mathrm{RT}}$$

$$\log\mathrm{K_X} - \log\mathrm{K_H} = -\frac{\Delta\mathrm{G_r}}{2.3\mathrm{RT}}$$

$$\mathrm{pka_X - pka_H} = \frac{\Delta\mathrm{G_r}}{2.3\mathrm{RT}}$$

$$\delta_\mathrm{x}\mathrm{pka_r} = \frac{\Delta\mathrm{G_r^0}}{2.3\mathrm{RT}}$$

$\delta_\mathrm{x}\mathrm{pka_r}$ is determined from experimental or theoretical $\mathrm{pka_r}$ values.

To analyze the substituent effects on $\delta_\mathrm{x}\mathrm{pka_r}$, we chose to use a linear regression model:

$$\mathrm{y_i} = \rho_0\mathrm{x}_0 + \sum_{\mathrm{j}=1}^{\mathrm{p}}\rho_\mathrm{j}\,\mathrm{x_j} + \varepsilon$$

- $\mathrm{y} = \delta_\mathrm{x}\mathrm{pka}$ is the observed variable, to explain ($\in$ R)
- x_0 is a fictitious parameter equal to 1.
- x1, …….xp are explanatory variables ($\in$ R)
- ε is the random error term of the model, also known as the model residual
- $\rho_0, \rho_1, \rho_2, \ldots\ldots \rho_\mathrm{p}$ are the parameters to estimate.

For N observations, we can write the N equations of the model in the following form:

$$y_i = \rho_0 x_0 + \sum_{j=1}^{p} \rho_j \, x_{ij} + \varepsilon_i \text{i=1, 2, 3 …. N}$$

With $$y_i = \hat{y}_i + \varepsilon_i$$

- $\hat{y}_i$ is the estimate (value predicted by the model) of y_i (experimental or calculated value),
- ρ_j are the unknown coefficients of the postulated mathematical model.

We can also write the model (3) in matrix form:

$$Y = X * \rho + \varepsilon$$

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_N \end{pmatrix}; X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots\cdot \\ 1 & x_{N1} & 1 & x_{Np} \end{pmatrix}; \rho = \begin{pmatrix} \rho_0 \\ \rho_1 \\ \cdots \\ \rho_p \end{pmatrix}; \varepsilon = \begin{pmatrix} \varepsilon_1 \\ y_2 \\ \cdots \\ y_N \end{pmatrix}$$

- Y : denotes the vector to be explained of size N,
- X : the explanatory matrix of size N- (p + 1),
- ε : the error vector of size N.

2.2. Estimation of statistical parameters ρ0,ρj (j=1, 2…p) and variance σ2

To estimate the coefficients ρ0,ρj (j=1, 2…p) we can use the principle of the least squares [19].

We search for $\overbrace{\rho}$ that minimizes the sum of the quadratic errors:

$$\begin{aligned} &\min \sum_{i=1}^{N} \varepsilon_i^2 = \min \varepsilon^t \, \varepsilon = \min\left(Y - \hat{Y}\right)^t \left(Y - \hat{Y}\right) \\ &= \min(Y - X\rho)^t \, (Y - X\rho) \\ &= \min S\,(\rho) \end{aligned}$$

Where εt is the transposed matrix of ε,

$\min \sum_{i=1}^{N} \varepsilon_i^2$ is the sum of squares of the residuals noted (SCR).

And $$S(\rho) = (Y - X\rho)^t(Y - X\rho) = Y^tY - \rho^tX^tY - Y^tX\rho + \rho^tX^tX\rho$$

The extremum condition of $S(\rho)$:

$$\frac{\partial S(\rho)}{\partial \rho} = -2X^tY + 2X^tX\rho \Rightarrow \hat{\rho} = (X^tX)^{-1}X^tY$$

Under the condition that the matrix (XtX) is invertible.

The model is written as follows: $y_i = \hat{\rho}_0 x_0 + \sum_{j=1}^{p} \hat{\rho}_j \, x_{ij} + \varepsilon_i \text{avec} \varepsilon_i = y_i - \hat{y}_i$

The parameter σ2 is estimated by:

$$s^2 = \frac{1}{(N-p-1)}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2 = \frac{\sum_{i=1}^{N}\varepsilon_i^2}{(N-p-1)} = \frac{SCR}{N-p-1}$$

- s2 is an unbiased estimator of σ2
- p+1 degree of freedom in the expression of σ2 is the "cost" of estimating the coefficients ρ0, ρ1, ... ρp needed to obtain the $\hat{y}_i$.

2.3. Hypothesis testing and range for coefficients ρj

- Test of model significance

Generally, we start by testing:

The Ho hypothesis: ρ1=ρ2=………ρp=0. Against the alternative hypothesis H1 : $\rho_j \neq 0, \forall j = 1,2,\ldots\ldots\ldots p$

We use the following Fisher statistic:

$$F \quad = \frac{\sum_{i=1}^{N}\frac{(\hat{y}_i-\bar{\hat{y}})^2}{p}}{\sum_{i=1}^{N}\frac{(y_i-\hat{y}_i)^2}{N-p-1}} = \frac{\frac{SCE}{p}}{\frac{SCR}{N-p-1}} = \frac{SCE}{SCR}\left(\frac{N-p-1}{p}\right)$$

Which is distributed under H0 according to a Fisher distribution with p and (N - p - 1) degrees of freedom.

SCE is the sum of explained squares, and SCR is the sum of residual squares, described in table 1.

We reject H0 with a risk $0 < \alpha < 1$ si :$F \quad \geq F_{1-\alpha/2}(p, N-p-1)$.

$F_{1-\alpha/2}(p, N-p-1)$ is the value of order 1-α of the Fisher distribution F(p, n - p - 1). Generally, α=0.05.

Table 1: Analysis of variance table (ANOVA)

Regression	**Sum of squares**	**Degree of freedom (ddl)**	**Mean of squares**	**F**
Explained (SCE)	$\sum_{i=1}^{N}(\hat{y}_i-\bar{\hat{y}})^2$	**p**	$\sum_{i=1}^{N}\frac{(\hat{y}_i-\bar{\hat{y}})^2}{p}$	$\frac{SCE}{SCR}\left(\frac{N-p-1}{p}\right)$
Residual (SCR)	$\sum_{i=1}^{N}(y_i-\hat{y}_i)^2$	**N-p-1**	$\sum_{i=1}^{N}\frac{(y_i-\hat{y}_i)^2}{N-p-1}$	
Total (SCT)	$\sum_{i=1}^{N}(y_i-\bar{\hat{y}})^2$	**N-1**	$\sum_{i=1}^{N}\frac{(y_i-\bar{\hat{y}}_i)^2}{N-1}$	

With the fundamental property SCT =SCE +SCR

The Fisher F statistic is used to test the fit of the model by the coefficient of determination[17] (or correlation) R2:

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT}$$

The R2 coefficient gives the proportion of variability in y that the model explains.

The closer the R2 is to 1, the better the fit of the model to the data.

If N < p, then the adjusted coefficient of determination is calculated:

$$\bar{R}^2 = R^2 - (1 - R^2)\frac{N-1}{N-p-1}$$

Test of significance of the coefficients ρj

A coefficient is considered significant if it is significantly different from zero for a certain risk. We will therefore test the hypothesis: H0: ρ1=ρ2=………ρp=0 against the hypothesis H1 : ρj≠ 0.

To do this, we use the Student statistic for: ρj=0, with: $|t| = \frac{|\hat{\rho}_j|}{\sigma_{\rho_j}}$

Where t follows a student's law; $\hat{\rho}_j$: the coefficient of the regression equation; σρj is the standard deviation or dispersion error of the coefficient ρj.

For a significance level of 0<α<1 and a number of degrees of freedom (N-p-1), the Student's table for a two-tailed test gives the value of t1-α/2(N-p-1).

If:

|t| > t1-α/2(N-p-1), we reject H0 at the accepted risk.

|t < t1-α/2 (N-p-1), we accept H0 at accepted risk.

If the H0 hypothesis is accepted, this means that the coefficient in question is not significantly different from zero at the α risk and, therefore, that the variable associated with it does not influence the y response.

To perform this test, we can also: look at the p-value, also called the significance level of the test:

If p-valeur ≤α on rejecting H0.

In the case of a bilateral test (H1: ρj≠ 0), with: p-valeur =P(|TN|) > |t| / H0). On reject H0 if p-valeur ≤α

These methods are incorporated in statistical software such as "Origin", SPSS, Excel stat ... or "R". In this work, multiple linear regression is performed using the "Origin v6" software.

Residuals analysis

The statistical analysis of the linear regression relies on assumptions related to the error term ε that embodies information absent from the model. It seems most appropriate to test these assumptions in order to validate the regression and interpret the results.

These assumptions are tested if the residuals: Are independent, are distributed according to a standard law of mean 0, and are distributed homogeneously, i.e., with a constant variance.

To do this, we examine the residuals $\hat{\varepsilon}_i$, the observed errors generated by the difference between the observed values of yi and the point predictions of the regression $\hat{y}_i$.

$$\hat{\varepsilon}_i = y_i - \hat{y}_i$$

With: $\hat{y}_i = \hat{\rho}_0 x_0 + \sum_{j=1}^{p} \hat{\rho}_j \, x_{i,j}$

The mean of the variances: $\bar{\hat{\varepsilon}} = \frac{1}{N}\sum_{i=1}^{N} \hat{\varepsilon}_i$

The variance of deviations: $\mathrm{Var}(\hat{\varepsilon}_i) = \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\varepsilon}_i - \bar{\hat{\varepsilon}}\right)^2$

The standard deviation is given by: $\sqrt{\frac{\mathrm{Var}(\hat{\varepsilon}_i)}{N}} = \sqrt{\frac{N}{N-1}\sum_{i=1}^{N}\left(\hat{\varepsilon}_i - \bar{\hat{\varepsilon}}\right)^2}$

The closer the mean of the residuals is to zero with a standard deviation close to zero, the better the linear regression.

3. Database and Methodology

The database collected in the following Table 2 includes pKa experimental values[18][19], [20], theoretically calculated [1] values by the DFT method of the 24 acids with different substituents, and the Taft parameter σ* of 130 neutral aliphatic substituents taken from Table 9.1 of the Lange's Handbook of Chemistry [21] including those of the 24 phosphonic acids studied.

This group is very diverse with a wide variety of atoms attached to the phosphonic function: substituents with a saturated carbon of sp3 hybridization, unsaturated of sp2 and sp hybridization; substituents with an amine or amide group, substituents with an O-alkyl group, a sulfur group hybridized sp3 and sp2, silanes and halogens (F, Cl, Br and I).

A data set of pKa constants of carboxylic acids for comparison with estimated pKa values of phosphonic acids, the proton transfer reactions studied are summarized as follows:

$$\mathbf{XPO_3H_2 + HPO_3H^- \xleftrightarrow{(1)} XPO_3H^- + HPO_3H_2\delta_X pka_1}$$
$$\mathbf{XPO_3H^- + HPO_3^{--} \xleftrightarrow{(2)} XPO_3^{2-} + HPO_3H^-\delta_X pka_2}$$

Where XPO3H2 is the substituted acid, and HPO3H2 is the reference acid.

$\delta_X$pka is the experimental or theoretical variable whose value depends on the nature and effect of the substituent on the dissociation reactions.

We used the simple and multiple linear regression model illustrated in the following sections to approximate these substituent effects.

Table 2: Taft parameter and acidity constant values of phosphonic

		XPO_3H_2					XCOOH
N°	X	σ*	pKa_1ex	pka2_exp	pka1_calc	pka2_calc	pka
1	H	0	1,3	6,7	1,3	6,7	375
	H	0	1,5	6,79			3.77
2	CH_3	-0,49	2,35	7,2	3,05	7,91	4.76
	CH_3	-0,49	2,12	7,29			4.76
	CH_3	-0,49	2,41	7,35			4.757
3	CH_2CH_3	-0,59	2,43	7,85	2,93	8,08	4.88
	CH_2CH_3	-0,59	2,45	7,6			
4	$CH_2CH_2CH_3$	-0,61	2,45	8,06	3,5	7,71	4.82
	$CH_2CH_2CH_3$	-0,61	2,49	8,18			4.976
5	$CH_2CH_2CH_2CH_3$	-0,62	2,59	8,19	3,21	7,82	4.84
6	$CH_2(CH_2)_4CH_3$	-0,49	2,6	7,9	2,84	8,2	4.89
7	$CH(CH_3)_2$	-0,68	2,55	7,75	3,46	8,01	4.84
	$CH(CH_3)_2$	-0,68	2,66	8,44			4.599
8	$CH_2CH(CH_3)_2$	-0,62	2,7	8,43	2,93	9,47	4,77
9	$CH(CH_3)CH_2CH_3$	-0,68	2,74	8,48	3,25	8	4,97
10	$C(CH_3)_3$	-0,79	2,79	8,88	3,77	8,11	5,03
	$C(CH_3)_3$	-0,79		8,71			5.025
11	$CH_2C(CH_3)_3$	-0,61	2,84	8,65	3,78	8,57	4,79
12	$C(CH_3)_2C_2H_5$	-0,49	2,88	8,96	3,32	7,92	
13	$CH_2C_6H_5$	-0,27	2,3	7,55	2,52	6,98	
14	$CH{=}CH_2$	0,07	3,48	8,54	1,94	7,02	4,25
15	$CH{=}CHC_6H_5$	-0,08	2	7,1	2,06	6,94	
16	C_6H_5	0,11	2,05	5,51	1,7	6,69	4,312
17	CH_2OH	-0,18	1,91	7,15	1,62	6,47	3,83
18	CF_3	2,12	1,16	3,93	-6,25	-0,63	5,2
19	CH_2Cl	0,56	1,51	6,17	-0,72	4,74	2.86
20	$CHCl_2$	1,45	1,14	4,97	-3,45	2,49	1,301
	$CHCl_2$	1,45		5,58			1.479
21	CCl_3	2,16	1,63	4,81	-5,96	-0,03	5,2
22	CH_2Br	0,51	1,14	5,62	0,1	6,06	2,889
23	CH_2I	0,36	1,3	6,72	0,99	8,13	2.901
24	$CH_2Si(CH_3)_3$	-0,74	3,22	8,7	1,79	8,56	2.86
25	$CH_2CH_2CH(CH_3)_2$	-0,66					
26	$CH_2CH_2CH_2CH_2CH_3$	-0,74					4,879
27	CH_2NO_2	0,91					1,48
28	CH_2OCH_3	0,03					3,4
29	$CH(OH)CH_3$	-0,37					3,856
30	$CH(OH)C_6H_5$	0,01					
31	CHF_2	1,56					
32	CH_2F	0,61					2,72
33	CH_2SH	0,13					3,55
34	CH_2CH_2Br	-0,49	2,25	7,3			
35	cyclohexyl	-0,64					4,914
36	$CH{=}C(CH_3)_2$	-0,3					5,137
37	$C{\equiv}CH$	1,69					1,887
38	$COCH_3$	1,16					2,105
39	$COOH$	1,59					1,52
40	NH_2	0,13	3,08	8,63	4,03(†)	9,37(†)	
41	OH	0,85	1,97	6,82	2,92(†)	7,56(†)	
42	OCH_3	1,32	1,54	6,31	2,49(†)	7,05(†)	3,53
43	OC_2H_5	1,19	1,6	6,62	2,55(†)	7,36(†)	
44	OC_3H_7	1,19	1,88	6,67	2,83(†)	7,41(†)	
45	OC_4H_9	1,19	1,8	6,84	2,75(†)	7,58(†)	

4. RESULTS AND DISCUSSIONS

4.1. Model M1- sample linear regression

The model M1 is represented by a simple linear model that can be expressed as:

$$\delta_x \mathrm{pk}\hat{a}_i^* = \hat{\rho}_0 + \hat{\rho}\sigma_i^*$$

Where σ*i is the Taft constant whose value is shown in the preceding table, we seek to estimate ρ0 and ρ are unknowns by simple linear regression (SLR). The statistical parameters of the correlation between observed values and the parameter σ* are recorded in Table 3.

Table 3: Statistical parameters of the linear regression $\delta_x pka_i$ vs σ* of phosphonic acids (experimental and theoretical)

	Coefficients	**ρ**	**Δρ**	**t-value**	**R²**	**DS**	**N**	**P**
$\delta_x p\hat{k}a_1^{*exp}$	**ρ₀**	**0,55**	**0,07**	**7,86**	**0,86**	**0,23**	**18**	**<0.0001**
	ρ	**-1,26**	**0,13**	**-9,69**				
$\delta_x p\hat{k}a_2^{*exp}$	**ρ₀**	**0,52**	**0,09**	**5,78**	**0,91**	**0,4**	**19**	**<0.0001**
	ρ	**-1,41**	**0,11**	**-12,82**				
$\delta_x p\hat{k}a_1^{*cal}$	**ρ₀**	**0,01**	**0,14**	**0,07**	**0,95**	**0,63**	**21**	**<0.0001**
	ρ	**-3,23**	**0,17**	**-19**				
$\delta_x p\hat{k}a_2^{*cal}$	**ρ₀**	**-0,25**	**0,13**	**-1,92**	**0,9**	**0,84**	**21**	**<0.0001**
	ρ	**-2,97**	**0,16**	**-18,56**				

The correlation of the experimental and theoretical values of $\delta_x pka_i$ with the parameter σ* is well elevated, attesting to the high dependence of this descriptor on the dissociation constant $\delta_x pka_i$ show in figure 5, the correlation coefficient R2 varies between 0.86 and 0.95, and the standard deviation (0.2<SD<0.6) is less than unity.

Regarding the significance of the coefficients, the absolute values of |t-Student|> |t|0.975 (2.12, 2.11, 2.10 for ddl=16, 17, 18 according to the statistical tables and therefore, the H0 hypothesis: ρ=0 is rejected. This shows that the coefficients are significant for experimental (ρ0 and ρ) and theoretical (ρ) pKa. However, for the latter, the coefficient ρ0 is insignificant and does not affect the variation of the theoretical constates.

The slopes of the correlation lines are of the negative sign with absolute values of 1.26 for $\delta_x pka_1$ and 1.41 for $\delta_x pka_2$ at the experimental level and 2.96 for $\delta_x pka_1$ and 3.24 for $\delta_x pka_2$ at the theoretical level.

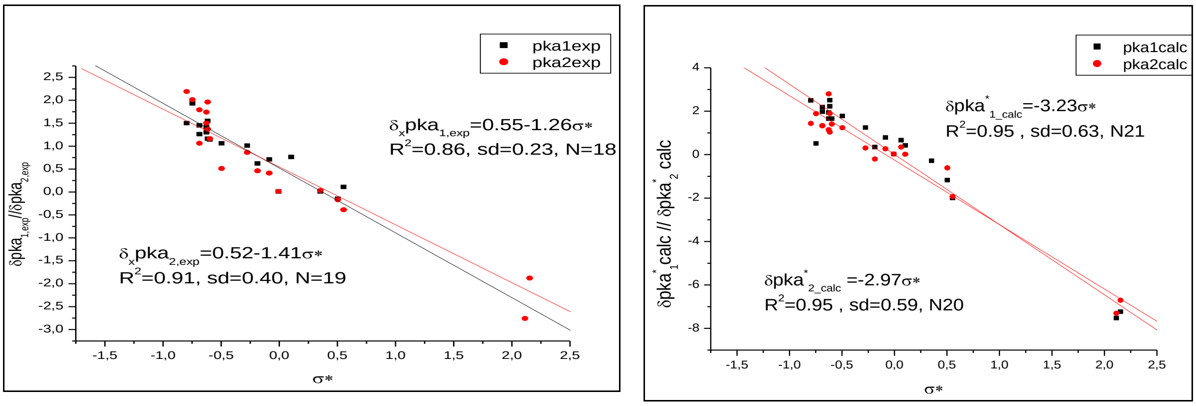

Figure 5: Correlation between the $\delta_x pka_i$ constants (experimental and theoretical) of phosphonic acids and the Taft constant σ*

These values are in the order of magnitude of the acidity constant prediction equations established by Taft for aliphatic carboxylic acids RCOOH (4.66-1.62σ*) [21-22]. The compounds excluded from the correlations have deviations from the experimentally or theoretically observed values greater than the pKa unit (Table 4).

Table 4: pKa values estimated from the M1 model equations with deviations from the observed values (experimental and theoretical).

	RPO_3H_2	This work M1-SLR							
		$pka_{1,i}^{*,\exp}$	$\hat{\varepsilon}_{1,i}^{\exp}$	$pka_{2,i}^{*,\exp}$	$\hat{\varepsilon}_{2,i}^{\exp}$	$pka_{1,i}^{*,calc}$	$\hat{\varepsilon}_{1,i}^{calc}$	$pka_{2,i}^{*,calc}$	$\hat{\varepsilon}_{2,i}^{calc}$
1	H	1,85	-0,55	7,22	-0,52	1,3	0	6,7	0
		1,85	-0,35	7,22	-0,43				
2	CH_3	2,47	-0,12	7,91	-0,71	3,05	-0,17	7,91	-0,24
		2,47	-0,35	7,91	-0,62				
		2,47	-0,06	7,91	-0,56				
		2,47	-0,12	7,91	-0,81				
3	CH_2CH_3	2,59	-0,16	8,05	-0,2	2,93	0,28	8,08	-0,37
		2,59	0.14	8,05	-0,45				
				8,05	0				
4	$CH_2CH_2CH_3$	2,62	-0,17	8,08	-0,02	3,5	-0,23	7,71	-0,8
		1,3	1.19	8,08	0,1				
				8,08	0,09				
5	$CH_2CH_2CH_2CH_3$	2,63	-0,04	8,09	0,1	3,21	0,09	7,82	-0,72
7	$CH(CH_3)_2$	2,71	-0,16	8,18	-0,43				
		1,3	1.36	8,18	0,26				
8	$CH_2CH(CH_3)_2$	2,63	0,07	8,09	0,34	2,93	0,37	9,47	0,93
9	$CH(CH_3)CH_2CH_3$	2,71	0,03	8,18	0,3	3,25	0,25	8	-0,71
10	$C(CH_3)_3$	2,85	-0,06	8,33	0,55	3,77	0,08	8,11	-0,93
				8,33	0,38				
11	$CH_2C(CH_3)_3$	2,62	0,22	8,08	0,57	3,78	-0,51	8,57	0,06
13	$CH_2C_6H_5$	2,19	0,11	7,6	-0,05	2,52	-0,35	6,98	-0,52
14	$CH{=}CH_2$	1,76	1,72	7,12	1,42	1,94	-0,87	7,02	0,53
15	$CH{=}CHC_6H_5$	1,95	0,05	7,33	-0,23	2,06	-0,5	6,94	0
16	C_6H_5	1,71	0,34	7,06	-1,55	1,7	-0,76	6,69	0,32
17	CH_2OH	2,08	-0,17	7,47	-0,32	1,62	0,26	6,47	-0,76
18	CF_3	-0,82	1,98	4,23	-0,3	-6,25	0,7	-0,63	-1,05
19	CH_2Cl	1,14	0,37	6,43	-0,26	-0,72	0,21	4,74	-0,3
				6,43	-0,13				
				6,43	0,16				
20	$CHCl_2$	0,02	1,12	5,18	-0,21	-3,45	0,07	2,49	0,08
				5,18	0,4				
				5,18	0,43				
21	CCl_3	-0,87	2,5	4,17	0,64	-5,96	0,28	-0,03	-0,34
22	CH_2Br	1,21	-0,07	6,5	-0,88	0,1	-0,45	6,06	0,87
23	CH_2I	1,4	-0,1	6,71	0,01	0,99	-0,85	8,13	2,5
24	$CH_2Si(CH_3)_3$	2,78	0,44	8,26	0,44	1,79	1,9	8,56	-0,33

Sum of deviations	**6,29**		**-3,81**		**1,15**		**-9,18**
Average of deviations	**0,2**		**-0,09**		**0,05**		**-0,44**
Standard deviation	**0,02**		**0,01**		**0,02**		**0,14**

Residue analysis: Table 4 shows the sum, mean, and standard deviation ε̂_between the observed and estimated values from the following equations :

$$\begin{aligned} \hat{y}_{1,i\_exp} &= -1.26\sigma_i^* \\ \hat{y}_{2,i\_exp} &= -1.41\sigma_i^* \\ \hat{y}_{1,i\_calc} &= -3.23\sigma_i^* \\ \hat{y}_{2,i\_cal} &= -2.96\sigma_i^* \end{aligned}$$

The relative constant of a phosphonic acid can be deduced from the following equation:

$$\left.\begin{matrix} \delta_x p\widehat{k}a_{k,i} = p\widehat{k}a_{k,i} - pka_{k,H} \\ = \rho_{0,i} + \rho\sigma_i^* \end{matrix}\right\} \Rightarrow p\widehat{k}a_{k,i} = pka_{k,H} + \rho_{0,i} + \rho\sigma_i^*$$

Where $pka_{r,H}$ is the acidity constant of the reference acid-base couples:

pKa1 (HPO3H2/HPO3H-) = 1.3 et pKa2 (HPO3H-/HPO32-) =6.7[23]

The estimated relative pKa* is thus deduced from the equations:

$$\boldsymbol{p\widehat{k}a_{1,i}^{*exp_i^*} = 1.85 - 1.26\sigma_i^*}$$

$$\boldsymbol{p\widehat{k}a_{2,i}^{*exp_i^*} = 7.22 - 1.41\sigma_i^*}$$

$$\boldsymbol{p\widehat{k}a_{1,i}^{*calc} = 1.3 - 3.23\sigma_i^*}$$

$$\boldsymbol{p\widehat{k}a_{2,i}^{*calc} = 6.7 - 2.96\sigma_i^*}$$

The domain of applicability of the model provided that ($\sigma_i^* \in [-0.84, 2.16]$), It is observed that the mean of the residuals is close to zero with a fairly low standard deviation. This result is a satisfactory indication of the good quality of the simple linear regression applied. The postulated M1 model is thus relevant and can be used to predict the relative acidity constants of the phosphonic derivatives with a confidence interval higher than 95%.

4.2. Model M2- Multiple linear regression

In Model M2, the quantitative variable (yi) to be explained (or response) is related to p explanatory quantitative variables (xij), called controls; here we try to explain yi through multiple linear regression (MLR), the general expression of the model 2 thus postulated is:

$$y_i = \rho_0 + \sum_{j=1}^{p} \rho_j x_{i,j} + \hat{\varepsilon}_i (i = 1,2,\ldots\ldots\ldots\ldots\ldots N) \text{ Or: } \quad y_i = \hat{y}_i + \varepsilon_i$$

ε_i is the difference between the observed value and the value predicted by the model. xi,j, the model parameters, and ρj are the coefficients to be determined.

In this study, the postulated model 2 is obtained from the linear combination of the empirical descriptors tabulated by Taft [ref] $\sigma_\alpha, \sigma_F et \sigma_R$, which are related to the polarizability, field, and resonance effects, respectively. The overall equation of this model 2 is thus written:

$$\hat{y}_i = \delta_x pka_i = \rho_0 + \rho_{\alpha,i}\sigma_\alpha + \rho_{F,i}\sigma_F + \rho_{R,i}\sigma_R$$

The model 2 developed from the variation $\delta_x pka_i$ (experimental and calculated) of studied phosphonic acids is represented by 4 equations whose statistical parameters are shown in Table 5.

Table 5: Statistical parameters of multiple linear regression, model 2 for phosphonic acids. Carboxylic acids are reported for comparison.

$\delta_x p\widehat{ka}_1^{exp}$						$\delta_x p\widehat{ka}_2^{exp}$					
Parameter	**Value**	**Error**	**t-Value**	**Prob>|t|**		**Parameter**	**Value**	**Error**	**t-Value**	**Prob>|t|**	
ρ_α	0,55	0,31	1,8084	0,0856		ρα	0,66	0,54	1,2315	0,2324	
ρ_χ	-1,02	0,44	-2,2974	0,0325		ρχ	-0,89	0,78	-1,1445	0,2659	
ρF	-2,91	0,6	-4,8222	0,0001		ρF	-7,55	1,06	-7,13	<0.0001	
ρR	-1,58	0,63	-2,4951	0,0215		ρR	-2,6	1,11	-2,3431	0,0296	
R2	**Adj. R2**	**SD**	**N**			**R2**	**Adj.R2**	**SD**	**N**		
0,64	0,59	0,42	24			0,75	0,71	0,73	24		
	Degrees of Freedom	**Sum of Squares**	**Mean Square**	**F Statistic**	**Prob>F**		**Degrees Freedom**	**Sum of Squares**	**Mean Square**	**F Statistic**	**Prob>F**
Model	3	6,22	2,07	11,93	<0.0001	Model	3	31,71	10,57	19,78	<0.0001
Error	20	3,48	0,17			Error	20	10,69	0,53		
Total	23	9,69				Total	23	42,4			
$\delta_x p\widehat{ka}_1^{calc}$						$\delta_x p\widehat{ka}_2^{calc}$					
Parameter	**Value**	**Error**	**t-Value**	**Prob>|t|**		**Parameter**	**Value**	**Error**	**t-Value**	**Prob>|t|**	
ρα	0,8	0,57	1,3951	0,1783		ρα	0,59	0,51	1,1535	0,2623	
ρχ	-0,42	0,83	-0,5112	0,6148		ρχ	-0,14	0,74	-0,1841	0,8558	
ρF	-17,49	1,13	-15,4625	<0.0001		ρF	-15,91	1,02	-15,6589	<0.0001	
ρR	-10,63	1,19	-8,965	<0.0001		ρR	-9,09	1,06	-8,5361	<0.0001	
R^2	**Adj. R^2**	**SD**	**N**			**R^2**	**Adj. R^2**	**SD**	**N**		
0,94	0,93	0,78	24			0,94	0,93	0,7	24		
	Degrees of Freedom	**Sum of Squares**	**Mean Square**	**F Statistic**	**Prob>F**		**Degrees of Freedom**	**Sum of Squares**	**Mean Square**	**F Statistic**	**Prob>F**
Model	3	180,97	60,32	98,98	<0.0001	Model	3	145,69	48,56	98,81	<0.0001
Error	20	12,19	0,61			Error	20	9,83	0,49		
Total	23	193,16				Total	23	155,52			
$\delta_x p\hat{k}a_{XCOOH}^{exp}$											
Parameter	**Value**	**Error**	**t-Value**	**Prob>|t|**							
ρα	0,32	0,14	2,3	0,04							
ρχ	-0,79	0,21	-3,72	0							
ρF	-9,13	0,28	-33,15	<0.0001							
ρR	-3,43	0,45	-7,61	<0.0001							
R_2	**Adj. R_2**	**SD**	**N**								
0,99	0,99	0,18	19								
	Degrees of Freedom	**Sum of Squares**	**Mean Square**	**F Statistic**	**Prob>F**						
Model	3	41,72	13,91	414,33	<0.0001						
Error	15	0,5	0,03								
Total	18	42,22									

These statistical data show a relatively high correlation coefficient (R2=0.94) for the theoretically observed dissociation constants $\delta_x pka_{1,i}^{cal}$ and $\delta_x pka_{2,i}^{cal}$ with a standard deviation of less than 0. 8 unit pKa and a Fisher index close to 100. In contrast, the values of these parameters are relatively low for the experimentally $\delta_x pka_{1,i}^{exp}$ and $\delta_x pka_{2,i}^{exp}$.

The coefficients ρF and ρR are significant in all four equations ([Prob>|t|] > 0.05), whereas ρ0 and ρα are insignificant in three equations where the H0 hypothesis is accepted with relatively small values. These have no significant effect on the $\delta_x$pka variation of phosphonic acids can be neglected with a 5% risk.

Furthermore, the coefficients ρα ,ρF, and ρR are negative. These negative values indicate that the dissociation constant pKa and substituent effects move in opposite directions. The field effect ρF contributes effectively at all times to dissociating phosphonic and carboxylic acids. This is shown by the data in Table 6 and represented by the histogram in Figure 6.

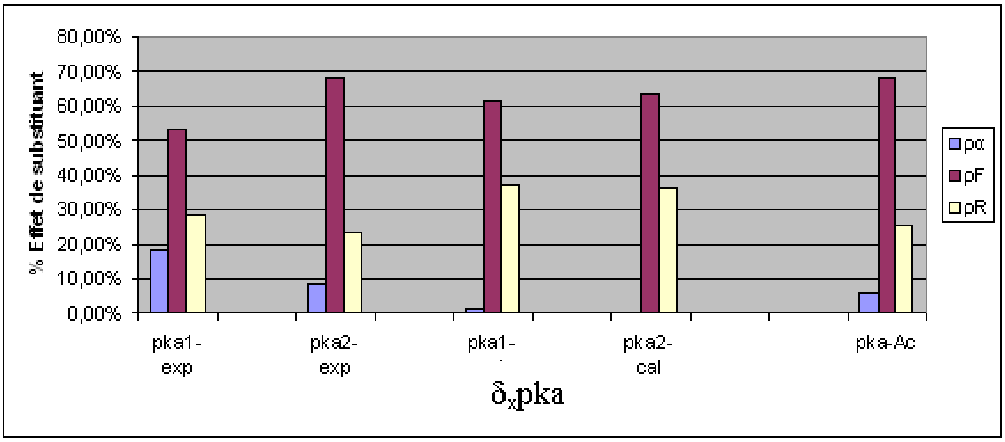

Figure 6: Percentage of polarization (ρα), field (ρF), and resonance (ρR) substituent effects on $\delta_x$pka of phosphonic and carboxylic acids

Table 6: Coefficients ρα, ρF, and ρR of linear regressions and relative percentage of substituent effects on the $\delta_x$pka variation of phosphonic and carboxylic acids.

	XPO(OH)$_2$								**pKa-XCOOH)**	
	pKa1-exp		**pKa2-exp**		**pKa1-cal**		**pKa2-cal**		**pKa-exp**	
ρα	**-1,02**	**18,46%**	**-0,89**	**8,05%**	**-0,42**	**1,48%**	**-0,14**	**0,55%**	**-0,79**	**5,92%**
ρF	**-2,91**	**52,87%**	**-7,55**	**68,43%**	**-17,49**	**61,21%**	**-15,91**	**63,31%**	**-9,13**	**68,37%**
ρR	**-1,58**	**28,66%**	**-2,60**	**23,56%**	**-10,63**	**37,18%**	**-9,09**	**36,16%**	**-3,43**	**25,70%**

The percentage related to the field effect varies between 50 and 70%. This is due to the coulombic interaction between the proton of the hydroxyl group (O-H) and the dipole moment induced at the substituent X by the permanent polarization linked to the electronegativity difference between the two atoms O and H.

The resonance effect also contributes with a relatively large share to the $\delta_x$pka variation with a percentage varying between 20 and 40% compared to the polarizability effect and the field effect. This is related to the 7/23 substituents (30%) possessing an sp2 hybridized carbon atom or heteroatom. The latter transmit their effects to the deprotonation site by conjugation with the π-system of the phosphonic function.

This transfer could also be due to hyper-conjugation of the C-H or C-C bond σ electrons of the alkyl groups. The magnitude of the resonance effect is relatively high in the case of phosphonic acids compared to carboxylic acids. This is probably related to the contribution of the 3d polarization orbitals on the phosphorus atom of similar energy to those of the 3s and 3p valence orbitals of this atom, and the polarization effect is clearly weak compared to the field effect and the resonance effect.

Prediction equations from experimental values:

$$\boldsymbol{\delta_x p\hat{k}a_{1,i}^{exp} = 0.55 - 1.02\sigma_{\alpha,i} - 2.91\sigma_{F,i} - 1.58\sigma_{R,i}}$$

$$\boldsymbol{\delta_x p\hat{k}a_{2,i}^{exp} = 0.66 - 0.89\sigma_{\alpha,i} - 7.55\sigma_{F,i} - 2.6\sigma_{R,i}}$$

From the theoretical values:

$$\boldsymbol{\delta_x p\hat{k}a_{1,i}^{cal} = 0.80 - 0.42\sigma_{\alpha,i} - 17.49\sigma_{F,i} - 10.53\sigma_{R,i}}$$

$$\boldsymbol{\delta_x p\hat{k}a_{2,i}^{cal} = 0.59 - 0.14\sigma_{\alpha,i} - 15.91\sigma_{F,i} - 9.08\sigma_{R,i}}$$

From the experimental values of carboxylic acids:

$$\delta_x p\hat{k}a_{i\_XCOOH} = 0.32 - 0.79\sigma_{\alpha,i} - 9.13\sigma_{F,i} - 3.43\sigma_{R,i}$$

It is noticeable that the deviations are small. The mean of the residuals is less than 0.02 pKa units except for the one corresponding $\hat{\varepsilon}_{2,i}^{exp}$=-0.91, which indicates an overestimation of $p\hat{k}a_{2,i}^{exp}$by the M2 model. To decrease the value of this deviation to -0.25, the value ρ0=0.66, which is insignificant, perhaps neglected.

Overall, Taft's model is successfully applied to the observed (experimental and theoretical) pKa of phosphonic and carboxylic acids in the aqueous phase, which gives confidence to this model to predict the pKa constants of phosphonic acids XPO3H2 and carboxylic acids with tabulated parameters σP, σF and σR of substituent X. Also, it can be used to estimate the parameters σP, σF, and σR for substituents for which the corresponding acid pKa values are available.

The relative constant of a phosphonic acid can be deduced from the following equation:

$$\left.\begin{aligned} &\delta_x p\hat{k}a_{k,i} = p\hat{k}a_{k,i} - pka_{k,H} \\ &= \rho_{0,i} + \rho_\alpha \sigma_{\alpha,i} + \rho_F \sigma_{F,i} + \rho_R \sigma_{R,i} \end{aligned}\right\} \Rightarrow p\hat{k}a_{k,i} = pka_{k,H} + \rho_{0,i} + \rho_\alpha \sigma_{\alpha,i} + \rho_F \sigma_{F,i} + \rho_R \sigma_{R,i}$$

CONCLUSION

The average of the deviations between the pKa values predicted by the two models is less than unity, with a standard deviation of less than 0.4 pKa units for the experimental values. For the theoretical values, the standard deviation reaches 2 pKa units because of the pKa values estimated by the M1 model for the XPO(OH)2 systems with X= O-Me, O-Et, O-nPr, and O-Ph.

The simple and multiple linear regression method allowed the construction of equations that predict the acidity constants pKa of phosphonic acids.

The analysis of statistical parameters and residuals allowed the validation of the linear regressions, which legitimizes our approach of substituent effects on the pKa values of phosphonic acids. Nevertheless, it is important to recall the limitations of these models to the substituents whose Taft parameters are tabulated. Despite these limitations, we now have a simple and efficient method for estimating the acidity constants pKa1 and pKa2 of XPO(OH)2 phosphonic acids, which we can apply with greater confidence intervals than 95%.

REFERENCES

1. [1] R. Bouaddi, L. El Maimouni, and A. El Hammadi, "A theoretical calculation of relative pKas of phosphonic acids in an aqueous solution using idodesmic reaction and the Polarisable Continuum Model," Int. J. Comput. Eng. Res., vol. 9, no. 2, 2019.
2. [2] M. Lew, C. Bagw, L. K. E. Hardebeck, and S. Wireduaah, "8-340-1-Pb," no. April, 2012.
3. [3] W. P. Ozimiński and J. C. Dobrowolski, "σ- and π-electron contributions to the substituent effect: Natural population analysis," J. Phys. Org. Chem., vol. 22, no. 8, pp. 769–778, 2009, doi: 10.1002/poc.1530.
4. [4] H. C. Shands, "Compte rendu," Semiotica, vol. 3, no. 3, pp. 280–287, 1971, doi: 10.1515/semi.1971.3.3.280.
5. [5] B. Cl, "Table A.l comprises an extensive compilation of Hammett."
6. [6] S. Rachuru, A. A. Skelton, and J. Vandanapu, "Application of Hammett equation to hydrogen bond interactions of benzoic acid in chloroform/water system and explanation for non-linear Hammett relation to partition coefficients for the same system," Comput. Theor. Chem., vol. 1190, no. August, p. 113024, 2020, doi: 10.1016/j.comptc.2020.113024.
7. [7] A. Campero and J. A. Diáz Ponce, "Averaged Scale in Electronegativity Joined to Physicochemical Perturbations. Consequences of Periodicity," ACS Omega, vol. 5, no. 40, pp. 25520–25542, 2020, doi: 10.1021/acsomega.0c00256.
8. [8] Y. Xing, J. Ku, W. Fu, L. Wang, and H. Chen, "Inductive effect between atomically dispersed iridium and transition-metal hydroxide nanosheets enables highly efficient oxygen evolution reaction," Chem. Eng. J., vol. 395, no. January, p. 125149, 2020, doi: 10.1016/j.cej.2020.125149.
9. [9] M. Manassir and A. H. Pakiari, "An electronic properties investigation to interpret the substituent constants of monosubstituted benzene derivatives," J. Mol. Graph. Model., vol. 92, pp. 201–207, 2019, doi: 10.1016/j.jmgm.2019.07.017.
10. [10] B. ying Wei, C. tun Cao, and C. zhong Cao, "Influences of polarizability effect of alkyl group and homoring competition effect of substituents on the fluorescence emission spectra of salen-type Schiff bases," J. Phys. Org. Chem., vol. 34, no. 11, pp. 1–8, 2021, doi: 10.1002/poc.4263.
11. [11] S. Herbers, P. Buschmann, J. Wang, K. G. Lengsfeld, K. P. R. Nair, and J. U. Grabow, "Reactivity and rotational spectra: The old concept of substitution effects," Phys. Chem. Chem. Phys., vol. 22, no. 20, pp. 11490–11497, 2020, doi: 10.1039/d0cp01145b.
12. [12] B. A. Saleh, S. A. Al-Shawi, and G. F. Fadhil, "Substituent Effect Study for 13Cβ SCS in a Meta-Styrene Series. Reynolds Dual Substituent Parameter Model Investigation," J. Phys. Chem. A, vol. 125, no. 36, pp. 7759–7768, 2021, doi: 10.1021/acs.jpca.1c02645.
13. [13] M. G. Zahl, R. Fossheim, K. J. Børve, L. J. Sæthre, and T. D. Thomas, "Electronic Properties of Chlorine, Methyl, and Chloromethyl as Substituents to the Ethylene Group-Viewed from the Core of Carbon," J. Phys. Chem. A, vol. 119, no. 36, pp. 9481–9493, 2015, doi: 10.1021/acs.jpca.5b05494.
14. [14] C. A. Jinfeng, R. D. Topsom, U. S. A. Ca, and A. G. Best, "ACIDITIES OF SUBSTITUTED," vol. 168, pp. 141–146, 1988.
15. [15] J. Palomar and J. L. G. De Paz, "Protonation study of some enamine systems," vol. 541, pp. 111–117, 2001.
16. [16] F. The, "AMERICAN CHEMICAL," vol. 112, no. 6, pp. 2047–2052, 1990.
17. [17] P. Taylor et al., "Critical Reviews in Analytical Chemistry The Correlation Coefficient : An Overview The Correlation Coefficient : An Overview," no. September 2013, pp. 37–41, 2007, doi: 10.1080/10408340500526766.
18. [18] L. D. Freedliaru, A. S. D. G. Dom, and N. Carolina, "ACIDS," 1956.
19. [19] E. Bosch, K. H. A. H. A.- Y, and K. B. S.-B. Oh-, "IONIC EQUILIBRIA IN NEUTRAL AMPHIPROTIC SOLVENTS ; RESOLUTION OF ACID STRENGTH IN tert -BUTY L ALCOHOL [H + l [A-I P + I @ -1 + W-1)," vol. 36, no. 6, 1989.
20. [20] "The Acid Dissociation Constants of Aromatic Phosphonic Acids. Ortho Substituents," vol. 540, no. 11, pp. 257–258, 1946.
21. [21] Z. Jinhua, T. Kleinöder, and J. Gasteiger, "Prediction of pKa values for aliphatic carboxylic acids and alcohols with empirical atomic charge descriptors," J. Chem. Inf. Model., vol. 46, no. 6, pp. 2256–2266, 2006, doi: 10.1021/ci060129d.