RAMReS Sciences des Structures et de la Matière Vol. 6, N°2 (2022) 186 - 211
http://publication.lecames.org/index.php/mat *ISSN: 2630-1180*

Determination of molecular descriptors influencing the first reduction potential of a family of Tetracyanoquinodimethane molecules at HF/6-31G(d,p) theory level

Fatogoma Diarrassouba[1], Kafoumba Bamba[1,*], Mawa Koné[2], K.K. Raymond Kouamé[1], Nahossé Ziao[1]

[1]Laboratoire de Thermodynamique et Physico-Chimie du Milieu (LTPCM), UFR-SFA, Université Nangui Abrogoua 02 BP 801 Abidjan 02, Côte d'Ivoire
[2]Laboratoire de Constitution et de Réaction de la Matière (LCRM), UFR-SSMT, Université Félix Houphouët Boigny 22 BP 582 Abidjan 22, République de Côte-d'Ivoire

Received: 28 June 2022 / Received in revised form: 27 August 2022 / Accepted: 14 September 2022

***Abstract*:**

In our study, we started with a series of forty (40) Tetracyanoquinodimethane (TCNQ) derivatives with their available experimental first reduction potentials, to develop a predictive QSPR (Quantitative Structure-Property Relationship) model. The model obtained relates the potential of the first reduction to three (03) molecular descriptors, namely the electronic affinity (EA), the sum of the absolute value of the Mulliken charges (Q) and the dipole moment (μ_D). This model displays very satisfactory statistical and validation parameters (R^2=0.9503; S=0.0577; F=165.5894 ; Q^2_{LOO}=0.9429 ; R^2_{ext}=0.9544 ; Q^2_{ext}=0.9394). These different parameters show that the QSPR developed model is validated and performs well in the prediction of first reduction potential. Thus, it can be used to effectively predict the potential for first reduction of future TCNQ of the same family that belongs to its domain of applicability with 95% of confidence level.

***Keywords*:** Tetracyanoquinodimethane (TCNQ); First reduction potential; QSPR model.

*Corresponding author:
Email address: kafoumba2001@yahoo.fr *(K. Bamba)*

1. Introduction

The development of plane geometry and conjugated-π electron acceptors is a prerequisite for the formation of organic charge transfer complexes. Many organic compounds function as electron acceptors in charge transfer complexes. Among the most common we can mention p-benzoquinone and its derivatives [1, 2], carboxylic acid anhydrides and their halogenated derivatives [2] and acid chlorides [2]. However, the presence of the powerful cyano group, an electron withdrawer with high electron affinity, makes of the molecules containing the cyano group potential electron acceptors. Tetracyanoethylene (TCNE) was one of the first to be studied [3]. It has formed organic charge transfer complexes with a variety of electron donors and was the precursor of cyano-based electron acceptor molecules. The extension of the conjugation between the cyano groups of TCNE led to the discovery of 7,7,8,8-Tetracyano-p-quinodimethane (TCNQ) [4,5] which is today one of the acceptors of electrons widely studied [6]. Wheland [7] reported that the electron acceptor must have moderate accepting power. To do this, he proposed that the value of the first reduction potential should be between -0.02 V and +0.35 V. Thus, the objective of this work is to develop a QSPR model (Quantitative Structure-Property Relationship) of the first reduction potential from a derived series from Tetracyanoquinodimethane (TCNQ) by ab initio methods. This model will therefore make it possible to explain and predict the first reduction potential of new derivatives of Tetracyanoquinodimethane (TCNQ) likely to have a large spatial extension.

2. Material and methods

2.1. Series of studied molecules

In the development of the predictive QSPR (Quantitative Structure-Property Relationship) model of the first reduction potential, we considered a series of forty Tetracyanoquinodimethane derivatives codified TCNQ [8-14]. The choice of these molecules is due to the availability of their experimental first reduction potentials. These properties have been all determined by cyclic voltammetry in acetonitrile. These molecules have constituted our database. Thirty (30) of which (75% of the database) were used for the training set and ten molecules (10) (25% of the database) were used for the test set. Table 1 presents these different molecules with their corresponding experimental first reduction potentials expressed in volts (V).

Table 1
Series of studied Tetracyanoquinodimethane (TCNQ) molecules.

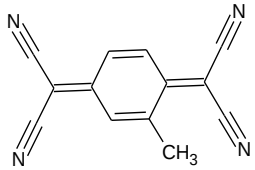
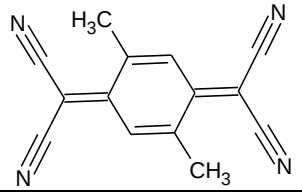
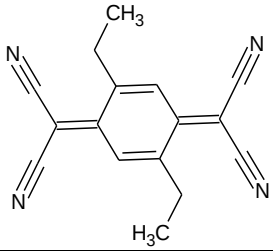
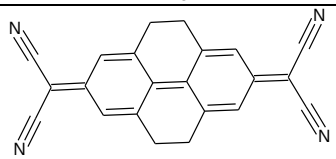
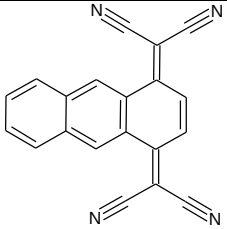
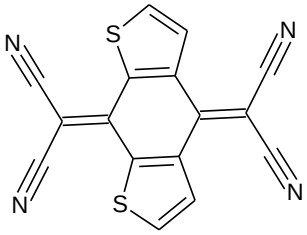
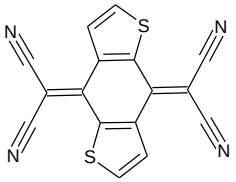
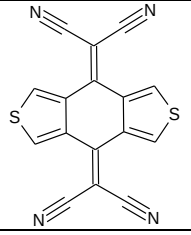
Training set			
Code	Molecule	E^{1}_{exp} (V)[(a)]	Reference
TCNQ_1		+0.170	[9]
TCNQ_2		+0.110	[9]
TCNQ_3		+0.120	[9]
TCNQ_4		+0.012	[8]
TCNQ_5		-0.180	[8]
TCNQ_6		+0.130	[8]
TCNQ_7		+0.130	[8]
TCNQ_8		-0.470	[8]

Table 1 (*continued*)

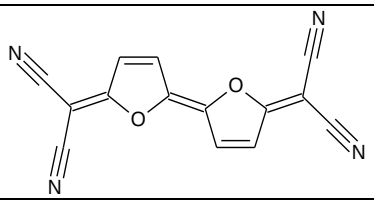
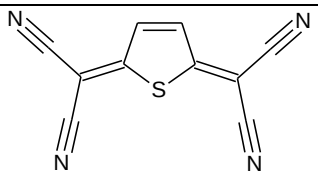
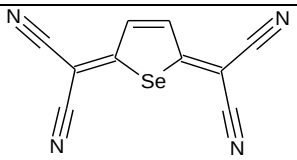
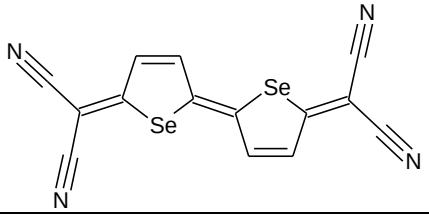
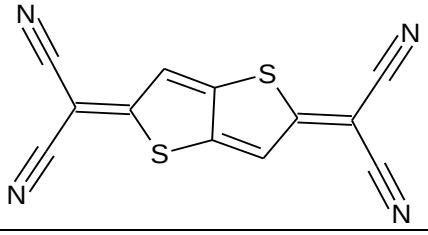
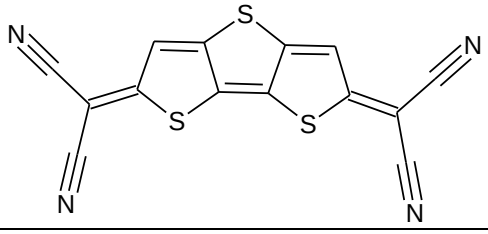
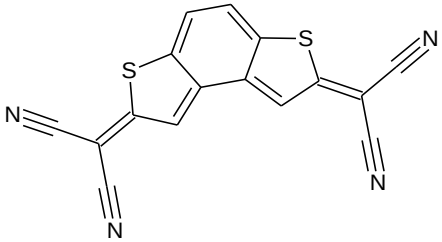
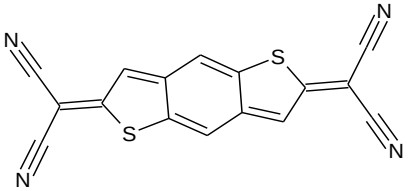
Training set			
Code	Molecule	$E^{1}_{exp}(V)^{(a)}$	Reference
TCNQ_9		-0.090	[8]
TCNQ_10		+0.068	[8]
TCNQ_11		+0.03	[8]
TCNQ_12		-0.05	[8]
TCNQ_13		+0.058	[8]
TCNQ_14		+0.048	[8]
TCNQ_15		+0.320	[8]
TCNQ_16		+0.200	[8]

Table 1 (*continued*)

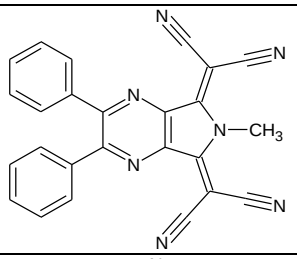
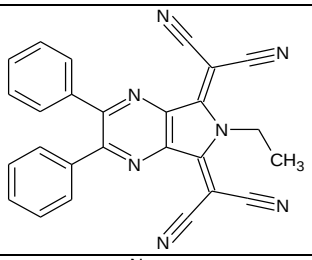
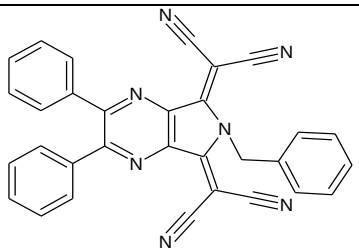
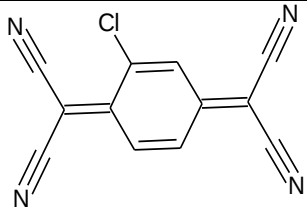
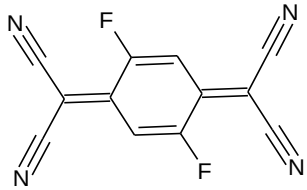
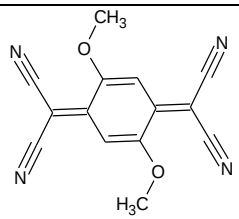
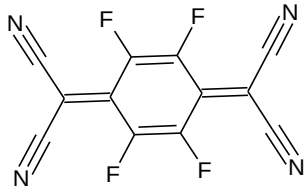
Training set			
Code	Molecule	$E^{1}_{exp}(V)^{(a)}$	Reference
TCNQ_17		-0.360	[8]
TCNQ_18		-0.370	[8]
TCNQ_19		-0.340	[8]
TCNQ_20		+0.290	[10]
TCNQ_21		+0.300	[12]
TCNQ_22		-0.010	[10]
TCNQ_23		+0.530	[10]

Table 1 (*continued*)

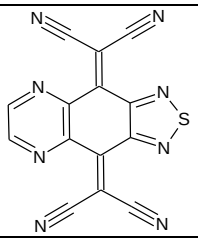
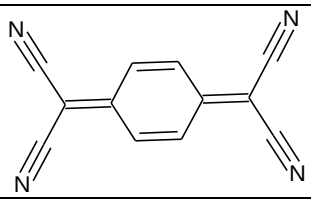
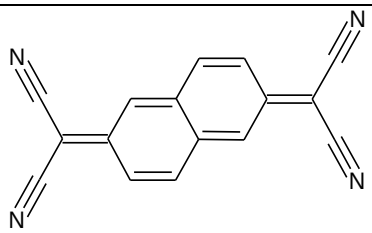
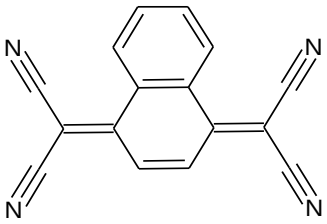
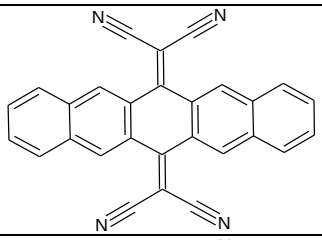
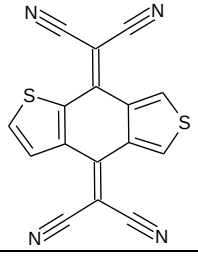
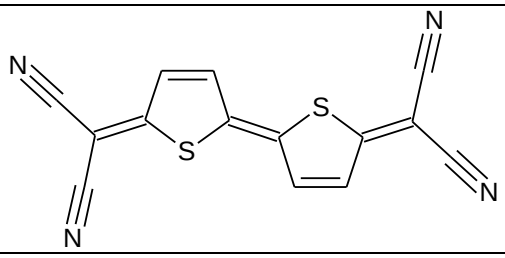
Training set			
Code	Molecule	$E^{1}_{exp}(V)^{(a)}$	Reference
TCNQ_24		-0.010	[13]
TCNQ_25		+0.080	[8]
TCNQ_26		+0.210	[8]
TCNQ_27		-0.040	[8]
TCNQ_28		-0.570	[8]
TCNQ_29		-0.140	[8]
TCNQ_30		-0.026	[8]

Table 1 (*continued*)

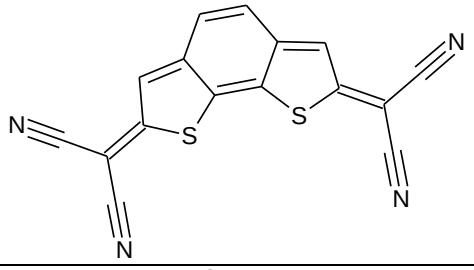
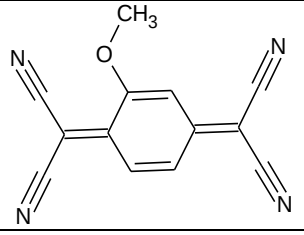
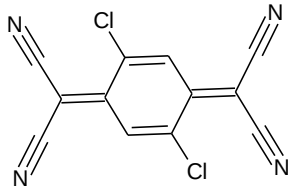
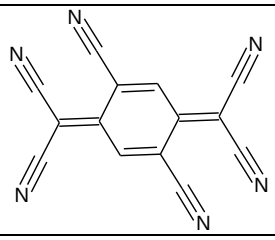
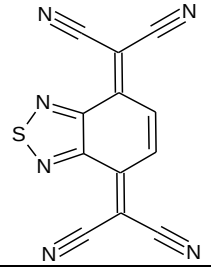
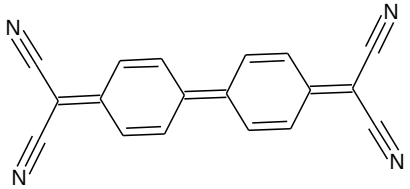
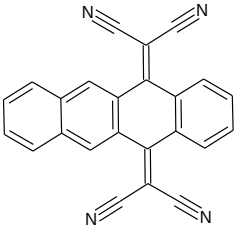
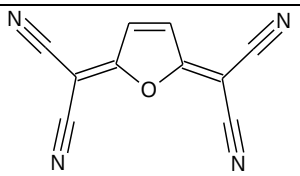
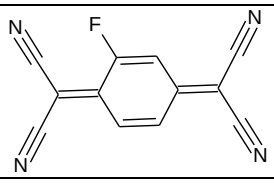
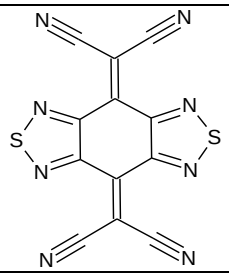
Test set			
Code	Molecule	$E^{1}_{exp}(V)^{(a)}$	Reference
TCNQ_31		+0.260	[8]
TCNQ_32		+0.070	[10]
TCNQ_33		+0.410	[10]
TCNQ_34		+0.650	[10]
TCNQ_35		+0.120	[14]
TCNQ_36		+0.130	[8]
TCNQ_37		-0.440	[8]

Table 1 (*continued*)

Test set			
Code	Molecule	$E^{1}_{exp}(V)^{(a)}$	Reference
TCNQ_38		+0.030	[8]
TCNQ_39		+0.260	[11]
TCNQ_40		-0.020	[13]

[(a)] $E^{1}_{exp}(V)$: first reduction potential.

2.2. Used theory Level and Software

The GaussView 5.0 [15] software was used to represent the 3D structure and to visualize the studied molecules. Then, the Gaussian 09 software [16] was used for optimization and frequency calculation (temperature: 298.15 Kelvin, pressure: 1 atmosphere, in vacuum). These predictions were performed at B3LYP/6-31G(d,p) level. As for 2D structures, they have been represented with Chemsketch [17]. The EXCEL software [18] was used for graphic representation. The XLSTAT software [19] was used for modeling and statistical tests. For the calculation of the observation levers, the Minitab 18 software [20] was used.

2.3. Calculated molecular descriptors

As part of our work, six (06) descriptors were calculated. These are: electronic energy (E_T), dipole moment (μ_D), Ionization Potential (IP), Electronic Affinity (EA), Nucleophilic index (N) and the sum of the absolute value of the Mulliken charges (Q). Table 2 presents the different approaches to these descriptors.

Table 2
List of molecular used descriptors.

Molecular descriptor	Notation	Expression
Electronic energy	E_T	
Dipole moment	μ_D	
Ionization potential	IP	IP=-E_{HOMO} [21]
Electronic affinity	EA	EA=-E_{LUMO} [21]
Nucleophilic index	N	$N\,(Nu) = E_{HOMO}(Nu) - E_{HOMO}(TCE)$[22]
Sum of the absolute value of the Mulliken charges	Q	

TCE stands for TetraCyanoEthene.

2.4. Statistical analysis

To develop a QSPR (Quantitative Structure-Property Relationship) 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 for the implementation of a model and the analysis of its statistical data. But the one we used in our study is Multiple Linear Regression (MLR) (Many explanatory variables). The generalized relation of the Multiple Linear Regression (MLR) equation [23] is as follows:

$$Y = a_0 + a_1X_1 + a_2X_2 + \cdots + a_nX_n \quad (1)$$

In this expression, Y is the response or the dependent variable, X_1, X_2,…, X_n are descriptors (characteristics or independent variables) present in the model with the corresponding regression coefficients a_1, a_2,…, a_n, respectively, and a_0 is the constant term of the model. For given values of X_1, X_2,…, X_n, the response Y must also follow a normal law. The descriptors involved in an MLR model should not correlate. For a statistically reliable model, the maximum number of descriptors should be on the order of one-fifth of the number of compounds in the training set. A MLR model that matches the data provided will lead to a scatterplot showing the deviation of the points from the fit line.

The selection of descriptors is a crucial step in QSPR modeling. In this study, the selection of descriptors was based on two criteria described below.

2.4.1. Criterion 1

There must be a linear dependence relationship between the first reduction potential and the descriptors. Under these conditions we shall have $|R| \geq 0.50$ [24], with R, the linear correlation coefficient of the line $E_{exp} = f(\text{Descripteur}_i)$.

2.4.2. Criterion 2

The descriptors must be independent on each other. To do this, the partial correlation coefficient a_{ij} between the descriptors i and j must be less than 0.70 ($a_{ij} < 0{,}70$) [24]. For a multilinear regression, the coefficients R and a_{ij} are expressed as follows:

$$R = \frac{COV(X,Y)}{S_X.S_Y} \quad (2)$$

$$a_{ij} = \frac{COV(X_i,X_i)}{Var(X_i)} \quad (3)$$

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

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

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

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

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

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

2.4.3. Statistical parameters

2.4.3.1. Determination coefficient(R^2) [25]

The determination coefficient is given by the following relationship:

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

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

2.4.3.2. Standard deviation (S) [26]

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

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

2.4.3.3. Adjusted determination coefficient ($R^2_{adjusted}$) [27]

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.

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

2.4.3.4. Fisher-Snedecor coefficient (F) [28]

It allows testing the global significance of linear regression. A globally significant regression equation contains at least a relevant explanatory variable to explain the dependent variable. The Fisher-Snedecor coefficient is related to the determination coefficient by the following relationship:

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

2.4.3.5. Cross-validation coefficient (Q^2_{LOO}) [29]

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

$$Q^2_{LOO} = 1 - \frac{\sum(y_{i,exp} - y_{i,pred})^2}{\sum(y_{i,exp} - \overline{y}_{exp})^2} = 1 - \frac{PRESS}{TSS} \quad (13)$$

According to Erickson et al. [30], a model with a value of $Q^2_{LOO} > 0.50$ is satisfactory in the prediction. In contrast, when a model displays $Q^2_{LOO} > 0.90$, it is said to be excellent.

2.4.3.6. Cross -validation criteria (PRESS) [29]

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

$$PRESS = \sum(y_{i,exp} - y_{i,pred})^2 \quad (14)$$

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:

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

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

2.4.3.7. Todeschini's parameter ($^cR_P^2$) [31]

$^cR_P^2$ is the corrected form of P.P. Roy's parameter noted R_P^2 [32]. It allows knowing if the model is due to chance correlations or not. If this parameter is greater than 0.50, then the model is not due to a chance correlations. It is defined as:

$$^cR_P^2 = R\sqrt{R^2 - R_r^2} \quad (16)$$

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

2.4.3.8. External validation coefficient (Q_{ext}^2) [33]

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

$$Q_{ext}^2 = 1 - \frac{n}{n_{ext}} \frac{PRESS(test)}{TSS} \quad (17)$$

Here, n_{ext} refers to the number of test set compounds

2.4.3.9. Parameter (RMSEP) [33]

External predictive ability of QSPR (Quantitative Structure-Property Relationship) model may further be determined by root mean square error in prediction given by:

$$RMSEP = \sqrt{\frac{\sum(y_{exp(test)} - y_{pred(test)})^2}{n_{ext}}} \quad (18)$$

2.4.3.10. Roy et al.'s parameters ($\overline{r_m^2}$ and Δr_m^2) [34]

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

$$\overline{r_m^2} = \frac{(r_m^2 + r'^2_m)}{2} \quad (19)$$

$$\Delta r_m^2 = |r_m^2 - r'^2_m| \quad (20)$$

Where,

$$r_m^2 = r^2(1 - \sqrt{r^2 - r_0^2}) \quad (21)$$

and $r'^2_m = r^2(1 - \sqrt{r^2 - r_0'^2})$ (22)

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.

2.4.3.11. External validation criteria or "Tropsha's criteria" [29, 35].

There are five criteria:

- Criterion 1: $R_{ext}^2 > 0.70$
- Criterion 2 : $Q_{ext}^2 > 0.60$
- Criterion 3: $\frac{|R_{ext}^2 - R_0^2|}{R_{ext}^2} < 0.1$ and 0.85<k<1.15
- Criterion 4: $\frac{|R_{ext}^2 - R_0'^2|}{R_{ext}^2} < 0.1$and 0.85<k'<1.15
- Criterion 5: $|R_{ext}^2 - R_0^2| < 0.3$

where, R_{ext}^2 stands for the determination coefficient of molecules for the test set; R_0^2 represents the determination coefficient of the

regression between predicted and experimental values for the test set without intercept; $R_0'^2$ is the determination coefficient of the regression between experimental and predicted values for the test set without intercept. k stands for the slope of the correlation line (predicted values according to the experimental values with intercept = 0) and k' is the slope of the correlation line (experimental values according to the predicted values with intercept = 0). Ouattara et al. [36] reported that if at least 3/5 of the Tropsha's criteria are verified, the developed QSPR (Quantitative Structure-Property Relationship) model is considered as a successful model in predicting of the studied property.

2.4.3.12. Lever (h_{ii}) [37]

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,

$$h_{ii} = x_i(X^TX)^{-1}x_i^T \quad (23)$$

where x_i is the line vector of the descriptors of compound i (i=1, 2… , n) 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 $\frac{3\,(p+1)}{n}$ [38]. Where n is the number of compounds in the training set and p is the number of model descriptors.

If the h_{ii} value of a compound in the training set is greater than the threshold value h*, the structure of this compound reinforces the developed model. If all the data points lie in the region of $0 \leq h_{ii} \leq h^*$ and $-3\sigma \leq R \leq 3\sigma$, the developed model can be considered statistically acceptable and valid. Any compound having a standardized residue greater than three standard deviation units ($R>3\sigma$) and which displays a leverage value less than the threshold value ($h_{ii} <h^*$), can be considered as outlier of the response (Y outlier). On the other hand, if the value of the leverage of a compound is greater than the threshold value ($h_{ii}> h^*$) while the value of R is less than 3 standard deviation units, this compound can be considered as structurally influential (X aberrant). The latter stabilizes the developed model and makes it more precise. It is said to be a good influential point.

2.4.3.13. Shapiro-Wilk test

Very popular, the Shapiro-Wilk test [39] is based on the W parameter. Compared to other tests, it is particularly powerful for small numbers ($n \leq 50$). The test statistic is written:

$$W = \frac{\left[\sum_{i=1}^{k=\left[\frac{n}{2}\right]} a_i(y_{n-i+1}-y_i)\right]^2}{\sum_i(y_i-\bar{y})^2} \quad (24)$$

With,

$a_i = a_{n-i+1}$, y_i corresponds to the series of sorted data;

$\left[\frac{n}{2}\right]$ is the whole part of the report $\frac{n}{2}$;

a_i are generated constants from the mean and the variance covariance matrix of the quantiles of a sample of size n according to the normal distribution. The constants a_{n-i+1} are provided in specific tables. The W statistic can therefore be interpreted as the coefficient of determination

between the series of generated quantiles from the normal distribution and the empirical obtained quantiles from the data.

The null and alternative hypotheses of the Shapiro-Wilk test are as follows:

H_0: The sample of size n follows a normal distribution;

H_1: The sample of size n does not follow a normal distribution.

The value of the W statistic is as high as the compatibility with the normal distribution is credible.

If $W < W_{critical}$, rejects H_0 (25)

If $W > W_{critical}$, do not reject H_0 (26)

2.4.3.14. Durbin-Watson test

Developed by Durbin and Watson (1950, 1951), the Durbin-Watson test [40-42] is used to detect autocorrelation between the residuals of a linear regression. In practice, the error terms are often autocorrelated, which can lead to poor estimation of the parameters. We assume that the residuals ε_i are stationary and distributed according to a normal distribution with mean 0. The null and alternative hypotheses of the Durbin-Watson test are as follows:

H_0: The residuals are not autocorrelated ($\rho = 0$);

H_1: The residuals are distributed according to an autoregressive process of order 1 (AR 1) ($\rho > 0$).

The d statistic of the test is written:

$$d = \frac{\sum_{i=2}^{n}(\varepsilon_i - \varepsilon_{i-1})^2}{\sum_{i=1}^{n}\varepsilon_i^2} \quad (27)$$

where $\varepsilon_i = y_i - \hat{y}_i$ and y_i and $\hat{y}_i$ are respectively the observed values and the predicted values of the response (dependent variable) for compound i. The upper and lower critical values, d_U (Upper) and d_L (Lower) were tabulated (Durbin-Watson table) for different values of k (number of explanatory variables) and n (sample size).

If $d < d_L$, reject H_0: $\rho = 0$ (28)

If $d > d_U$, do not reject H_0: $\rho = 0$ (29)

If $d_L < d < d_U$, the test is inconclusive (30)

2.5. Contribution of an explanatory variable to the prediction of a property

The calculation of the contributions is a very important step because it allows to know the contribution of each explanatory variable in the prediction of the studied property. For any explanatory variable X_i, the contribution C_{xi} to the prediction of the property associated with it is based on the t statistic (Student's test) [43, 44]. The contribution of the variable X_i is given by the following relation:

$$C_{Xi} = \frac{|t(X_i)|}{\sum|t(X_i)|} \times 100 \quad (31)$$

C_{xi} expressed as a percentage (%);

$|t(X_i)|$: absolute value of the t-test of the variable X_i;

$\sum|t(X_i)|$: sum of the absolute values of the t-tests of all the explanatory variables X_i;

3. Results and discussion

3.1. Values of calculated molecular descriptors

In this study, six molecular descriptors were calculated and subjected to two basic selection criteria. Among these, the most relevant have been retained. The selected molecular

descriptors will be used to establish a predictive QSPR model. Table 3 reports the values of these different descriptors.

Table 3
Values of the calculated molecular descriptors.

Training set						
Compound	E_{HF}(ua)	μ_D(D)	IP (eV)	EA (eV)	N(eV)	Q(e)
TCNQ_1	-713.5112	1.2956	9.5405	1.3899	2.5695	5.2086
TCNQ_2	-752.5457	0.0048	9.4458	1.3011	2.6642	5.8597
TCNQ_3	-830.6153	1.1663	9.4003	1.1920	2.7096	6.8178
TCNQ_4	-1057.8633	0.0007	8.2606	1.3047	3.8494	7.3640
TCNQ_5	-979.7865	7.0566	8.8813	0.6940	3.2286	6.4581
TCNQ_6	-1620.9555	5.3314	9.4455	1.0323	2.6644	6.8626
TCNQ_7	-1620.9555	5.2417	9.4621	1.0187	2.6478	6.9503
TCNQ_8	-1620.9652	5.8346	9.7824	-0.0087	2.3275	7.0703
TCNQ_9	-899.9120	0.0009	8.9094	1.1041	3.2006	7.8243
TCNQ_10	-995.0777	3.2992	9.9220	1.1041	2.1879	4.3068
TCNQ_11	-2995.1456	2.5805	9.8140	1.0464	2.2960	4.0793
TCNQ_12	-5545.3515	0.0008	8.6467	0.9386	3.4632	5.1305
TCNQ_13	-1468.3189	0,0025	9.2999	1.1596	2.8100	5.2730
TCNQ_14	-1941.5576	3.7802	8.7349	1.1928	3.3750	6.0698
TCNQ_15	-1620.9586	7.5197	8.3741	1.6256	3.7359	6.0658
TCNQ_16	-1620.9677	0.0001	8.6198	1.5181	3.4901	6.0281
TCNQ_17	-1335.3861	7.0203	8.7387	0.4912	3.3712	11.2990
TCNQ_18	-1374.4254	5.9488	8.7072	0.4765	3.4028	11.8564
TCNQ_19	-1564.9354	7.0959	8.6876	0.4003	3.4224	12.8765
TCNQ_20	-1133.3549	0.6473	9.7514	1.6805	2.3585	4.5277
TCNQ_21	-872.1519	0.0013	9.9027	1.8052	2.2072	5.6343
TCNQ_22	-902.2406	0.0003	9.2234	1.0568	2.8865	7.3445
TCNQ_23	-1069.8062	0.0003	10.1011	2.1407	2.0088	6.3119
TCNQ_24	-1364.3140	4.8625	9.9789	0.8581	2.1310	8.0415
TCNQ_25	-674.4769	0.0021	9.6496	1.4810	2.4603	4.6139
TCNQ_26	-827.1226	0.0000	8.7850	1.6236	3.3250	5.6701
TCNQ_27	-827.1288	5.7800	9.6308	0.8752	2.4791	5.4756
TCNQ_28	-1285.1043	9.4298	8.7151	-0.1976	3.3949	8.3138
TCNQ_29	-1620.9606	5.6378	9.5770	0.5840	2.5330	7.0014
TCNQ_30	-1545.2181	0.0002	8.7107	1.0135	3.3993	5.7957

Table 3 (*continued*)

Test set						
Compound	E_{HF}(ua)	μ_D(D)	IP (eV)	EA (eV)	N(eV)	Q(e)
TCNQ_31	-1620.9663	7.0390	8.4998	1.5061	3.6102	6.0774
TCNQ_32	-788.3593	2.2953	9.3892	1.2772	2.7208	5.9068
TCNQ_33	-1592.2326	0.0060	9.8505	1.8656	2.2595	4.4763
TCNQ_34	-857.8923	0.0021	10.3050	2.3642	1.8050	5.1098
TCNQ_35	-1179.6872	0.2582	9.8981	1.2878	2.2119	6.3227
TCNQ_36	-904.0091	0.0009	8.2744	1.6443	3.8355	6.0517
TCNQ_37	-1132.4472	9.0068	8.7627	-0.0397	3.3473	7.6001
TCNQ_38	-672.4243	5.6929	10.1493	1.0663	1.9607	5.6617
TCNQ_39	-773.3150	0.9722	9.7702	1.6506	2.3398	5.0955
TCNQ_40	-1684.8960	2.5177	10.1839	0.8423	1.9261	8.0745

3.2. Selection of relevant molecular descriptors

3.2.1. Submission of molecular descriptors to selection criterion 1

The submission of molecular descriptors to selection criterion 1 is presented in table 4. The examination of the data in table 4 shows that there is a linear dependency relationship between the first reduction potential and the descriptors: μ_D, EA, and Q as $|R| > 0.50$. Regarding the electronic energy E(HF), the ionization potential EI and the nucleophilicity index N, it is clear that $|R| < 0.50$. Thus, they must be systematically rejected. Among these different calculated descriptors, only EA, Q and μ_D will therefore be subject to criterion 2.

3.2.2. Submission of molecular descriptors to selection criterion 2

The submission of molecular descriptors to selection criterion 2 is presented in table 5. On analysis of the results of table 5, the partial correlation coefficients are all less than 0.70. This implies that the descriptors: EA, Q and μ_D are independent two by two. They can therefore coexist in the same QSPR model. Also, in order to identify the descriptors which actually contribute to the prediction of the first reduction potential, statistical tests were also carried out.

Table 4
Submission of molecular descriptors to criterion 1.

Equation	Correlation coefficient $\|R\|$	Descriptor rejected if $\|R\| < 0.50$
$E_{exp} = f(E(HF))$	0.1210	Rejected
$E_{exp} = f(\mu_D)$	0.6025	Retained
$E_{exp} = f(EI)$	0.3155	Rejected
$E_{exp} = f(EA)$	0.9615	Retained
$E_{exp} = f(N)$	0.3155	Rejected
$E_{exp} = f(Q)$	0.6191	Retained

Table 5
Submission of descriptors to criterion 2.

Correlation between:	Coefficient a_{ij}	Independent descriptors if $a_{ij} < 0.7$
AE et μ_D	-0.1168	Independent
AE et Q	-0.1367	Independent
Q et μ_D	0.3514	Independent

3.3. Development of the QSPR model: $\mathbf{E^1_{theo} = f(EA, Q\ et\ \mu_D)}$

3.3.1. Regression equation and Student's test of the QSPR model

The regression coefficients in table 6 attributed to the different explanatory variables lead to the following regression equation:

$$E^1_{theo} = -0.4237 + 0.4769 \times AE - 0.0180 \times Q + 0.0143 \times \mu_D \quad (32)$$

The regression equation indicates that the coefficients of the variables EA and μ_D are positive when that of Q is negative. Under these conditions, the first reduction potential evolves in the same direction as EA and μ_D when it evolves in the opposite direction of Q. Indeed, the increase in the parameters EA and μ_D leads to an increase in the theoretical redox potential while a large value of the charge Q leads to a reduction in the redox potential. For the constant at the origin and the explanatory variable EA, we note values of the *p-values* belong to the interval [0; 0.001[showing that the latter have a highly significant influence on the potential of first reduction. Regarding the variables Q and μ_D, we record values of *p-values* belong to] 0.001; 0.01]. This shows that these last two variables have a very significant influence on the first reduction potential of the family of studied molecules. The absolute values of the t statistical test and the values of the contributions reveal that the electronic affinity EA still makes the strongest contribution (73.01%) indicating that this is the main predictor of the potential for first reduction of the studied TCNQ family. The sum of the charges of Mulliken Q brings a contribution of 13.46% when the dipole moment μ_D brings a contribution of 13.52%.

3.3.2. Analysis of Variance table (ANOVA) and overall Fisher test of the model

The summary of the determined parameters from the ANOVA test is reported in table 7. We note that the *p-value* belongs to the interval [0; 0.001[indicating that the model regression equation is highly significant for the prediction of the first reduction potential of the series of studied molecules. This significance is confirmed by the very high Fischer value (F = 165.5894) which is very large than the significance limit value (F_{limite} = 2.98). To be more precise, at least one of the explanatory variables is relevant to explain the dependent variable (redox potential). Moreover, the experimental variance is equal to TSS = 1.7407 when the theoretical variance due to the model is equal to ESS = 1.6541.

3.3.3. Statistical parameters of model

We note in table 8 that the correlation coefficient is very high (R=0.9748), which means that the potential for first reduction is strongly correlated with the selected descriptors. The coefficient of determination R^2=0.9503 translates that 95.03% of the experimental variance of the first reduction potential is explained by the descriptors of the model. In addition, the very low standard deviation (S=0.0577) shows a good fit and a high reliability of the prediction. Certainly, the statistical parameters gave satisfactory results but it is necessary to carry out the validation tests of this model.

Table 6
Values of the regression coefficients of model.

Source	Coefficient	Standard error	t	C_{xi}(%)	p-value
Constante	-0.4237	0.0662	-6.3987	-	< 0,0001 (***)
AE	0.4769	0.0308	15.4791	73.01	< 0,0001 (***)
Q	-0.0180	0.0063	-2.8545	13.46	0,0084 (**)
μ_D	0.0143	0.0050	2.8669	13.52	0,0081 (**)

(**)Very significant (***)Highly significant $|t_{critique}| = 2.06$

Table 7
ANOVA output.

Source	DDL	Sum of squares	Average of squares	F	*p-value*
Modèle	3	1.6541	0.5514	165.5894	< 0.0001 (***)
Erreur	26	0.0866	0.0033		
Total corrigé	29	1.7407			

Table 8
Statistical parameters of the model.

n	R	R^2	$R^2_{ajusté}$	S	F
30	0.9748	0.9503	0.9445	0.0577	165.5894

3.4. Internal validation of the model

3.4.1. Leave-One-Out cross-validation of the model

In table 9, statistical parameters of the LOO cross-validation of the model are gathered. We note remarkably that the cross validation coefficient LOO of the model is greater than 0.90 ($Q^2_{LOO} > 0.90$). That proves that the model is excellent in the prediction of the potential of first reduction of the family of studied molecules according to Ericsson et al. [30]. Also, out of 100 molecules in the training set, 94.29 have their redox potentials predicted. The model therefore has a very high predictive capacity with respect to the molecules of the training set. This result therefore shows that our developed QSPR model is very insensitive to an operation to set apart a molecule and put it back in the training series (Leave-One-Out) because the cross-validation coefficient Q^2_{LOO} is close to that of determination R^2. This justifies its robustness. Regarding the coefficient ($\overline{r^2_m(LOO)}$), its value is greater than 0.50 when that of $\Delta r^2_m(LOO)$ is less than 0.2. Therefore, for the prediction of the redox potential of the training set, the model is acceptable. To find out whether the established QSPR model is hazardous, the randomization test for the studied property was carried out.

3.4.2. Model Y-randomization test

In the case of randomization, a circular permutation (i.e. 29 iterations) was performed. The summary of the mean values of the randomization parameters are given in table 10. The mean value of the randomized coefficient of determination R^2_r is very low ($R^2_r = 0.1397$) indicating that the equation of the regression line determines only 13.97% of the distribution of points (potential for first reduction). In addition, there is a strong scatter of the scatter plot around the regression line confirmed by a high randomized standard deviation (s_r=0.2397). The very low value of the statistic ($F_r = 1.5218$) of the randomized model shows the equation of the randomized model is not significant. Regarding the corrected parameter ${}^cR^2_P$ of Todeschini, its value is much greater than 0.50 (${}^cR^2_P > 0.50$). The parameter ${}^cR^2_P$ being greater than 0.50, the developed QSPR model is not due to chance correlations.

Table 9
Statistical parameters of the LOO cross-validation of the model.

n	Q^2_{LOO}	$\overline{r^2_m(LOO)}$	$\Delta r^2_m(LOO)$	PRESS	SDEP
30	0.9429	0.9259	0.0039	0.0994	0.0576

Table 10
Mean values of the parameters of the randomization of model.

Randomized parameter	R^2_r	s_r	F_r	${}^cR^2_P$
Average value	0.1397	0.2397	1.5218	0.8777

3.5. External validation of the model

3.5.1. External model validation parameters

Statistical parameters of external model validation are presented in table 11.

From the analysis of the data in table 11, it is to be understood that the model has a very high predictive power due to the high value of the external validation coefficient ($Q^2_{ext} = 0.9394$). This translates that out of 100 molecules in the test set, 93.94 have their redox potentials predicted by the model. Likewise, 95.44% of the experimental variance of the first reduction potential is explained by the descriptors of the model. Regarding $\overline{r^2_m(test)}$, we note a value is greater than 0.50 while that of $\Delta r^2_m(test)$ is less than 0.2. Thus, the model is acceptable for predicting the redox potential of test set molecules. In addition, the five (05) criteria of Tropsha were verified.

3.5.2. Verification of the model's Tropsha criteria

Criterion1: $R^2_{ext} = 0.9544 > 0{,}70$

Criterion 2: $Q^2_{ext} = 0.9394 > 0{,}60$

Criterion 3: $\frac{|R^2_{ext} - R^2_0|}{R^2_{ext}} = 0.0001 < 0.1$

and $k = 1.0007$ with $0.85 < k < 1.15$

Criterion 4: $\frac{|R^2_{ext} - R'^2_0|}{R^2_{ext}} = 0.0001 < 0.1$

and $k' = 0.9639$ with $0.85 < k' < 1.15$

Criterion 5: $|R^2_{ext} - R^2_0| = 0.0001 < 0.3$

It can be seen that all five (05) Tropsha criteria are verified. In turn, the model is very efficient in predicting the first reduction potential of molecules in the test set of the experimental database.

3.6. Comparison between predicted values by the model and experimental values of the redox potential of the test set.

The theoretical values of the redox potentials were compared with the experimental values based on the regression constants of the graphs $E^1_{theo} = f(E^1_{exp})$ of figure 1 and $E^1_{exp} = f(E^1_{theo})$ of figure 2.

From figures 1 and 2, the values of the guiding coefficients k = 1.0007 and k'= 0.9639 are very close to unity. Which means, the value of the calculated redox property is very close to the experimental value ($E^1_{theo} \approx E^1_{exp}$). Which leads to a report closer to unity. The results of the external validation further prove that the model performs very well in the prediction of the first reduction potential of the series of studied molecules. It can be used effectively for the prediction of the first reduction potential of new TCNQ within its area of applicability.

Table 11

Statistical parameters of external model validation.

n_{ext}	R^2_{ext}	Q^2_{ext}	$\overline{r^2_m(test)}$	$\Delta r^2_m(test)$	PRESS (test)	RMSEP
10	0.9544	0.9394	0.9449	0.0000	0.0352	0.0593

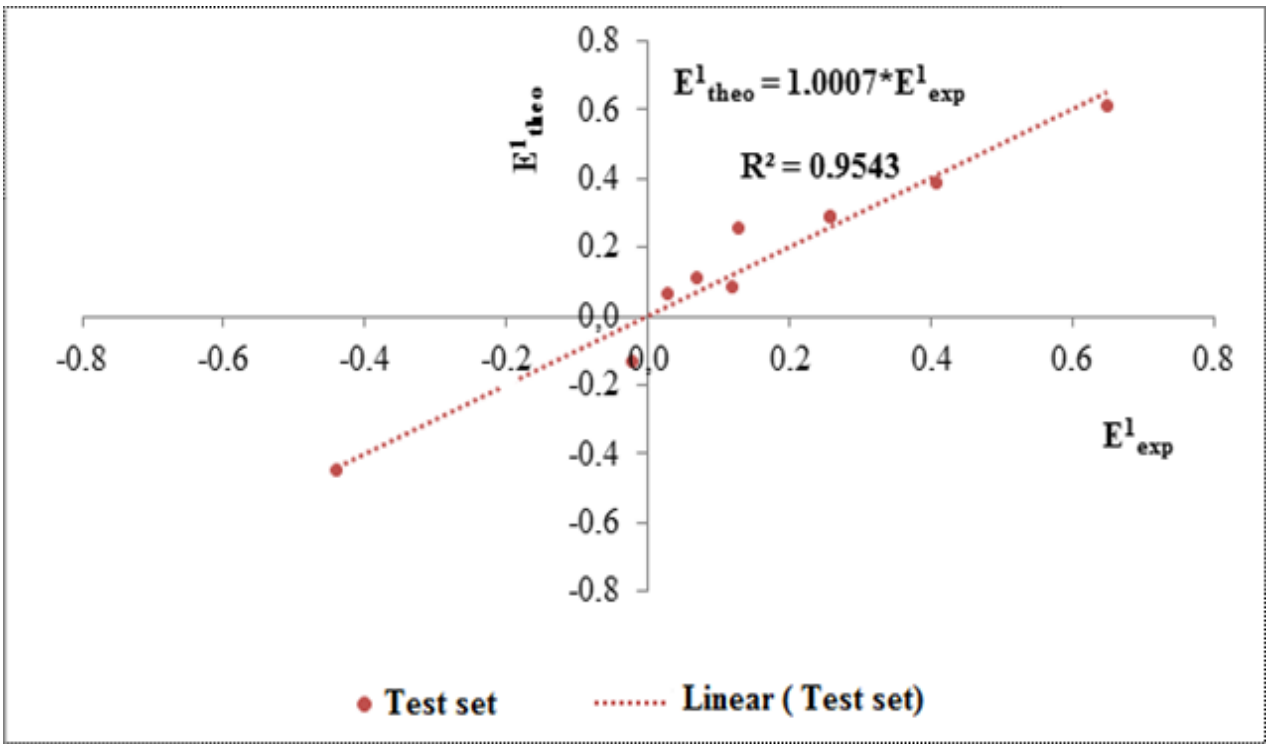


Fig. 1. $E^1_{theo}=f(E^1_{exp})$ graph of the model test series (intercept = 0).

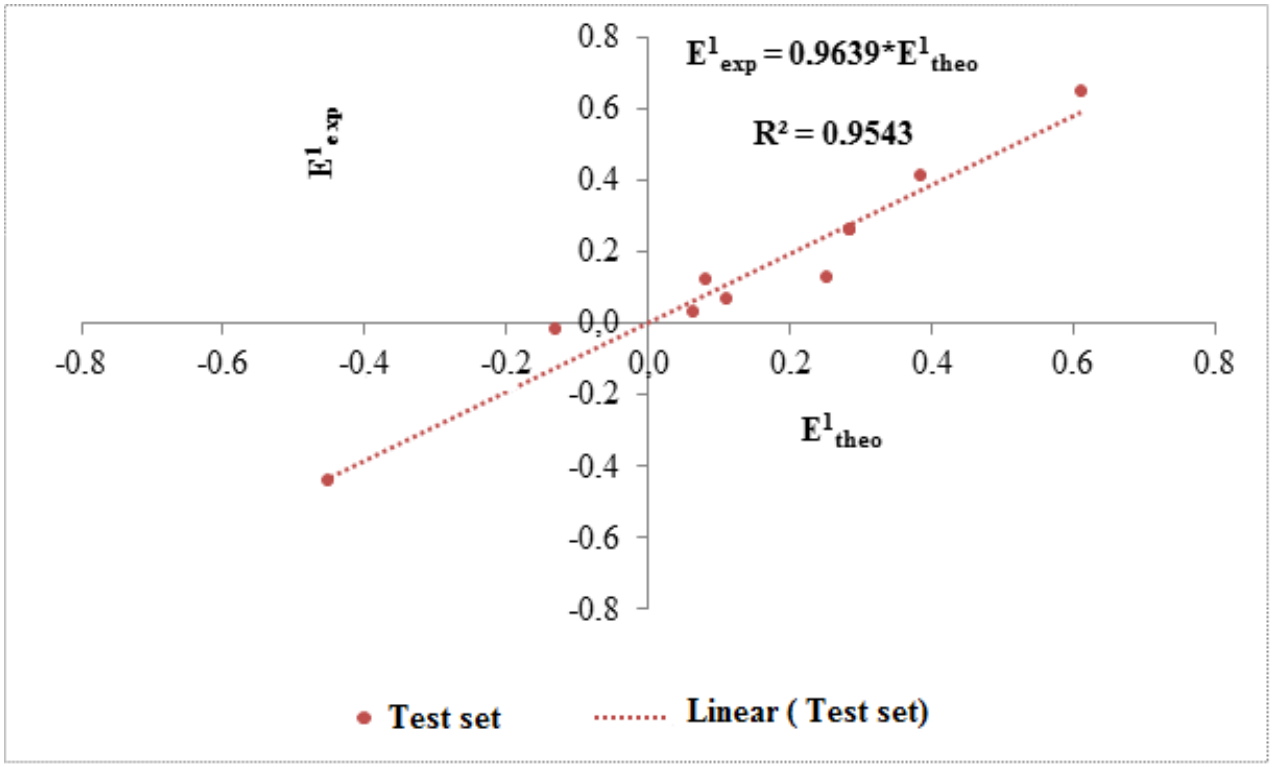


Fig. 2. $E^1_{exp}=f(E^1_{theo})$ graph of the model test series (intercept = 0).

3.7. Correlation between predicted values by the model and experimental values of the redox potential

Through the distribution of the scatter plot around the regression line in figure 3, we note that there is a strong linear correlation between the predicted values by the model and the experimental values of the first reduction potential. In figure 4, we see that there is a similarity between the curves of the predicted values by the model and the experimental values, especially for the test set. Consequently, these graphs confirm that the model is validated and is very efficient in the prediction of the redox potential. This also reflects the adequacy of the level of used theory for the development of this QSPR model.

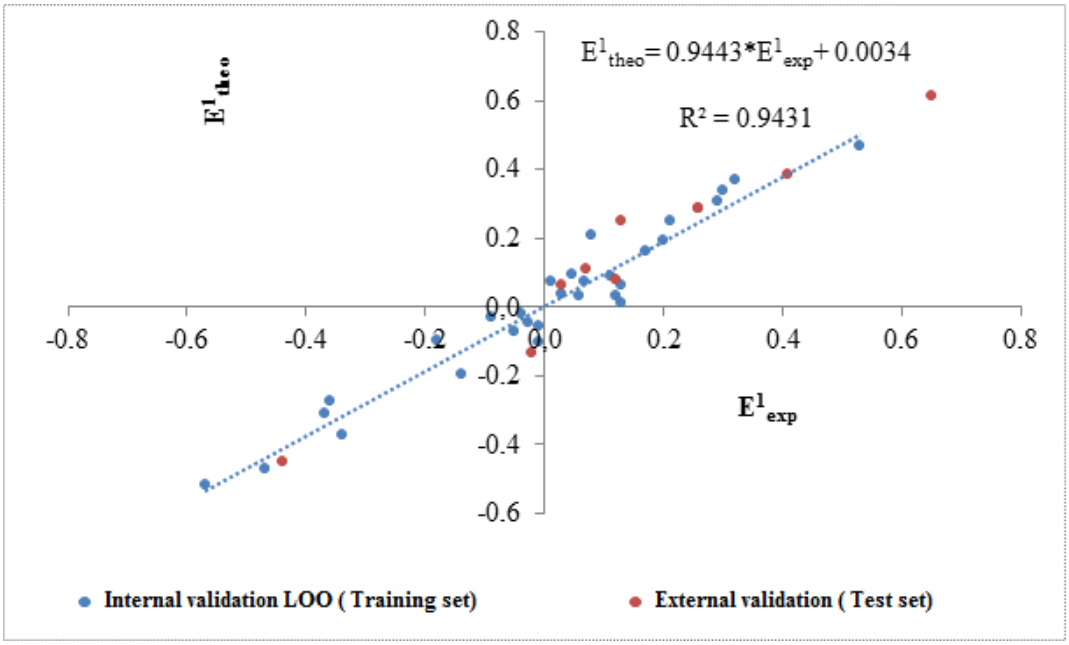


Fig. 3. $E^1_{theo} - E^1_{exp}$ scatterplot of the model.

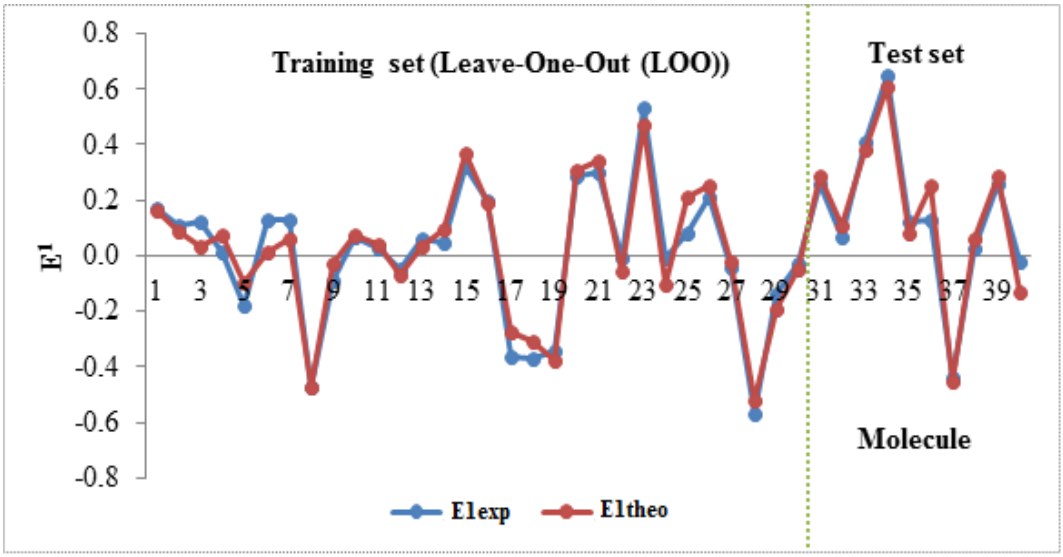


Fig. 4. Similarity between predicted values by model and experimental values.

3.8. Statistical tests

3.8.1 Shapiro-Wilk test of the model

Parameter values of the Shapiro-Wilk test of the model are presented in table 12. Analysis of the values in this table shows that the calculated p-value is greater than $1-\alpha = 0.05$ (5% significance level). As regards $W_{critical}$, we note that the value is less than that of calculated W. As a result, the normality assumption is compatible with our data. This normal distribution is confirmed by the distribution of the point cloud along the first bisector (the equation line y = x) (Fig. 5).

3.8.2. Durbin-Watson test of the model

The data in table 13 show the calculated statistical test d is greater than the maximum critical ($d_U = 1.66$). Also, the calculated p-value is greater than $1-\alpha = 0.05$ (significance level of 5%). It is therefore clear that the residuals are not autocorrelated. These residuals do not contain any information that could influence the prediction of the first reduction potential by the model. This interpretation is confirmed by the random distribution of the point cloud in figure 6.

Table 12
Parameter values of the Shapiro-Wilk test of the model.

W	$W_{critique}$	p-value	$1-\alpha$
0.9778	0.940	0.6086	0.05

Table 13
Parameter values of the model Durbin-Watson test.

d	d_L	d_U	p-value	$1-\alpha$
1.8706	1.34	1.66	0.3060	0.05

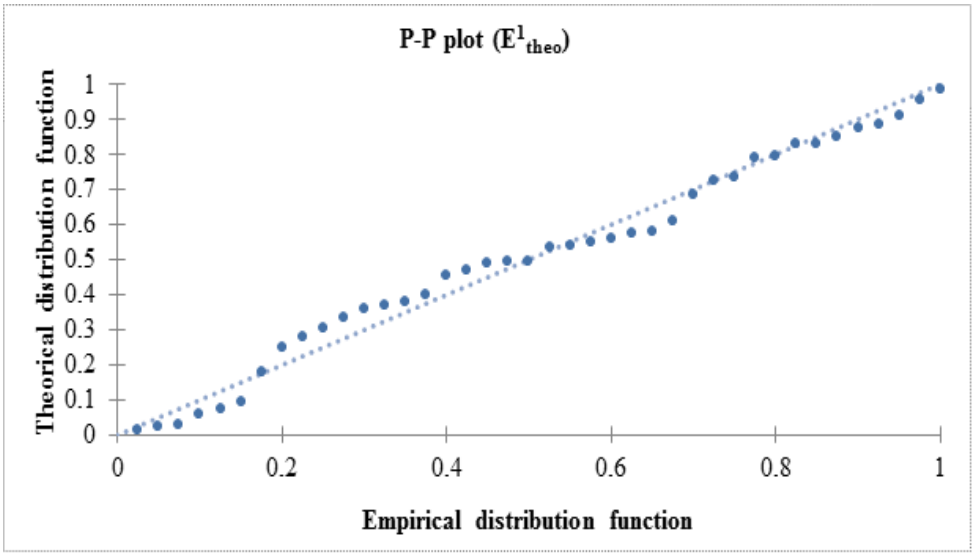


Fig. 5. P-P plot (E^1_{theo}) graph of the model.

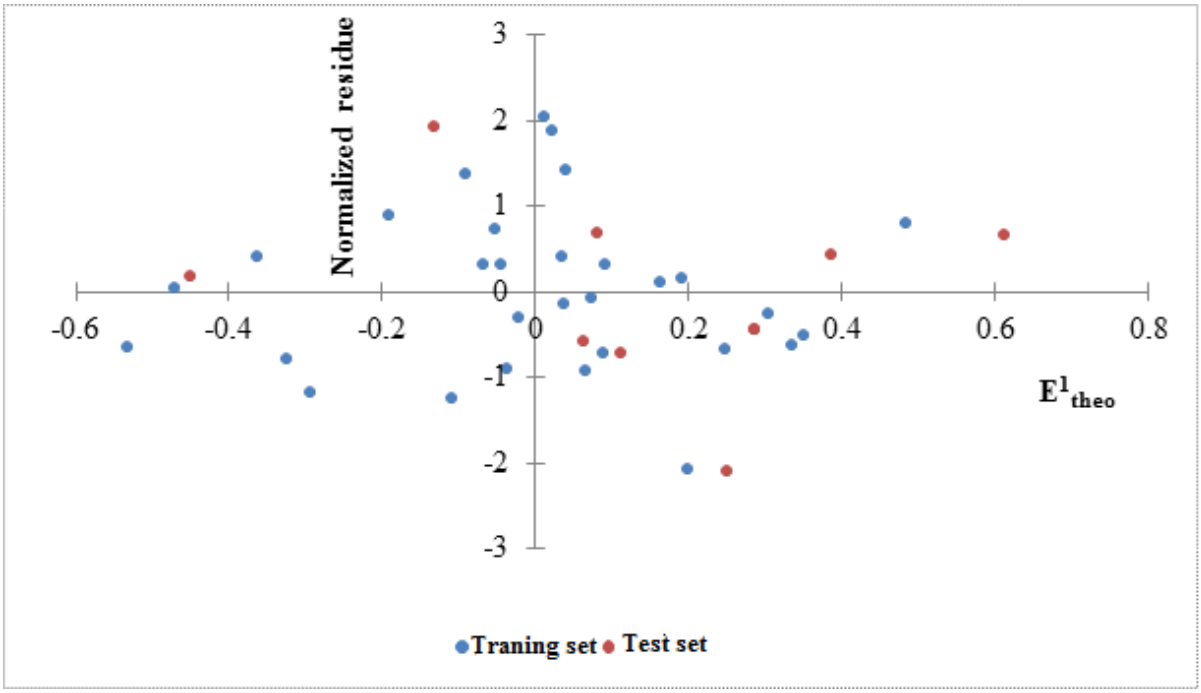


Fig. 6. Normalized residual graph =$f(E^1_{theo})$ of the model.

3.9. Applicability Domain of the model

The Williams diagram (graphical representation of the standardized residuals as a function of the levers) was represented to define the Domain of Applicability (Fig. 7).

On the Williams diagram, we notice all the observations of the training set have their standardized Residuals figure between ± 3 standard deviation units. It is the same for the test set. This shows that no observation of the two sets is aberrant. In addition, the levers obtained are all lower than the critical value ($h^* = 0.4000$) except that of TCNQ_15. The TCNQ_15 observation has its lever greater than the critical value but with a low standardized residual (value between ± 3 standard deviation units) indicating that the latter reinforces the elaborated QSPR model and increases its precision in the prediction: It is a good influential point of view that it belongs to the training set. The results of the external validation reveal that the model can be used as a model for predicting the redox potential of future TCNQ belonging to the same family in its domain of applicability.

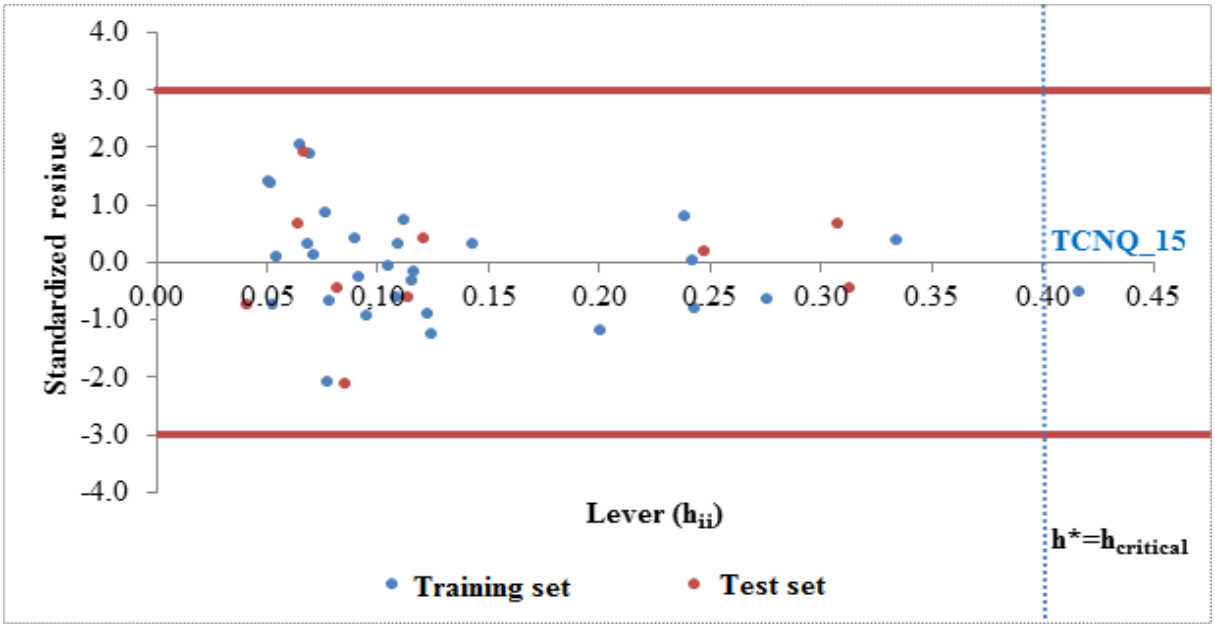


Fig. 7. Williams diagram.

4. Conclusion

The main objective of this work was to make a study of Quantitative Structure-potential first reduction of a series of forty (40) Tetracyanoquinodimethane (TCNQ) derivatives. A predictive QSPR model depending on three (03) molecular descriptors was established. They are: electron affinity (EA), the sum of the absolute value of the Mulliken charges (Q) and the dipole moment (μ_D). This model displays various statistical and validation parameters very satisfactory (R^2=0.9503 ; S=0.0577 ; F=165.5894; Q^2_{LOO}=0.9429;Q^2_{ext}=0.9394; $\overline{r^2_m(LOO)}$=0.9259; $\Delta r^2_m(LOO)$=0.0039). These different parameters show that the developed QSPR model is validated and performs well in the prediction of first reduction potential. It is acceptable as a prediction model. Consequently, it can now be used to predict the potential for first reduction of future TCNQ of the same family that belongs to its domain of applicability. Therefore, we plan to exploit this model to design new derivatives of Tetracyanoquinodimethane (TCNQ) with moderate oxidizing power.

References

[1] J.B. Torrance, J.J. Mayerle, V.Y. Lee, K. Bechgaard, *A new class of highly conducting organic materials: charge-transfer salts of the tetrathiafulvalenes with the tetrahalo-p-benzoquinones*, J. Am. Chem. Soc. 101 (1979) 4747-4748.

[2] V.P. Parini, *Organic charge-transfer complexes*, Russ. Chem. Rev. 31 (1962) 408.

[3] R.E. Merrifield, W.D. Phillips, *Cyanocarbon Chemistry. II.[1] Spectroscopic Studies of the Molecular Complexes of Tetracyanoethylene*, J. Am. Chem. Soc. 80 (1958) 2778–2782.

[4] D.S. Acker, R.J. Harder, W.R. Hertier, W. Mahler, L.R. Melby, R.E. Benson, W.E. Mochel, *7,7,8,8-Tetracyanoquinodimethane and its electrically conducting anion-radical derivatives*, J. Am. Chem. Soc. 82 (1960) 6408–6409.

[5] D.S. Acker, W.R. Hertier, *Substituted Quinodimethans. I. Preparation and Chemistry of 7,7,8,8-Tetracyanoquinodimethan*, J. Am. Chem. Soc. 84 (1962) 3370–3374.

[6] L.R. Melby, R.J. Harder, W.R. Hertler, W. Mahler, R.E. Benson, W.E. Mochel, *Substituted Quinodimethans. II. Anion-radical Derivatives and Complexes of 7,7,8,8-Tetracyanoquinodimethan*, Journal of the American Chemical Society 84 (1962) 3374-3387.

[7] R.C. Wheland, *Correlation of Electrical Conductivity in Charge-Transfer Complexes with Redox Potentials, Steric Factors, and Heavy Atom Effects*, J. Amer. Chem. Soc. 98 (1976) 3926-3930.

[8] J. James, B.Sc. Delaney, Synthesis of New Heterocyclic TCNQ Analogues, Doctorate of Philosophy, Dublin City University (School of Chemical Sciences) (1997).

[9] J.R. Andersen, O. Jorgensen, *Organic Metals; Mono- and 2,5-Di-substituted 7,7,8,8-tetracyano-pquinodimethanes and Conductivities of their Charge-transfer Complexes*, J. Chem. Soc. Perkin Trans. 1 (1979) 3095-3098.

[10] R.C. Wheland, J.L. Gillson, *Synthesis of Electrically Conductive Organic Solids*, J. Am. Chem. Soc. 98 (1976) 3916-3925.

[11] J.P. Ferraris, G. Saito, *Organic Metals with Asymmetric Acceptors: The Monofluorotetra-cyanoquino-dimethane Anion*, J. Chem. Soc. Chem. Comm. 22 (1978) 992-993.

[12] G. Saito, J.P. Ferraris, *Difluorotetra-cyanoquinodimethane : Electron Affinity Cut -off for 'Metallic' Behaviour in a Tetrathiafulvalene Salt*, J. Chem. Soc. Chem. Comm. 22 (1979) 1027-1029.

[13] Y. Tsubata, T. Suzuki, Y. Yamashita, T. Mukai, T. Miyashi, *Tetracyanoquinodimethanes fused with 13, s-thiadiazole and pyrazine units*, Heterocycles 33 (1992) 337-348.

[14] Y. Yamashita, *Novel Electron Acceptors and Donors Containing Fused-heterocycles*, J. Org. Syn. Chem. 47 (1989) 1108-1117.

[15] R Dennington, T. Keith, J Millam, GaussView Version 5, Semichem Inc., Shawnee Mission, KS (2009).

[16] M.J. Frisch et al., Gaussian 09, Revision A.02, Gaussian, Inc., Wallingford CT (2009).

[17] ACDLABS 10, Advanced Chemistry Development Inc., Toronto, ON, Canada (2015).

[18] Microsoft ® Excel ® (2010).

[19] XLSTAT Version 2014.5.03, Copyright Addinsoft 1995-2014 (2014).

[20] Minitab® 18, Minitab, Inc. (2017).

[21] Koopmans, T., *Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den Einzelnen Elektronen Eines Atoms*, Physica, 1 (1934) 104–113.

[22] L.R. Domingo, E. Chamorro, P. Pérez, *Understanding the Reactivity of Captodative Ethylenes in Polar Cycloaddition Reactions. A Theoretical Study*, J. Org. Chem. 73 (2008) 4615-4624.

[23] G.W Snedecor, W.G. Cochran, Statistical methods, Wiley-Blackwell (1991).

[24] A. Vessereau, Méthodes statistiques en biologie et en agronomie, Lavoisier (Tec & Doc), Paris, (1988).

[25] S. Chatterje, A.S. Hadi, Regression Analysis by Example, 4th Edition, John Wiley & Sonc, Inc., Hoboken (2006).

[26] A.F. Siegel, Practical Business Statistics, IRWIN, 3rd edition (1997).

[27] P. Besse, Pratique de la modélisation statistique, Publications du laboratoire de statistique et Probabilité (2003).

[28] R.D. Cook, S. Weisberg, An introduction to Regression Graphics, Wiley Series in probability and mathematical statistics (1994).

[29] A. Golbraikh, A. Tropsha, *Beware of q2!*, Journal of Molecular Graphics and Modelling 20 (2002) 269-276.

[30] L. Erikson, J. Jaworska, A. Worth, M. Cromin, R.M. Mc Dowell, P. Gramatica., *Methods for reliability, uncertainty assessment, and applicability evaluations of regression based and classification QSPRs*, Environmental Health Perspective 111 (2003) 1361-1375.

[31] T.R. Milano, *Chemometrics and QSAR Research Group,* University of Milano Bicocca, Milano, Italy (personal communication) (2010).

[32] P.P. Roy, S. Paul, I. Mitra, K. Roy, *On Two Novel Parameters for Validation of Predictive QSAR Models*, Molecules 14 (2009) 1660-1701.

[33] V. Consonni, D. Ballabio, R. Todeschini, *Evaluation of model predictive ability by external validation techniques,* J. Chemometrics 24 (2010), 194-201.

[34] K. Roy, I. Mitra, S. Kar, P.K. Ojha, R.N. Das, H. Kabir H., *Comparative studies on some metrics for external validation of QSPR models,* J. Chem. Inf. Model 52 (2012) 396-408.

[35] A. Tropsha, P. Gramatica, V.K. Gombar, *The Importance of Being Earnest: Validation Is the Absolute Essential for Successful Application and Interpretation of QSPR Models*, QSAR & Comb. Sc. 22 (2003) 69-77.

[36] O. Ouattara, N. Ziao, *Quantum Chemistry Prediction of Molecular Lipophilicity Using Semi-Empirical AM1 and Ab Initio HF/6-311++G Levels*, Computational Chemistry 5 (2017) 38-50.

[37] P. Gramatica, *Principles of QSAR models validation: internal & external*, QSAR and Combinatorial Sciences 26 (2007) 694–701.

[38] T.I. Netzeva, A.P. Worth, T. Aldenberg, R. Benigni, M.T.D. Cronin, P. Gramatica, J.S. Jaworska, S. Kahn, G. Klopman, C.A. Marchant, G. Myatt, N. Nikolova-Jeliazkova, G.Y. Patlewicz, R. Perkins, D.W. Roberts, T.W. Schultz, D.T. Stanton, J.J.M. Van De Sandt, W.Tong, G. Veith, C. Yang, *Current status of methods for defining the applicability domain of (quantitative) structure–activity relationships*, Alternatives to Laboratory Animals 33 (2005) 155–173.

[39] S.S. Shapiro, M.B. Wilk, *An analysis of variance test for normality (Complete Samples)*, Biometrika 52 (1965) 591-611

[40] J. Durbin, G.S. Waston, *Testing for serial correlation in Least Squares Regression*, Biometrika 38 (1951) 159-179.

[41] J.G. MacKinnon, Durbin-Watson Statistic. The New Palgrave Dictionary of Economics, Palgrave Macmillan (ed.) (2008) 1-3.

[42] S. Chattefuee, A.S. Hadi, Regression analysis by example, fourth edition (2006).

[43] D. Laffly, Régression multiple : principes et exemples d'application (2006).

[44] G. Saporta, Probabilité, Analyse de données et statistique, 2ème Edition, Edition Technip. (2006).