Quantitative Structure antifungal Activity Relationship (QSAR) study of a series of Schiff bases derivatives from4-aminobenzenesulphonamide by DFT method

[1]N'dri Jean Stéphane, [1]Koné M. Guy-Richard, [1,2*]Kodjo Charles Guillaume, [1]Affi Sopi Thomas, [3]Kablan Ahmont Landry Claude, [2]Ouattara Zana Adama and [1]Ziao Nahossé

[1]Laboratoire de Thermodynamique et de Physico-Chimie du Milieu, UFR SFA, Université Nangui Abrogoua 02 BP 801 Abidjan 02, Côte-d'Ivoire,
[2]Laboratoire de Chimie BioOrganique et de Substances Naturelles, Université Nangui Abrogoua, UFR-SFA, 02 B.P. 801 Abidjan 02 Côte-d'Ivoire,
[3]UFR des Sciences Biologiques, Université Péléforo Gon Coulibaly de Korhogo, BP 1328 Korhogo, Côte d'Ivoire

ABSTRACT : In this work we will focus on a series of Schiff bases derived from 4-aminobenzenesulphonamide. By implementing quantum chemistry methods, at B3LYP/6-311 G (d, p) level, we have identified two QSAR models from quantum descriptors and antifungal activities against Candida albicans and Aspergillus niger strains that were accredited of good statistical performance. For model 1 statistical indicators were: correlation coefficient $\mathbf{R^2}$= 0.984, standard deviation $\mathbf{S}$=0.514 , Fischer test $\mathbf{F}$=90.753 and the correlation coefficient of cross-validation $\mathbf{Q^2_{CV}}$ = 0.983. Wereas the statistical datas of model 2 were: correlation coefficient $\mathbf{R^2}$=0.952, standard deviation $\mathbf{S}$=0.596, Fischer test F=79.022 and the correlation coefficient of cross-validation $\mathbf{Q^2_{CV}}$ = 0.951. These two models showed that quantum descriptors namely dipole moment and index of electrophilie were the bases for antifungal activity of these Schiff bases. These models were validated with the criteria of acceptance Eriksson et al. for training set and Tropsha et al. for the validation set.

Keywords: Schiff base, Quantum chemical descriptors, QSAR, Antifungal activity

I. INTRODUCTION

Fungus of the genus *Candida* cause serious illnesses those frequency remains constant despite the development of new therapeutic means[1], especially in immune compromised patients[2]. Among the *Candida* kind, 54.3% of infections are due to *albicans* species[3]. The thrush, due to fungus of the genus *Candida*, are the most common opportunistic infections, and their frequency has doubled between the years 80 and 90. Indeed, they represent more than 80% of infections fungus. Thus, *Candida Albicans* is responsible of infections those frequency and severity, are at the fore front of the fungal infections[4-7]. *Aspergillus* are most of the time of respiratory pathogens[8]. For example, the genus *Aspergillus Niger* is the cause of mycoses such as keratitis, otomycoses, onyxis, cutaneous lesions but also otitis and sinusitis. It is rarely found in the immunodeprime, where it is responsible for skin, lung or widespread infections [8]. To combat these fungi, several biologically active molecules including mainly Schiff bases derived from 4-aminobenzenesulphonamide are used. In the specific case of the QSAR study, a series of nine derived from 4-aminobenzenesulphonamide Schiff bases have been used among them six were used to training set and the other three remaining external validation set (**Figure 1**). These Schiff bases derived have been synthesized and tested by Santosh et al [9]. for their antifungal activities. These molecules were synthesized from various aromatic aldehydes in ethanol in the presence of acetic acid as a catalyst to get the Schiff bases (**a**-**i**). These derived Schiff bases have been developed to fight against the fungi *Aspergillus niger* and *Candida albicans*. However, for more than two decades, the permanent appearance of new breeds of fungi, resistant, worried more and more agricultural, medical and health communities. It appears necessary to develop an antifungal that offers a broad spectrum of action, a high degree of efficiency, a good margin of safety and flexibility of use in order to identify the problems of resistance. QSAR study is used to determine models based on the quantum descriptors. The QSAR study is more and more used, due to the growth of the means of calculations, to explain or predict the properties of molecular in order to limit the excessive number of experiences, sometimes lengthy, expensive and reduce the cost of the production of medicines by pharmaceutical companies[10-11]. The general objective of this work is to make descriptive and predictive study of antifungal activity. The obtained models will identify the best analogues called 'touches' that will be able to fight more effectively against fungi *Aspergillus niger* and *Candida albicans*.
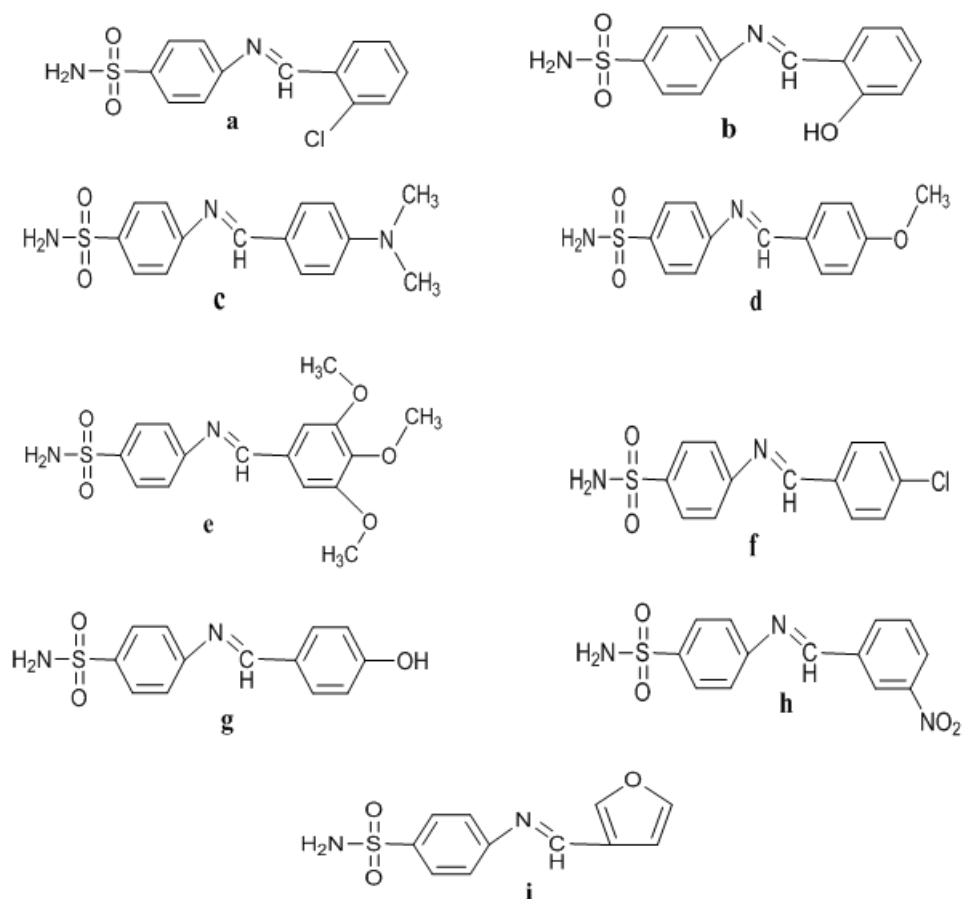
Figure1: Molecular structures of test and validation of the Schiff bases used for QSAR models

II. MATERIALS AND METHODS

Computational details

In order to find a link between the antifungal activity values of the studied molecules and their molecular structures, quantum chemistry calculations were made using Gaussian 03[12]. The DFT methods were generally considered to generate a variety of molecular properties in QSAR studies that increase predictability [13-20], reduce the computing time and costs involved in the design of new drugs [21-22]. The B3LYP/6-311 G (d, p) theoretical level has been used to determine the molecular descriptors. Concerning the split-valence and triple-dzeta bases choices, they were sufficiently large and taking into account polarization functions was important in the quantification of molecular descriptors obtained. Modelling was done using the method of multilinear regression implemented in *Excel* spreadsheets and *XLSTAT* [23-24].

Quantum Descriptors

For the development of QSAR models, some theoretical descriptors related to the conceptual DFT were determined. In particular, the HOMO (Higher Occupied Molecular Orbital) energy, the LUMO (Lower Unoccupied Molecular Orbital) energy, the chemical electonegativity ($\square$), chemical hardness ($\square$), chemical softness ($\square$), the electrophilicity index ($\square$), the total exchange energy ($\square E_T$) and the dipole moment (μ). These descriptors were all based on the optimized molecules. It should be noted that frontier molecular orbitals-related descriptors have been calculated very simply in the context of the approximation of Koopmans [25]. The LUMO energy characterizes the sensitivity of the molecule to a nucleophilic attack, and concerning HOMO energy, it characterized the susceptibility of a molecule in an electrophilic attack. The electronegativity was the parameter which reflects the ability of a molecule to not let escape its electrons.The chemical hardness ($\square$) and overall softness ($\square$) express the resistance of a system to change its number of electrons. The parameter which reflects the ability of the molecule to accept electrons was the electrophilie index ($\square$). The other studied parameter, the total energy of exchange ($\square E_T$) provides information on the existence of mutual exchange of electrons between the studied molecules. The dipole moment   indicates the stability of a molecule in water. Thus, a strong dipole

moment translated a low solubility in organic solvents and strong solubility in water. These parameters, except the dipole moment were calculated from equations (1):

$$\chi = -1/2\,(\varepsilon_{LUMO} + \varepsilon_{HOMO})$$

$$\eta = (E_{LUMO} + E_{HOMO})/2$$

$$\sigma = 1/\eta \qquad (1)$$

$$\omega = \chi^2/2\eta \qquad \Delta E_T = -\eta/4$$

Estimation of the Predictive Ability of QSAR Model

The six molecules of the training set and the three other validation set molecules used in this study, had various diameters of inhibition from 11 to 22 mm. This range of activities allowed us to define a quantitative relationship between the diameters of inhibition and theoretical descriptors. The quality of a model was determined by taking account many statistic criteria of analysis such as the correlation coefficient R^2, the standard deviation **S**, the correlation coefficientof cross-validation Q_{CV}^2 and Fischer **F**. R^2, **S** and **F** relate to the adjustment of the calculated and experimental values: they describe the predictive ability within the limit of model and allow to estimate the accuracy of the values calculated on the learning game [26-27]. As the correlation coefficient cross-validation Q_{CV}^2, it provided information on the power of the model forecast. This predictive power was said to be 'internal' because it was calculated from the structures used to build this model.
The squared correlation coefficient gave an evaluation of the dispersion of the theoretical values around the experimental values. The quality of modeling was better when the points were close to the fitted line [28].The adjustment of points in this straight line can be evaluated by the correlation coefficient. The correlation coefficient R^2 was given by the following equation (2):

$$R^2 = 1 - \frac{\Sigma(y_{i,exp} - \hat{y}_{i,theo})^2}{\Sigma(y_{i,exp} - \bar{y}_{i,exp})^2} \qquad (2)$$

Where:
$y_{i,exp}$: The experimental value of the inhibition diameter
$\hat{y}_{i,theo}$: The theoretical value of the inhibition diameter
$\bar{y}_{i,exp}$: The average valueof the experimental values of the inhibition diameter

More the R^2 value will be closer to 1 more the theoretical and experimental values were correlated.In addition, the variance σ^2 was determined by the relationship 3:

$$\sigma^2 = s^2 = \frac{\Sigma(y_{i,exp} - y_{i,theo})^2}{n - k - 1} \qquad (3)$$

Where **k** was the number of independent descriptors, **n** was the number of molecules of the training set and **n-k-1** was the degree of freedom.
The root mean square error S was another statistical indicator used. It allows to evaluate the reliability and accuracy of a model. It was obtained from equation (4):

$$s = \sqrt{\frac{\Sigma(y_{i,exp} - y_{i,theo})^2}{n - k - 1}} \qquad (4)$$

The Fisher test **F** was also used to measure the level of statistical significance of the model, i.e. quality of the choice of descriptors constituting the model.The Fisher test **F** was defined from equation (5):

$$F = \frac{\Sigma(y_{i,theo} - y_{i,exp})^2}{\Sigma(y_{i,exp} - y_{i,theo})^2} * \frac{n - k - 1}{k} \qquad (5)$$

The correlation coefficientof cross-validation Q_{cv}^2, to assess the accuracy of the prediction on the training set was calculated using the following relationship:

$$Q_{cv}^2 = \frac{\sum(y_{i,theo} - \bar{y}_{i,exp})^2 - \sum(y_{i,theo} - y_{i,exp})^2}{\sum(y_{i,theo} - \bar{y}_{i,exp})^2} \quad (6)$$

The performance of a mathematical model, for Eriksson et *al.* [29], was characterized by a value of $Q_{cv}^2 > 0.5$ for a satisfactory model and for an excellent model when $Q_{cv}^2 > 0.9$. According to them, for a given training set, a model will be performant if the acceptance criterion $R^2 - Q_{cv}^2 < 0.3$ was respected.

Moreover, the prediction power of a model can be obtained from five Tropsha's criteria[30-31], if at least three of the criteria were satisfied, then the model will be considered efficient in predicting the activity studied. These criteria were the following:

$$\textbf{1)} \; R_{Test}^2 > 0.7 \qquad \textbf{2)} \; Q_{CvTest}^2 > 0.6 \; , \quad \textbf{3)} \; |R_{Test}^2 - R_0^2| \leq 0.3 \; ,$$

$$\textbf{4)} \; \frac{|R_{Test}^2 - R_0^2|}{R_{Test}^2} < 0.1 \; \text{ and } \; 0.85 \leq k \leq 1.15, \quad \textbf{5)} \; \frac{|R_{Test}^2 - R_0^2|}{R_{Test}^2} < 0.1 \; \text{ and } \; 0.85 \leq k' \leq 1.15$$

III. RESULTS AND DISCUSSION

The trainnig set of the six (6) molecules of Schiff bases derived from 4-aminobenzenesulphonamide and the three (3) molecules of the validation set were shown in **Table 1**.

Table 1: Quantum descriptors and experimental antifungal activities of the training and validation set

Code	χ (eV)	η (eV)	σ (eV^{-1})	ω(eV)	ΔE_T(eV)	μ (Debye)	*A. niger* Inhibition Diameters (mm)	*C. albicans* Inhibition Diameters (mm)
a	4.599	4.267	0.234	2.479	-1.067	6.208	14	12
b	4.272	4.271	0.234	2.136	-1.068	5.840	15	13
c	3.694	3.826	0.261	1.784	-0.957	5.902	13	16
d	4.121	4.227	0.237	2.009	-1.057	6.400	15	11
e	4.293	4.279	0.234	2.154	-1.070	10.137	15	20
f	4.565	4.292	0.233	2.428	-1.073	5.487	19	13
g	4.191	4.273	0.234	2.055	-1.068	5.706	12	14
h	4.990	4.005	0.250	3.108	-1.001	9.269	22	20
i	4.312	4.223	0.237	2.201	-1.056	3.861	17	18

Quantitative Structure Activity Relationship (QSAR) Validation Model

It should be noted that the negative or positive sign of the coefficient of a descriptor of the model reflected the effect of proportionality between the evolution of the biological activity of interest and the parameter of the regression equation. Thus, the negative sign indicated that when the value of the descriptor was high, the biological activity decreased while the positive sign meant the opposite effect.

Model 1

The model equation obtained by correlating the descriptors theoretical of Schiff bases derived from 4-aminobenzenesulphonamide optimized and inhibition diameter of *Aspergillus niger*. The statistical indicators were presented below:

$$d = 1.446 + 7.301*\omega - 0.228*\mu \quad (7)$$

$$N = 6 \qquad R^2 = 0.984 \quad Q_{CV}^2 = 0.983 \qquad S = 0.514 \qquad F = 90.753$$

The negative and positive signs of the coefficient of the electrophilie index of the dipole moment respectively translated that the diameter inhibition of *Aspergillus niger* will be improved for a high value of the electrophilie index and for a low value of the dipole moment. The significance of this model was given by a Fisher F-test **F** = 90.753:For this model, the correlation coefficient of cross-validation Q_{CV}^2 equals 0.983. This value greater than 0.9 indicated a model known as great according Erikson *et al.* [29]. This model was acceptable as it was consistent with the criteria of acceptance of the authors: $R^2 - Q_{CV}^2 = 0.984 - 0.983 = 0.001 < 0.3$. External validation of this model was made with the compound **a, f** and **g**. **Figure 2**, illustrate the

correlation between experimental and theoretical diameters of inhibition of the training set (blue points) and validation set (red points).
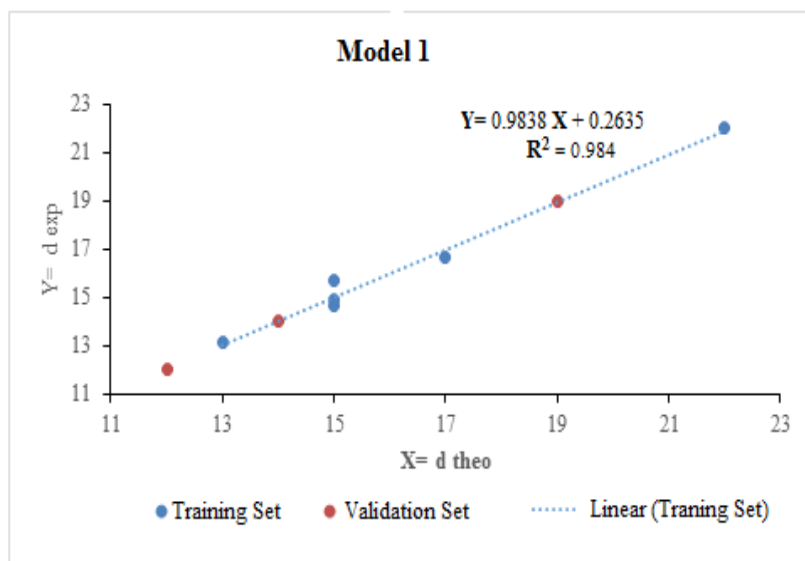

Figure 2: Plot of correlation equation between the experimental and theoretical diameters of inhibition of Model

Figure 2 shows that there was a strong correlation between the experimental and theoretical diameters of inhibition according Model 1

Verification of the criteria of Tropsha for model 1 with external validation

$$R^2_{Test} = 1 > 0.7 \qquad Q^2_{CvTest} = 0.999 > 0.6 \qquad |R^2_{Test} - R^2_0| = 0.00 \leq 0,3$$

$$\frac{|R^2_{Test} - R^2_0|}{R^2_{Test}} = 0.00 < 0.1 \text{ and } 0.85 \leq k = 1 \leq 1.15 \ ;$$

$$\frac{|R^2_{Test} - R'^2_0|}{R^2_{Test}} = 0.00 < 0.1 \text{ and } 0.85 \leq k' = 1 \leq 1.15$$

All values satisfy Tropsha's criteria. Model 1 was retained as predictive model of antifungal activity (*Aspergillus niger)*.

Model 2
As for model 2, the model obtained by correlating the theoretical descriptors of the databases of Schiff bases derived from 4-aminobenzenesulphonamide optimized and diameter of inhibition of *Candida albicans* whose statistical indicators were presented below:

$$d = 28.083 - 2.611 * \mu \quad (8)$$

$$N = 6 \qquad \mathbf{R^2} = 0.952 \qquad \mathbf{Q^2_{CV}} = 0.951 \qquad \mathbf{S = 0.596} \qquad \mathbf{F = 79.022}$$

The negative coefficient of the dipole moment here indicated that of the inhibition diameter of *Candida Albicans* will be improved for any molecule of the series of Schiff bases derived from 4-aminobenzenesulphonamide which will have a low value of the dipolar moment. The significance of the model was translated by the Fischer factor **F**= 79.022: the correlation coefficient of cross-validation $\mathbf{Q^2_{CV}}$= 0.951. This value greater than 0.9 indicates a model known as great according Erikson *et al.*.[29] This model was acceptable with $\mathbf{R^2} - \mathbf{Q^2_{CV}} = 0.952 - 0.951 = 0.001 < 0.3$. External validation of this model was made with the compound **c, e** and **h. Figure 3** represents the correlation between experimental and theoretical diameters ofinhibition of the training set (blue points) and validation set (red points).
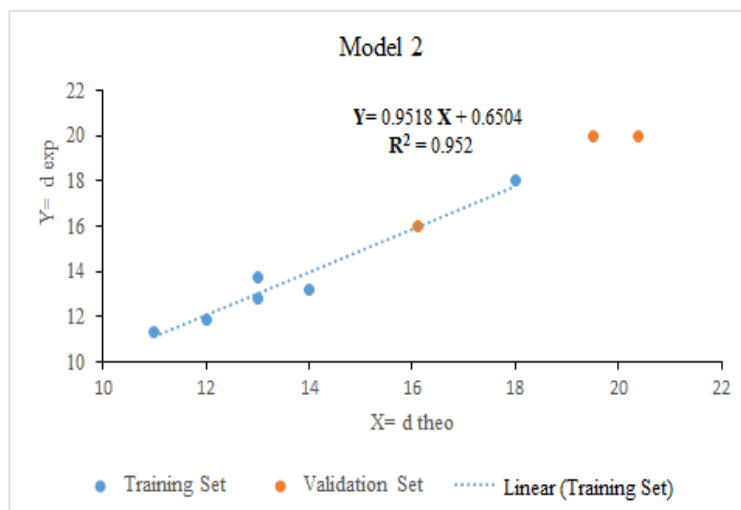
Figure 3: Plot of correlation equation between the experimental and theoreticaldiameters of inhibition of Model 2

Figure 3 shows that there was a strong correlation between the experimental and theoretical diameters of inhibition according Model 2.

Verification of the criteria of Tropsha for model 2 with external validation

$$R^2_{Test} = 0.962 > 0.7 \qquad Q^2_{CvTest} = 0.961 > 0.6 \qquad |R^2_{Test} - R^2_0| = 0.033 \leq 0.3$$

$$\frac{|R^2_{Test} - R^2_0|}{R^2_{Test}} = 0.034 < 0.1 \text{ and } 0.85 \leq k = 0.999 \leq 1.15,$$

$$\frac{|R^2_{Test} - R'^2_0|}{R^2_{Test}} = 0.040 < 0.1 \text{ and } 0.85 \leq k' = 1 \leq 1.15$$

All values satisfy Tropsha's criteria. Model 2 was retained as predictive model of antifungal activity (*Candida albicans)*.

IV. CONCLUSION

This study highlighted the relationship between the antifungal activity (*Candida albicans* and *Aspergillus niger*) and quantum descriptors of the molecules studied. The descriptors of the optimized molecules were the dipole moment (**μ**) and electrophilie index (□). These parameters allowed to predict the behavior of Schiff bases studied in the presence of fungi *Candida albicans* and *Aspergillus niger*. There were strong correlations between the calculated and the experimental diameter of inhibition. The QSAR models obtained allowed to predict the activity of the best analogues called lead. These proposed models permit to identify the descriptors that enhance antifungal activity giving guidance to design new molecules more active against fungi *Candida albicans* and *Aspergillus niger*. The robustness of these models has been verified with the help of avalidation set including three molecules.

REFERENCES
[1]   M. A. Pfaller and D. J. Diekema,Clin. Microbiol Rev, 20, 2007, 133-163.
[2]   S. Benedict and J. Colagreco,Cancer Nurs, 17, 1994, 411-417.
[3]   M. A. Pfaller, R. N. Jones, G. V. Doern, H. S. Sader, S. A. Messer, A. Houston, S. Coffman and R J. Hollis,Antimicrob Agents Chemother , 44,2000, 747-751.
[4]   G. P. Bodey, M. Mardani, H. A. Hanna, M. Boktour, J. Abbas, E. Girgawy, R. Y. Hachem, D. P. Kontoyiannis and I.I. Raad, Am J Med ,112, 2002, 380-385.
[5]   C. A. Hage, M. Goldman and L. J. Wheat, Eur J Med Res , 7,2002, 236-241.
[6]   R. Martino and M. Subira, Ann Hematol , 81,2002, 233-243.
[7]   L. P. Samaranayake, P. L. Fidel, J. R. Naglik, S. P. Sweet, R. Teanpaisan, M. M. Coogan, E. Blignaut and P. Wanzala, Oral Dis ,8(2), 2002, 151-160.
[8]   N. Quatresous, aspergillose humaine. epidemiologie, diagnostic biologique, contrôle, 2011,14-21.

[9]   K. Santosh, M. S. Niranjan, K. C. Chaluvaraju, C. M. Jamakhandi. And K. Dayanand, Journal of Current Pharmaceutical Research, 2010.

[10]   T. I. Oprea, (Ed.) (WILEY-VCH Verlag, Germany, 2005).

[11]   E. A. Rekka, P. N. Kourounakis, (Ed) (Taylor & Francis Group, LLC., USA, 2008).

[12]   Gaussian 03, Revision B.05, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani,N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima,Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K Raghavachari, J B Foresman, J V Ortiz, Q Cui, A G Baboul, S Clifford, J Cioslowski, B BStefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, Gaussian, Inc., 2003, P.A. Pittsburgh.

[13]   P. K. Chattaraj, A. Cedillo and R. G. Parr, J. Phys. Chem.,103, 1991, 7645

[14]   P. W. Ayers and R. G. Parr, 2000, J. Am Chem., Soc, 122, 2010, 2000.

[15]   F. De Proft, J. M. L. Martin, P. Geerlings, Chem. Phys Let.,250, 1996, 393.

[16]   P. Geerlings, F. De Proft, J. M.L. Martin, J. Ed.; Elsevier; Amsterdam, 4, 1996, 773.

[17]   F. De Proft, J. M. L. Martin, P. Geerlings, Chem. Phys Let. , 256, 1996, 400.

[18]   F. De Proft, P. Geerlings, J Chem Phys, 106, 1997, 3270,

[19]   P. Geerlings, F. De Proft, W. Langenaeker, Adv. Quantum Chem., 33, 1996, 303.

[20]   R. G. Parr, R. A. Donnelly, M. Levy, W. E. Palke, J. Chem. Phys.,68, 1978, 3801,

[21]   C. Hansch, P. G. Sammes, and J. B. Taylor, (Eds.) (Pergamon Press, Oxford, 4,1990), 33-58,.

[22]   R. Franke, Theoretical Drug Design Methods, (Elsevier, Amsterdam, 1984).

[23]   Microsoft ® Excel ® 2013 (15.0.4420.1017) MSO (15.0.4420.1017) 64 Bits , Partie de Microsoft Office Professionnel Plus 2013.

[24]   XLSTAT Version 2014.5.03 Copyright Addinsoft 1995-2014, XLSTAT and Addinsoft are Registered Trademarks of Addinsoft 2014.

[25]   T. Koopmans, Úber die Zuordnung von Wellenfunktiomen und Eigenwerten zu den einzelnen Elektronen eines Atoms, Physica, 1, 1934, 104-113,

[26]   G. W. Snedecor, W. G. Cochran, Statistical Methods, (Oxford and IBH, New Delhi, India,1967), 381.

[27]   M. V. Diudea, Nova Science, (Huntingdon , New York, USA, 2000)

[28]   E. X. Esposito, A. J. Hopfinger, J. D. Madura, Methods in Molecular Biology, 275, 2004, 131-213.

[29]   L. Eriksson, J. Jaworska, A. Worth, M. T. D. Cronin, R. M. Mc Dowell, P. Gramatica, Environmental Health Perspectives, 111(10), 2003,1361-1375.

[30]   A. Golbraikh, A. Tropsha, Beware of qsar, J. Mol. Graph. Model, 20, 2002, 269.

[31]   (a) 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. Sci., 22, 2003, 69. (b) A. Tropsha, Best practices for QSAR model development, validation, and exploitation. Molecular Informatics, 29, 2010, 476.