

## Quantitative Structure-Property Relationship (QSPR) modeling of the second reduction potential of a family of Tetracyanoquinodimethane (TCNQ) molecules using descriptors of quantum chemistry

Fatogoma DIARRASSOUBA<sup>1</sup>, KafoumbaBAMBA<sup>1</sup>, Mawa KONÉ<sup>2</sup>,  
AhissanDonatien EHOUMAN<sup>1</sup>, Koffi Kan Raymond KOUAMÉ<sup>1</sup> and Nahossé ZIAO<sup>1</sup>

<sup>1</sup>Laboratoire de Thermodynamique et Physico-Chimie du Milieu, UFR SFA, Université Nangui Abrogoua 02 BP 801 Abidjan 02, Côte-d'Ivoire

<sup>2</sup>Laboratoire de Chimie Organique et de Substances Naturelles, UFR-SSMT, Université Félix Houphouët Boigny  
22 BP 582 Abidjan 22, République de Côte-d'Ivoire

### ABSTRACT:

The general objective of this work was to model the second reduction potential of a series of Tetracyanoquinodimethane (TCNQ) molecules. With a database of twenty-eight (28) analogous TCNQ molecules, twenty molecules or about 2/3 of the database were used for the training set and eight molecules or about 1/3 of the database were used for the test set. The molecular descriptors were calculated after optimization and frequency calculation using the DFT method in 6-31G(d,p) basis set. Using statistical analysis methods, a QSPR model for predicting the second reduction potential dependent on the electrophilic index only was developed. The statistical and validation parameters linked to this model were determined and considered interesting. These various parameters revealed that this model can be used to effectively predict the second reduction potential of future TCNQ of this same family belonging to its domain of applicability at the 95% confidence level.

**Keywords:** Tetracyanoquinodimethane, Second reduction potential, Statistical analysis.

Date of Submission: 04-09-2020

Date of Acceptance: 19-09-2020

### I. INTRODUCTION

Coulomb intramolecular repulsion of electron donors and acceptors is a phenomenon that can prevent charge complexes from being excellent electrical conductors. Minimal Coulomb repulsion is required to reach a metallic state. Theoretical studies have suggested that extending the TCNQ  $\pi$ -system lowers Coulomb's intramolecular repulsion and thus maintains a metallic state in organic systems [1]. With regard to the TCNQ  $\pi$ -system in particular, the degree of this repulsion can be predicted based on the difference between the potential of first reduction and that of second reduction. Indeed, the difference between the first and second reduction potentials ( $\Delta E = E^1 - E^2$ ) is a measure of Coulomb repulsion. Here smaller  $\Delta E$  values are desired as they increase the electrical conductivity of the organic charge complex. However, one of the major concerns in the synthesis of organic charge complexes is that organic chemists are faced with the development of

thermodynamically stable radical systems. Which is not an easy task. Hence, it requires alternative methods to experimentation. Among these methods, the Quantitative Structure-Property Relationship (QSPR) has become of great interest. It allows the physicochemical properties of a family of given molecules to be linked to their molecular structures. We can then explain the origin of these properties and predict them for other molecules of this same family whose experimental data are non-existent. QSPR models predicting the first reduction potential of the tetracyanoquinodimethane family have already been developed. They have even been the subject of publications by Peter W. Kenny [2] and F. Diarrassouba et al. [3]. However, up to date, there does not appear to be any study on QSPR modeling of the second reduction potential of this same family of molecules. Therefore, the main objective in this work is to develop a QSPR model predictive of the second reduction potential using quantum chemistry

descriptors based on an experimental database of twenty-eight (28) derivatives of Tetracyanoquinodimethane.

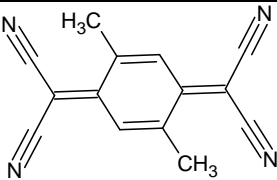
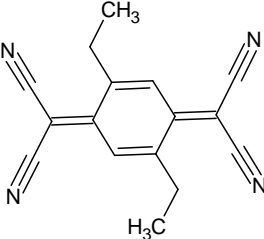
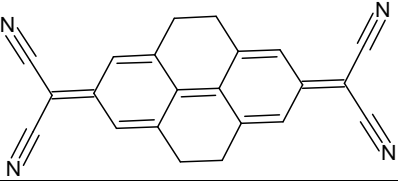
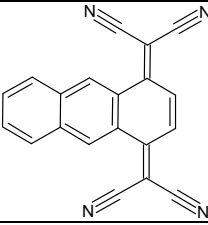
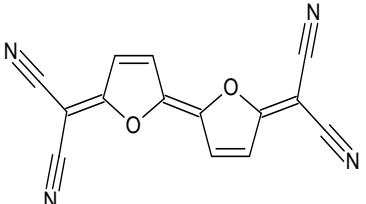
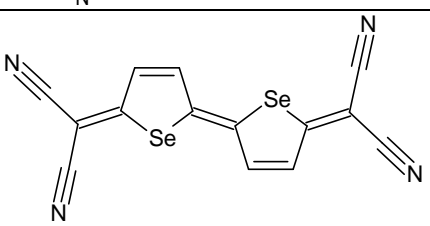
## II. MATERIAL AND METHODS

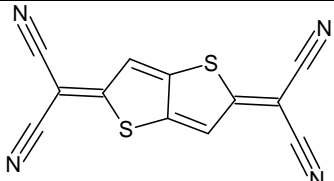
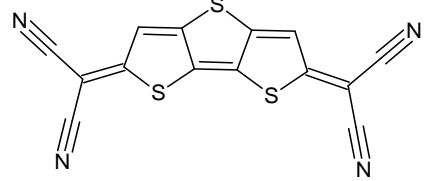
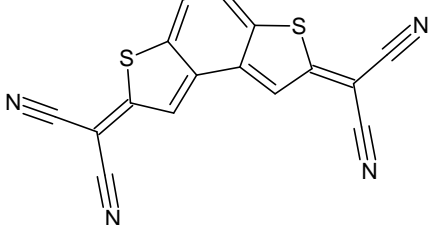
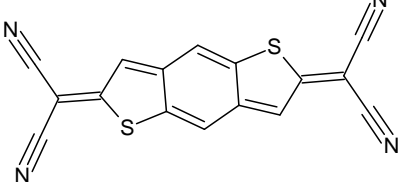
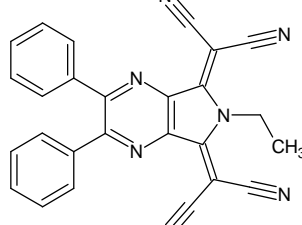
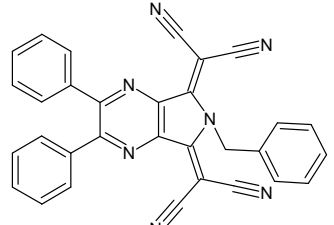
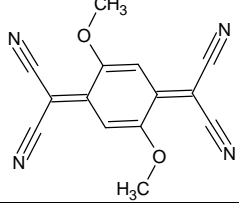
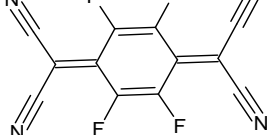
### 2.1. Database

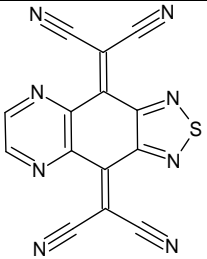
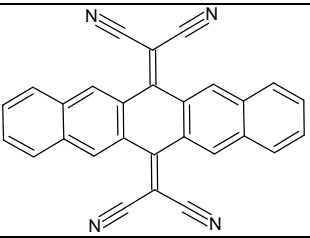
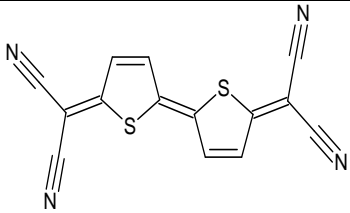
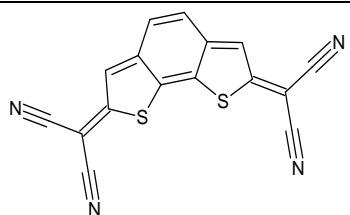
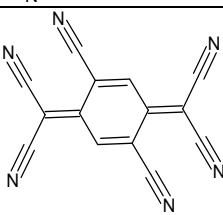
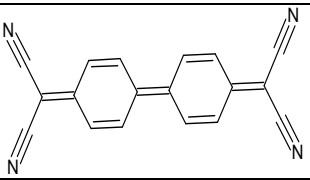
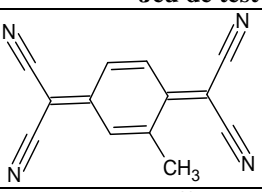
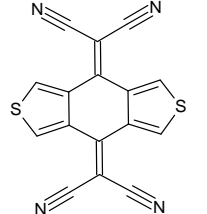
In the development of the predictive QSPR model of the second reduction potential, we considered a series of twenty-eight Tetracyanoquinodimethane derivatives codified TCNQ [4-11]. The choice of these molecules is due

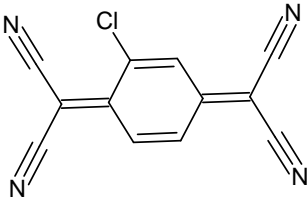
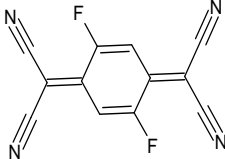
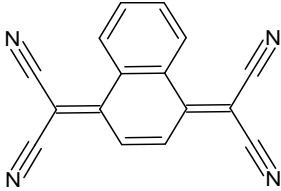
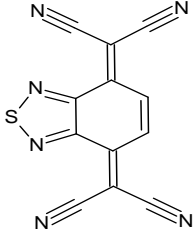
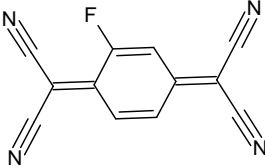
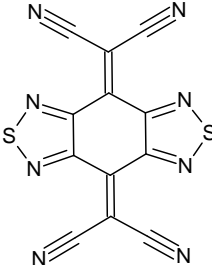
to the availability of their experimental second reduction potentials. These properties have been all determined by cyclic voltammetry in acetonitrile. These molecules have constituted our database. Twenty of which (or about 2/3 of the database) were used for the training set and eight molecules (or about 1/3 of the database) were used for the test set. **Table 1** presents these different molecules with their corresponding experimental second reduction potentials expressed in volts (V).

Table 1. Series of studied molecules

Training set			
Code[3]	Molecule	$E_{exp}^2$ (V)	Reference
TCNQ_2		-0.350	[4]
TCNQ_3		-0.365	[4]
TCNQ_4		-0.214	[5]
TCNQ_5		-0.480	[5]
TCNQ_9		-0.310	[5]
TCNQ_12		-0.250	[5]

TCNQ_13		-0.355	[5]
TCNQ_14		-0.288	[5]
TCNQ_15		0.090	[5]
TCNQ_16		-0.050	[5]
TCNQ_18		-0.530	[5]
TCNQ_19		-0.490	[5]
TCNQ_22		-0.470	[6]
TCNQ_23		0.020	[6]

TCNQ_24		-0.460	[10]
TCNQ_28		-0.910	[5]
TCNQ_30		-0.259	[5]
TCNQ_31		-0.010	[5]
TCNQ_34		0.090	[5]
TCNQ_36		-0.030	[5]
<b>Jeu de test</b>			
TCNQ_1		-0.340	[4]
TCNQ_8		-0.730	[5]

TCNQ_20		-0.220	[6]
TCNQ_21		-0.144	[8]
TCNQ_27		-0.410	[4]
TCNQ_35		-0.380	[11]
TCNQ_39		-0.185	[7]
TCNQ_40		-0.490	[9]

## 2.2. Computational Theory Level and Softwares

GaussView 5.0[12] software was used to represent the 3D structure and visualize the studied molecules. Then, Gaussian 09 software[13] was used for optimization and frequency calculation (Temperature 298.15 Kelvin, pressure 1 atmosphere, in vacuum). The theory level used is B3LYP/6-31G(d,p). As for 2D structures, they have been represented with chemsketch[14]. EXCEL[15] and XLSTAT[16] softwares were used for graphic representation and modeling respectively. For the calculation of the observation levers, the minitab 18[17] software was used.

## 2.3. Statistical analysis

To develop a QSPR model, a data analysis method is required. This method quantifies the relationship between the studied property and the molecular structure (descriptors). There are several methods to build a model and analyze its statistical data. But the method we used in our study is Simple Linear Regression (SLR) (a single explanatory variable). Generally speaking, the equation of the simple regression is of the form:

$$Y = a_0 + a_1X \quad (1)$$

with Y: studied property;

X: explanatory variable in correlation with the studied property;

$a_0, a_1$ : model regression constants.

The selection of descriptors is a crucial step in QSPR modeling. In this study, the selection of

descriptors was based on two criteria described as follows:

- **Criterion 1**

There must be a linear dependence relationship between the second reduction potential and the descriptors. Under these conditions,  $|R| \geq 0.50$  [18]. R stands for the linear correlation coefficient of the line  $E_{exp} = f(\text{Descripteur}_i)$ .

- **Criterion 2**

The descriptors must be independent of each other. To do this, the partial correlation coefficient  $a_{ij}$  between the descriptors i and j must be less than 0.70 ( $a_{ij} < 0,70$ ) [18]. For a multilinear regression, the coefficients R and  $a_{ij}$  are expressed as follows:

$$R = \frac{\text{COV}(X,Y)}{S_X \cdot S_Y} \quad (2) \quad \text{and} \quad a_{ij} = \frac{\text{COV}(X_i, X_j)}{\text{Var}(X_i)} \quad (3)$$

The relationships 4, 5, 6 and 7 were used to calculate many statistical and validation parameters:

$$ESS = \sum (Y_{i,cal} - \bar{Y}_{exp})^2 \quad (4)$$

$$TSS = \sum (Y_{i,exp} - \bar{Y}_{exp})^2 \quad (5)$$

$$RSS = \sum (Y_{i,exp} - Y_{i,cal})^2 \quad (6)$$

$$TSS = ESS + RSS \quad (7)$$

where TSS: Total Sum of Squares; ESS: Extended Sum of Squares; RSS: Residual Sum of Squares.

- **Determination coefficient  $R^2$  [19]**

The determination coefficient is given by the following relationship:

$$R^2 = 1 - \frac{\sum (Y_{i,exp} - Y_{i,cal})^2}{\sum (Y_{i,exp} - \bar{Y}_{exp})^2} = 1 - \frac{RSS}{TSS} \quad (8)$$

$$\text{with } R = \sqrt{\frac{\sum (Y_{i,cal} - \bar{Y}_{exp})^2}{\sum (Y_{i,exp} - \bar{Y}_{exp})^2}} = \sqrt{\frac{ESS}{TSS}} \quad (9)$$

- **Standard deviation S [20]**

It is an indicator of dispersion. It provides information on how the distribution of data is spread around the average. The closer its value is to 0, the better the adjustment and the more reliable the prediction will be.

$$S = \sqrt{\frac{\sum (Y_{i,exp} - Y_{i,cal})^2}{n-p-1}} = \sqrt{\frac{RSS}{n-p-1}} \quad (10)$$

- **Adjusted determination coefficient ( $R^2_{adjust}$ ) [21]**

It allows to measure the robustness of a model unlike  $R^2$ . This coefficient is used in multiple regressions because it takes into account the number of parameters (descriptors) of the model.

$$R^2_{adjust} = 1 - \frac{(n-Intercept)}{n-p-1} \cdot \frac{RSS}{TSS} = 1 - \frac{(n-Intercept)}{n-p-1} \cdot (1 - R^2) \quad (11)$$

- **Fisher-Snedecor coefficient (F) [22]**

It allows to test the global significance of linear regression. The Fisher-Snedecor coefficient is related to the determination coefficient by the following relationship:

$$F = \frac{n-p-1}{p} \cdot \frac{ESS}{RSS} = \frac{n-p-1}{p} \cdot \frac{R^2}{1-R^2} \quad (12)$$

- **Kubinyi Criterion (FIT) [23]**

It measures the size or robustness of the model. The smaller the FIT, the more robust the model is, i.e. the model has more variables.

$$FIT = \frac{(n-p-1)}{(n+p^2)} \cdot \frac{R^2}{(1-R^2)} \quad (13)$$

- **Cross-validation coefficient ( $Q^2_{LOO}$ ) [24]**

It measures the accuracy of the prediction on the data of the training set.

$$Q^2_{LOO} = 1 - \frac{\sum (y_{i,exp} - y_{i,pred})^2}{\sum (y_{i,exp} - \bar{y}_{exp})^2} = 1 - \frac{PRESS}{TSS} \quad (14)$$

- **Cross-validation criteria (PRESS) [24]**

The sum of the quadratic prediction errors, PRESS (*Prediction Sum of Squares*) is defined by the relationship:

$$PRESS = \sum (y_{i,exp} - y_{i,pred})^2 \quad (15)$$

This criterion is used to select models with good predictive power. (we always look for the smallest PRESS). A standard deviation of error of prediction (SDEP) is calculated from PRESS:

$$SDEP = \sqrt{\frac{\sum (y_{i,exp} - y_{i,pred})^2}{n}} = \sqrt{\frac{PRESS}{n}} \quad (16)$$

In these expressions,  $n$  is the number of molecules in the training set,  $p$  is the number of explanatory variables.  $y_{i,\text{exp}}$  and  $y_{i,\text{pred}}$  are respectively the experimental and predicted values of property for molecule  $i$  and  $\bar{y}_{\text{exp}}$  is average value of the property for the training set.

- **Todeschini's parameter ( ${}^cR_p^2$ )[25]**

${}^cR_p^2$  is the corrected form of P.P. Roy's parameter noted  $R_p^2$ [26]. It allows to know if the model is due to chance correlations or not. If this parameter is greater than 0.50, the model is assumed not due to a chance correlations.

$${}^cR_p^2 = R\sqrt{R^2 - R_r^2} \quad (17)$$

with  $R_r^2$ , the average value of  $R_{ri}^2$  of the models obtained with the randomized property.

- **External validation coefficient ( $Q_{\text{ext}}^2$ )[27]**

It measures the accuracy of the prediction on the test set data.

$$Q_{\text{ext}}^2 = 1 - \frac{n}{n_{\text{ext}}} \frac{\text{PRESS}(\text{test})}{\text{TSS}} \quad (18)$$

here,  $n_{\text{ext}}$  refers to the number of test set compounds

- **Parameter (RMSEP)[27]**

External predictive ability of QSPR model may further be determined by root mean square error in prediction given by:

$$\text{RMSEP} = \sqrt{\frac{\sum (y_{\text{exp}}(\text{test}) - y_{\text{pred}}(\text{test}))^2}{n_{\text{ext}}}} \quad (19)$$

- **Roy K. and al. parameters ( $\bar{r}_m^2$  and  $\Delta r_m^2$ )[28]**

For the acceptable prediction, the value of  $\Delta r_m^2$  should preferably be lower than 0.20 when the value of  $\bar{r}_m^2$  is more than 0.50.

$$\bar{r}_m^2 = \frac{(r_m^2 + r'_m{}^2)}{2} \quad (20)$$

$$\Delta r_m^2 = |r_m^2 - r'_m{}^2| \quad (21)$$

$$\text{herer}_m^2 = r^2(1 - \sqrt{r^2 - r_0^2}) \quad (22) \quad \text{and } \text{dr}'_m^2 = r^2(1 - \sqrt{r^2 - r_0^2}) \quad (23)$$

The parameters  $r^2$  and  $r_0^2$  are the determination coefficients between the observed and predicted values of the compounds (training set or test set) with and without intercept, respectively. The parameter  $r_0^2$  bears the same meaning but uses the reversed axes.

- **External validation criteria or "Tropsha's criteria"[24][29].**

There are five such criteria:

❖ Criterion 1:  $R_{\text{ext}}^2 > 0.70$

❖ Criterion 2:  $Q_{\text{ext}}^2 > 0.60$

❖ Criterion 3:  $\frac{|R_{\text{ext}}^2 - R_0^2|}{R_{\text{ext}}^2} < 0.1$  and  $0.85 < k < 1.15$

❖ Criterion 4:  $\frac{|R_{\text{ext}}^2 - R_0^2|}{R_0^2} < 0.1$  and  $0.85 < k' < 1.15$

❖ Criterion 5:  $|R_{\text{ext}}^2 - R_0^2| < 0.3$

where,  $R_{\text{ext}}^2$ : determination coefficient of molecules for the test set;  $R_0^2$ : determination coefficient of the regression between predicted and experimental values for the test set without intercept;  $R_0^2$ : determination coefficient of the regression between experimental and predicted values for the test set without intercept;  $k$ : slope of the correlation line (values predicted according to the experimental values with intercept = 0);  $k'$ : slope of the correlation line (experimental values according to the predicted values with intercept = 0). Ouanlo Ouattara et al.[30] reported that if at least 3/5 of the Tropsha's criteria are verified, then the QSPR model developed is considered a successful model in predicting of the studied property.

- **Lever ( $h_{ij}$ )[31]**

The lever is a kind of distance from the barycentre of the points in the space of the explanatory variables. It identifies observations that are abnormally far from others. For observation  $i$

$$h_{ij} = x_i(X^T X)^{-1} x_i^T \quad (i=1, \dots, n) \quad (24)$$

where  $x_i$  is the line vector of the descriptors of compound  $i$  and  $X$  is the matrix of the model derived from the values of the descriptors of the training set. The index T refers to the transposed matrix/vector. The critical value of lever  $h^*$  is, in general, set to  $\frac{3(p+1)}{n}$ [32], where  $n$  is the number of compounds in the training set and  $p$  is the

number of model descriptors. If a compound has a residual and a lever that exceeds the critical value  $h^*$ , this compound is considered outside the applicability domain of the developed model.

#### 2.4. Calculated molecular descriptors

As part of our work, six (06) descriptors were calculated. These are: electronic energy ( $E_T$ ), dipole moment ( $\mu_D$ ), ionization potential ( $IP$ ), electronic affinity ( $EA$ ), electrophile index ( $\omega$ ) and the sum of the absolute value of the Mulliken charges ( $Q$ ). **Table 1** presents the different approaches to these descriptors.

**Table 1.** List of molecular descriptors used

Molecular descriptors	Notation	Expressions
Electronic energy	$E_T$	
Dipole moment	$\mu_D$	
Ionization potential[33]	$IP$	$IP = -E_{HOMO}$
Electronic affinity[33]	$EA$	$EA = -E_{LUMO}$
Electrophilic index[34]	$\omega$	$\omega = \frac{\chi^2}{2\eta}$
Sum of the absolute value of the Mulliken charges[30]	$Q$	

with  $\eta = \frac{IP - EA}{2}$ [35]  $\chi = \frac{IP + EA}{2}$ [36]

### III. RESULTS AND DISCUSSION

#### 3.1. Values of calculated molecular descriptors

The values of the six molecular descriptors are summarized in **Table 2**.

**Table 2.** Values of calculated descriptors

Training set						
COMPOUND	$E_T$	$\mu_D$	$IP$	$EA$	$\omega$	$Q$
TCNQ_1	-757.2186	0.0007	7.1298	4.6219	13.7670	5.8730
TCNQ_2	-835.8476	0.0014	7.0645	4.5302	13.2619	6.7581
TCNQ_3	-1064.5157	0.0036	6.2638	4.5376	16.8964	7.2240
TCNQ_4	-985.8761	6.1457	6.8895	4.2251	11.5914	5.6736
TCNQ_5	-905.2187	0.0008	6.6228	4.4380	13.9988	6.6582
TCNQ_6	-5553.5498	0.0003	6.5370	4.4094	14.0793	4.4161
TCNQ_7	-1473.7483	0.0014	7.0441	4.5582	13.5381	4.7414
TCNQ_8	-1948.1462	3.4780	6.6484	4.5253	14.7019	5.3336
TCNQ_9	-1627.3871	6.9138	6.3604	4.8269	20.4026	5.4517
TCNQ_10	-1627.3934	0.0008	6.6043	4.7253	17.0789	5.5583
TCNQ_11	-1382.9963	5.5556	6.6141	3.8297	9.7933	9.7223
TCNQ_12	-1574.7312	5.5486	6.5966	3.8085	9.7076	10.4844
TCNQ_13	-907.6375	0.0018	6.8465	4.3152	12.3044	6.9352
TCNQ_14	-1075.4622	0.0016	7.6118	5.2544	17.5554	5.3776
TCNQ_15	-1370.7582	2.0047	7.4634	4.4573	11.8177	7.6879
TCNQ_16	-1293.1709	8.3638	6.6323	3.3972	7.7736	6.9336
TCNQ_17	-1551.1661	0.0001	6.5858	4.4714	14.4562	2.4880
TCNQ_18	-1627.3928	6.1661	6.5087	4.7153	17.5607	5.6189
TCNQ_19	-863.0285	0.0006	7.9057	5.5402	19.1068	5.5363
TCNQ_20	-909.6358	0.0018	6.3906	4.8048	19.7592	5.3941

Test set



TCNQ_21	-717.9018	1.2456	7.2245	4.7196	14.2385	5.0988
TCNQ_22	-1627.3872	4.4817	7.2533	3.7080	8.4726	6.9263
TCNQ_23	-1138.1647	0.4703	7.4028	4.9545	15.5928	4.3909
TCNQ_24	-877.0332	0.0012	7.4926	5.0348	15.9630	5.1691
TCNQ_25	-832.2260	3.7464	7.2008	4.5060	12.7141	5.2247
TCNQ_26	-1185.0515	0.4964	7.4885	4.7109	13.3953	6.2528
TCNQ_27	-777.8097	0.7489	7.4060	4.9333	15.3935	4.7732
TCNQ_28	-1691.5156	0.0015	7.6401	4.4605	11.5131	8.0094

### 3.2. Selection of molecular descriptors

#### 3.2.1. Submission of descriptors to selection criterion 1

The molecular descriptors calculated were subjected to selection criterion 1 which stipulates that there must be a linear dependence relationship between the second potential reduction and the molecular descriptors.

**Table 3.** Submission of molecular descriptors to selection criterion 1

Equation	Correlation coefficient $ R $	Descriptor rejected if $ R  \geq 0.50$
$E_{\text{exp}}^2 = f(E_T)$	0.0413	rejected
$E_{\text{exp}}^2 = f(\mu_D)$	0.3599	rejected
$E_{\text{exp}}^2 = f(IP)$	0.1004	rejected
$E_{\text{exp}}^2 = f(EA)$	<b>0.8952</b>	<b>Retained</b>
$E_{\text{exp}}^2 = f(\omega)$	<b>0.9592</b>	<b>Retained</b>
$E_{\text{exp}}^2 = f(Q)$	0.4548	rejected

Examination of the data in **Table 3** shows that there is a linear dependence relationship between the second reduction potential and the descriptors EA and  $\omega$  since their linear correlation coefficient exceeds 0.50. Thus, these two descriptors will be subject to selection criterion 2.

#### 3.2.2. Submission of molecular descriptors retained for selection criterion 2

With regard to selection criterion 2, it stipulates that the molecular descriptors selected according to criterion 1 must be independent of each other in pairs.

**Table 4.** Submission of molecular descriptors to selection criterion 2

Correlation between	Coefficient $a_{ij}$	Independent descriptors $a_{ij} < 0.70$
$\omega$ et EA	6.3859	dependent

Note in **Table 4** that the partial correlation coefficient between the electronic affinity (EA) and the electrophilic index ( $\omega$ ) is much greater than 0.70. It is therefore clear that these two descriptors are dependent. These cannot exist in the same QSPR model. However, the electrophilic index which displays the highest linear correlation coefficient ( $|R| = 0.9592$ ) with the property studied, will be used to establish a QSPR model with a single descriptor.

### 3.3. Study of the developed model

In the QSPR developed model, the potential of second theoretical reduction

( $E_{\text{theo}}^2$ ) is dependent on the electrophilic index ( $\omega$ ) whose regression equation is:

$$E_{\text{theo}}^2 = -1.2799 + 0.0691 * \omega$$

The parameters regarding this equation are:

$n = 20$  ;  $R = 0.9592$  ;  $R^2 = 0.9201$  ;  $R_{\text{adjusted}}^2 = 0.9157$  ;  $S = 0.0726$  ;  $F = 207.2814$  ;

$\text{FIT} = 9.8705$ ;  $p - \text{value} < 0.0001$ ;  $\text{TSS} = 1.1885$ ;  $\text{ESS} = 1.0935$ ;  $\alpha = 95\%$

	t	p-value
Constant	-17,9626	< 0,0001
$\omega$	14,3973	< 0,0001

The sign of the coefficient of the explanatory variable  $\omega$  is positive. This reflects the existence of a direct correlation between the explanatory variable ( $\omega$ ) and the dependent variable (second reduction potential). A compound with a high electrophilic index will necessarily have a high second reduction potential. Regarding the p-value, we see that its value is less than 0.001. Thus, the original constant and the explanatory variable have a highly significant influence on the second reduction potential. This high significance of the original constant and the explanatory variable ( $\omega$ ) is even confirmed by the very high value of the absolute value of the Student

coefficient which is much higher than the critical value ( $t_{\text{critical}}=2,10$ )[37].

### 3.3.1. Internal validation of the model

In the context of internal validation, the Leave-One-Out (LOO) cross-validation technique and the Y-randomization test were used.

- **Leave-One-Out (LOO) cross validation**

This validation is carried out by omitting a molecule in each cycle. The statistical parameters related to the Leave-One-Out cross-validation are listed in **Table 5**.

**Table 5.** Statistical parameters of the LOO cross validation of the model

n	$Q_{\text{LOO}}^2$	$\bar{r}_m^2(\text{LOO})$	$\Delta r_m^2(\text{LOO})$	PRESS	SDEP
20	0.8850	0.8445	0.0255	0.1366	0.0827

The data in **Table 5** indicate that the value of  $Q_{\text{LOO}}^2$  is satisfactory because  $Q_{\text{LOO}}^2 > 0.5$ [38]. This result also shows that our QSPR model is not too sensitive to this operation of removing a molecule and putting it back into the training set. In addition, out of 100 molecules in the training set, 88.50 have their redox potentials predicted. The established model therefore has a high predictive capacity towards the molecules of the training set. This high predictive capacity is even confirmed by the low value of the SDEP (0.0827) which tends

towards 0. This indicates that there is a good agreement between the observed values and the predicted values by the model.

- **Model Y-randomization test**

To ensure that the developed QSPR model was not due to chance, we looked at the property of the randomization test. The calculated average statistical indicators are summarized in **Table 6**.

**Table 6.** Average values of the model randomization parameters

Randomizationparameter	$R_r^2$	$S_r$	$F_r$	${}^cR_p^2$
Average value	0.0682	0.2479	1.3957	0.8853

According to the data in Table 6, the mean value of the coefficients of determination of the randomized models is very low ( $R_r^2 = 0.0682$ ). This value means that on average, only 6.82% of experimental variance of second reduction potential is explained by the descriptor of the randomized model. As for the average standard deviation of the randomized models, we note a very high value compared to that of the non-randomized model. This indicates a strong dispersion of the points around the mean value. Whereas the mean Fisher coefficient  $F_r = 1.3957$ , it is low compared to the significance level. Under these conditions, the regression equation of the randomized model is not

significant. In view of the value of the Todeschini parameter, we see that its value is greater than 0.50 ( ${}^cR_p^2 = 0.8853$ ). It is therefore clear that our established QSPR model is not due to chance correlations.

### 3.3.2. External validation of the model

External validation always concerns the molecules of the test set. It groups the molecules that were not used to build the model. **Table 7** reports the statistical parameters of the external validation of the model.

- **External validation parameters of the model**

**Table 7.** Statistical parameters of the external validation of the model

$n_{\text{ext}}$	$R_{\text{ext}}^2$	$Q_{\text{ext}}^2$	$\bar{r}_m^2(\text{test})$	$\Delta r_m^2(\text{test})$	PRESS (test)	RMSEP
8	0.9839	0.9867	0.9251	0.0086	0.0063	0.0281

From the data in **Table 7**, it is clear that the established model has a very high predictive power because  $Q_{\text{ext}}^2 = 0.9867$ . This shows that out of 100 molecules in the test set, 98.67 have their second reduction potentials predicted by the model. Also, 98.39% of the experimental variance of the second reduction potential is explained by the descriptor of the model. To judge the performance of the QSPR model developed, other criteria were also checked. These are the external validation criteria or Tropsha criteria. Verification of the five (05) criteria of Tropsha.

**Criterion 1:**  $R_{\text{ext}}^2 = 0.9839 > 0.70$

**Criterion 2:**  $Q_{\text{ext}}^2 = 0.9867 > 0.60$

**Criterion 3:**  $\frac{|R_{\text{ext}}^2 - R_0^2|}{R_{\text{ext}}^2} = 0.0048 < 0.1$  and  $k=0.9617$  with  $0.85 < k < 1.015$

**Criterion 4:**  $\frac{|R_{\text{ext}}^2 - R_0^2|}{R_{\text{ext}}^2} = 0.0026 < 0.1$  and  $k'=1.0360$  with  $0.85 < k < 1.015$

**Criterion 5:**  $|R_{\text{ext}}^2 - R_0^2| = 0.0047 < 0.3$

We can see that all five (05) criteria of Tropsha are verified. Consequently, the model performs very well in predicting the second potential reduction.

- **Comparison between observed values and values predicted by the QSPR model from the ratio**

$$\tau = \frac{E_{\text{theo}}^2}{E_{\text{exp}}^2}$$

The performance of the model was also studied by comparing the ratio  $\tau = \frac{E_{\text{theo}}^2}{E_{\text{exp}}^2}$  of the test set. The values

obtained are collated in **Table 8**.

Values of the ratio  $\tau = \frac{E_{\text{theo}}^2}{E_{\text{exp}}^2}$

CODE	$E_{\text{exp}}$	$E_{\text{theo}}$	$\tau$
TCNQ_21	-0.3400	-0.2962	0.8711
TCNQ_22	-0.7300	-0.6945	0.9514
TCNQ_23	-0.2200	-0.2026	0.9210
TCNQ_24	-0.1440	-0.1770	1.2295
TCNQ_25	-0.4100	-0.4015	0.9793
TCNQ_26	-0.3800	-0.3544	0.9327
TCNQ_27	-0.1850	-0.2164	1.1696
TCNQ_28	-0.4900	-0.4845	0.9887

In this table, we note that the values of the ratio  $\tau = \frac{E_{\text{theo}}^2}{E_{\text{exp}}^2}$  are substantially equal to the unit. This shows

that the model is very efficient in predicting the redox potential of the series of studied molecules and can be used to predict the second reduction potential of other TCNQ molecules of this same family.

- **Comparison between observed values and values predicted by the QSPR model from the regression curve**

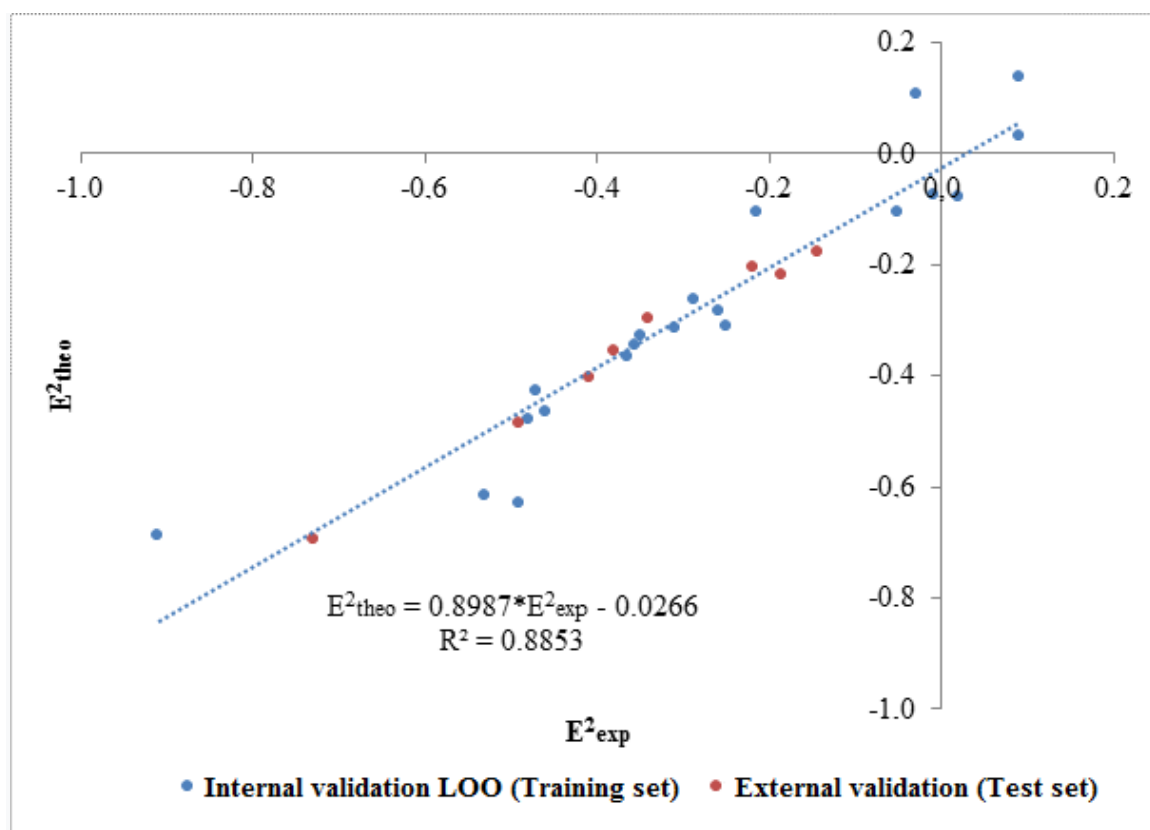


Figure 1. Dispersion diagram of the established model

In **Figure 1**, we can see that the points tend to approach the regression line. This observation reveals that there is a good agreement between the observed values and the predicted values by the developed model. With regard to **Figure 2**, the curves of the observed values and of the theoretical

values evolve in a similar manner since there is an almost superposition of these curves in particular for the test set. This confirms that there is a good agreement between the observed values and the predicted values by the model.

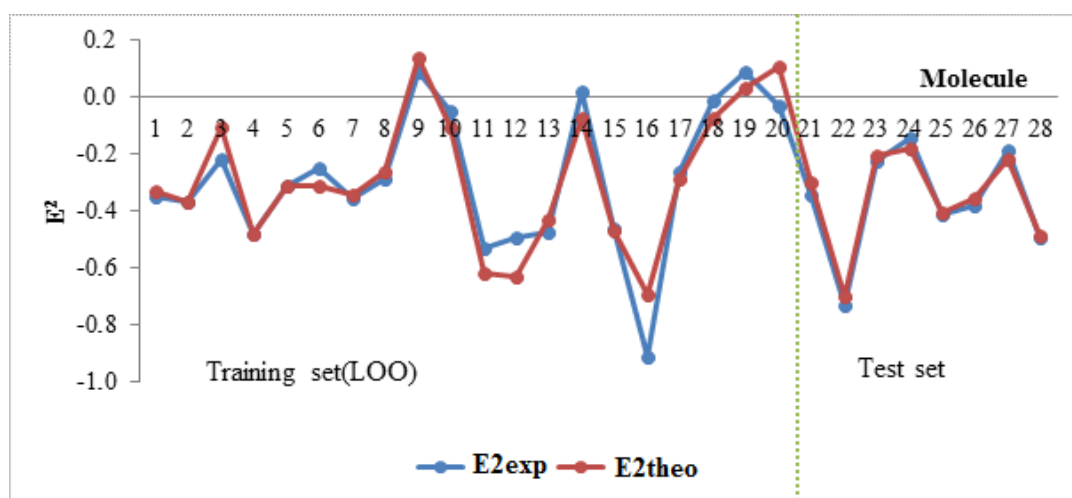


Figure 2. Similarity curves between predicted values by the model and observed values

### 3.4. Statistical tests

#### 3.4.1. Shapiro-Wilk test [39]

**Table 9.** Shapiro-Wilk test parameters

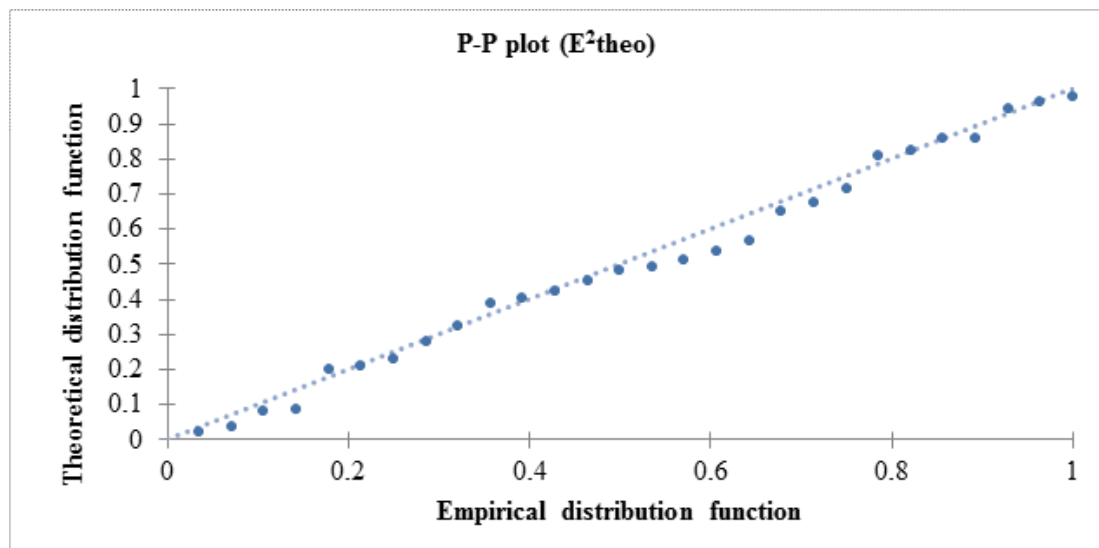
<b>W</b>	0.9837
<b>p-value</b>	0.9271
<b>1 - <math>\alpha</math></b>	0.05

#### Interpretation of the test

**H<sub>0</sub>:** The variable from which the sample comes follows a Normal law

**H<sub>a</sub>:** The variable from which the sample comes does not follow a Normal law  
 Since the calculated p-value is greater than the threshold significance level  $1-\alpha = 0.05$ , we cannot

reject the null hypothesis H<sub>0</sub>. The risk of rejecting the null hypothesis H<sub>0</sub> when it is true represents 92.71%. Thus, the sample follows a normal law. This normal distribution is even confirmed by the distribution of the points along the main diagonal (**Figure 3**).



**Figure 3.** P-P plot graph (E<sup>2</sup><sub>exp</sub>) of the model

#### 3.4.2. Durbin-Watson test[40]

**Table 10.** Parameters of the Durbin-Watson test

<b>U</b>	1.9339
<b>p-value</b>	0.4317
<b><math>\alpha</math></b>	0.05

#### Interpretation of the test

**H<sub>0</sub>:** Residues are not autocorrelated

**H<sub>a</sub>:** The residues describe an AR (1) process (the residues are autocorrelated)

Since the calculated p-value is greater than the threshold significance level  $1-\alpha = 0.05$ , we cannot reject the null hypothesis H<sub>0</sub>. The risk of rejecting

the null hypothesis H<sub>0</sub> when it is true is 43.17%. Therefore, the residuals are not autocorrelated and do not contain any information that could influence the prediction. This is confirmed by the random distribution of points in **Figure 4**.

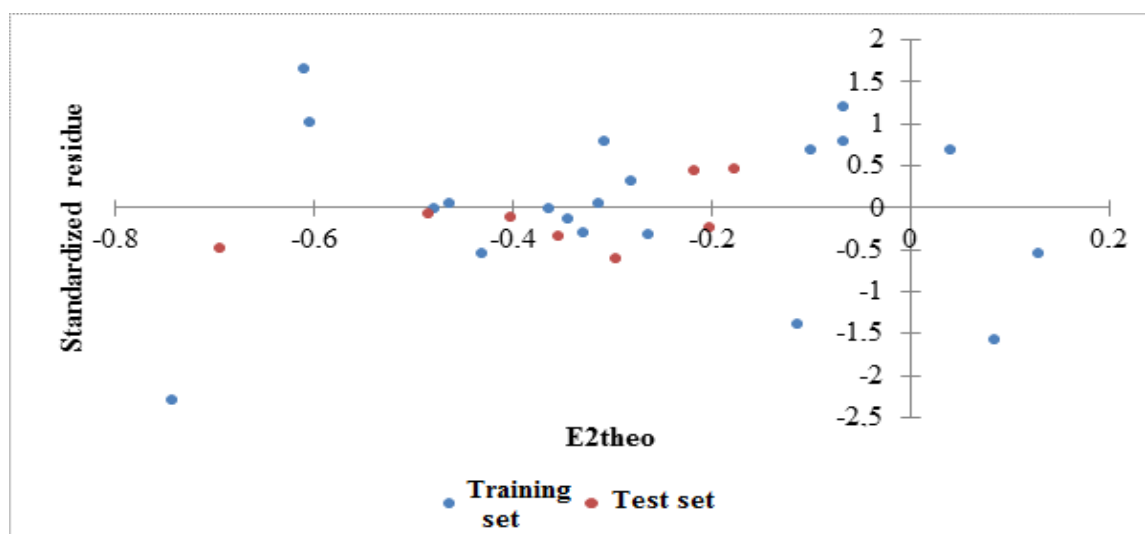


Figure 4. Standardized residuals graph =  $f(E_{theo}^2)$  of the model

### 3.5. Applicability of the model

The model's Applicability Domain (DA) was defined by analyzing the Williams diagram below (Figure 5).

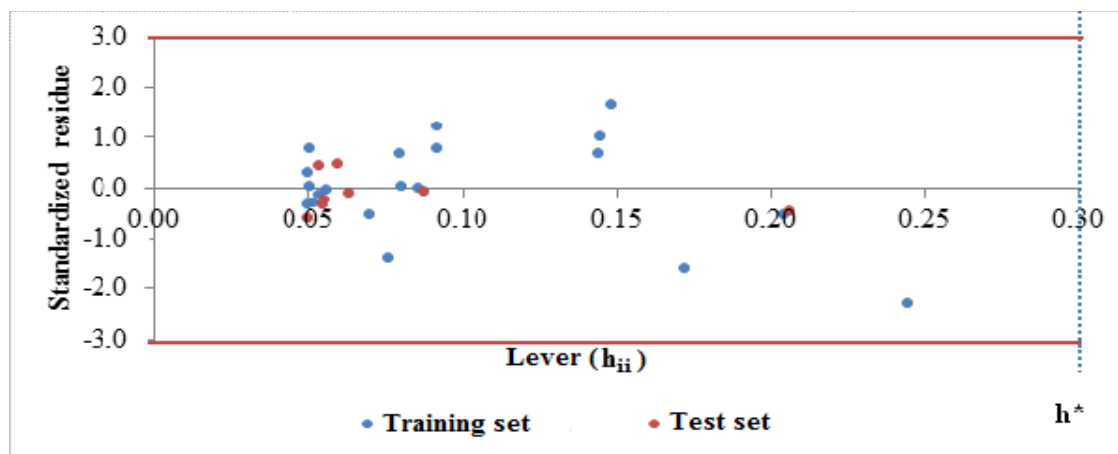


Figure 5. Williams diagram the model

The analysis of Williams diagram shows that all the observations have standardized residues between  $-3\sigma$  and  $+3\sigma$ . In addition, the levers obtained are less than the threshold value  $h^* = 0.3$ . This proves the absence of outliers. The results of the external validation and the domain of applicability show that the model can be used for the prediction of the second reduction potentials of future TCNQ belonging to its applicability domain.

## IV. CONCLUSION

Tetracyanoquinodimethane (TCNQ) molecules and existing TTF-TCNQ charge transfer complexes generally have remarkable redox and electrical properties. However, improving these properties or finding new molecules that still display more attractive properties remains a concern in

scientific research. In this dynamic, we have carried out the work with the aim of developing a predictive QSPR model linking the second reduction potential from a series of Tetracyanoquinodimethane molecules analogous to quantum descriptors derived from the conceptual density functional theory. In the established model, the theoretical of second potential reduction is a function of the electrophilic index. This model displays interesting statistical parameters such like the Fisher coefficient ( $F = 207.2814$ ) which shows that the regression equation of the model is globally significant. In addition, the correlation coefficient ( $R = 0.9592$ ) shows a strong correlation between the second potential reduction and the electrophilic index. The standard deviation ( $s = 0.0726$ ) tends to 0, which shows a good fit and a high reliability of the prediction. Also, certain



statistical parameters reflect the high predictive capacity of this model. Indeed, the internal validation parameters ( $Q_{L00}^2 = 0.8850$ ; SDEP = 0.0827) show that the model has an acceptable predictive capacity towards the molecules of the training set. As for the external validation parameters ( $Q_{ext}^2 = 0.9867$ ; RMSEP = 0.0281), they reveal the model has a high predictive capacity towards molecules of the test set. The verification of the five (05) criteria of Tropsha shows that the model is validated and very efficient in predicting the potential for second reduction. This model is therefore suitable for predicting the second reduction potential of this same family of molecules belonging to its applicability domain.

### REFERENCES

- [1]. Garito, A. F., & Heeger, A. J. (1974). *Design and synthesis of organic metals. Accounts of Chemical Research*, 7(7), 232–240.
- [2]. Kenny, P. W. (1995). *Prediction of planarity and reduction potential of derivatives of tetracyanoquinodimethane using ab initio molecular orbital theory. Journal of the Chemical Society, Perkin Transactions 2*, (5), 907-909.
- [3]. Diarrassouba, F., Koné, M., Bamba, K., Traoré, Y., Koné, M.G.-R. and Assanvo, E.F. (2019) Development of Predictive QSPR Model of the First Reduction Potential from a Series of Tetracyanoquinodimethane (TCNQ) Molecules by the DFT (Density Functional Theory) Method. *Computational Chemistry*, 7(4), 121-142.
- [4]. Andersen, J. R., & Jørgensen, O. (1979). *Organic metals; mono- and 2,5-disubstituted 7,7,8,8-tetracyano-p-quinodimethanes and conductivities of their charge-transfer complexes. J. Chem. Soc., Perkin Trans. 1*, (0), 3095–3098.
- [5]. James J. Delaney B.Sc. (1997), "Synthesis of New Heterocyclic TCNQ Analogues", Doctorate of Philosophy, Dublin City University (School of Chemical Sciences).
- [6]. Wheland, R. C., & Gillson, J. L. (1976). *Synthesis of electrically conductive organic solids. Journal of the American Chemical Society*, 98(13), 3916–3925.
- [7]. Ferraris, J. P., & Saito, G. (1978). "Organic metals" with asymmetric acceptors: the monofluorotetracyanoquinodimethane anion. *J. Chem. Soc., Chem. Commun.*, (22), 992–993.
- [8]. Saito, G., & Ferraris, J. P. (1979). *Difluorotetracyanoquinodimethane: electron affinity cut-off for "metallic" behaviour in a tetrathiafulvalene salt. J. Chem. Soc., Chem. Commun.*, (22), 1027–1029.
- [9]. Suzuki, T., Fujii, H., Yamashita, Y., Kabuto, C., Tanaka, S., Harasawa, M., ... Miyashi, T. (1992). *Clathrate formation and molecular recognition by novel chalcogen-cyano interactions in tetracyanoquinodimethanes fused with thiadiazole and selenadiazole rings. Journal of the American Chemical Society*, 114(8), 3034–3043.
- [10]. Tsubata, Y., Suzuki, T., Yamashita, Y., Mukai, T. and Miyashi, T. (1992) Tetracyanoquinodimethanes Fused with 13, s-Thiadiazole and Pyrazine Units. *Heterocycles*, 33(1), 337-348.
- [11]. Suzuki, T., Yamashita, Y., Kabuto, C., & Miyashi, T. (1989). *Preparation, properties, and crystal structure of a 1,2,5-thiadiazolotetra-cyanoquinodimethane. Journal of the Chemical Society, Chemical Communications*, (16), 1102.
- [12]. R Dennington, T Keith and J Millam., GaussView Version 5, Semicem Inc., Shawnee Mission, KS, 2009.
- [13]. Gaussian 09, Revision A.02, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2009.
- [14]. ACDLABS 10, Advanced Chemistry Development Inc., Toronto, ON, Canada, 2015.
- [15]. Microsoft ® Excel ® 2010.
- [16]. XLSTAT Version 2014.5.03, Copyright Addinsoft 1995-2014, 2014.
- [17]. Minitab® 18

- [18]. Vessereau A.(1988), Méthodes statistiques en biologie et en agronomie, *Lavoisier (Tec & Doc). Paris*, 538 p.
- [19]. Chatterje S. and Hadi A.S.(2006), Regression Analysis by Example. 4th Edition, John Wiley & Sonc, Inc., Hoboken, 366 p
- [20]. A. F. Siegel (1997), Practical Business Statistics, *IRWIN*, 3rd edition
- [21]. P. Besse (2003), Pratique de la modélisation statistique, Publications du laboratoire de statistique et Probabilité, Université Paul Sabatier 31062, Toulouse.
- [22]. R. D. Cook, S. Weisberg (1994), An introduction to Regression Graphics, *Wiley Series in probability and mathematical statistics*, 265 p
- [23]. Kubiny, H. (1994). *Variable Selection in QSAR Studies. I. An Evolutionary Algorithm. QSAR & Combinatorial Science*, 13(3), 285–294.
- [24]. Golbraikh, A., & Tropsha, A. (2002). *Beware of  $q^2$ ! Journal of Molecular Graphics and Modelling*, 20(4), 269–276.
- [25]. Todeschini R. Milano (2010), Chemometrics and QSAR Research Group, University of Milano Bicocca, Milano, Italy (personal communication)
- [26]. Pratim Roy, P., Paul, S., Mitra, I., & Roy, K. (2009). *On Two Novel Parameters for Validation of Predictive QSAR Models. Molecules*, 14(5), 1660–1701.
- [27]. Consonni, V., Ballabio, D., & Todeschini, R. (2010). *Evaluation of model predictive ability by external validation techniques. Journal of Chemometrics*, 24(3-4), 194–201.
- [28]. Roy, K., Mitra, I., Kar, S., Ojha, P. K., Das, R. N., & Kabir, H. (2012). *Comparative Studies on Some Metrics for External Validation of QSPR Models. Journal of Chemical Information and Modeling*, 52(2), 396–408.
- [29]. Tropsha, A., Gramatica, P., & Gombar, V. (2003). *The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. QSAR & Combinatorial Science*, 22(1), 69–77.
- [30]. O. Ouattara, N. Ziao (2017). Quantum Chemistry Prediction of Molecular Lipophilicity Using Semi-Empirical AM1 and *Ab Initio* HF/6-311++G Levels, *Computational Chemistry*, 05(01), 38-50.
- [31]. Gramatica, P. (2007). *Principles of QSAR models validation: internal and external. QSAR & Combinatorial Science*, 26(5), 694–701.
- [32]. Netzeva, T. I., Worth, A. P., Aldenberg, T., Benigni, R., Cronin, M. T. D., Gramatica, P., ... Yang, C. (2005). *Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure-Activity Relationships. Alternatives to Laboratory Animals*, 33(2), 155–173.
- [33]. Koopmans, T. (1934). *Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den Einzelnen Elektronen Eines Atoms. Physica*, 1(1-6), 104–113.
- [34]. Parr, R. G., Szentpály, L. v., & Liu, S. (1999). *Electrophilicity Index. Journal of the American Chemical Society*, 121(9), 1922–1924.
- [35]. Reed, A. E., Curtiss, L. A., & Weinhold, F. (1988). *Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint. Chemical Reviews*, 88(6), 899–926.
- [36]. Parr, R. G., Donnelly, R. A., Levy, M., & Palke, W. E. (1978). *Electronegativity: The density functional viewpoint. The Journal of Chemical Physics*, 68(8), 3801–3807.
- [37]. Mathieu R., probabilités, statistiques et analyses multicritères (Version Courte), Date de parution : juillet 2013 ; Révision : janvier 2017
- [38]. Eriksson, L., Jaworska, J., Worth, A. P., Cronin, M. T. D., McDowell, R. M., & Gramatica, P. (2003). *Methods for Reliability and Uncertainty Assessment and for Applicability Evaluations of Classification- and Regression-Based QSARs. Environmental Health Perspectives*, 111(10), 1361–1375. doi:10.1289/ehp.5758
- [39]. Shapiro, S. S., & Wilk, M. B. (1965). *An Analysis of Variance Test for Normality (Complete Samples). Biometrika*, 52(3/4), 591-611
- [40]. Durbin, J., & Watson, G. S. (1951). *Testing for serial correlation in least squares regression. II. Biometrika*, 38(1-2), 159–178.