Predictive Modeling of Breast Anticancer Activity of a Series of Coumarin Derivatives Using Quantum Descriptors

ABSTRACT

We focused on a series of coumarin derivatives in this work. The method of Density Functional Theory (DFT) of quantum chemistry has been used at B3LYP / 6-31G (d, p) level in order to identify molecular descriptors which are useful for this study. The analysis of the statistical indicators allowed to obtain a QSAR model based on quantum descriptors and anti-cancer activity against breast cancer (MCF-7) that were accredited for good statistical performance. For the model, the statistical indicators were: correlation coefficient R^2 = 0.904, standard deviation S = 0.102, Fischer test coefficient F = 18.779 and correlation coefficient of cross validation Q^2_{cv}= 0.894. This model has shown that the quantum descriptors namely the dipole moment, the energetic gap, the distance d(c-c) and the dihedral angle D(o=c-c-h) are at the origin of the anticancer activity of these coumarinic compounds. The validation of this model has been done according to Eriksson et al.'s acceptance criteria for the training set and according to Tropsha et al.'s five criteria for the validation set. The last used tool of validation is the ratio of theoretical and experimental activities values of the validation set which must tend towards the unit.

Keywords *anticancer activity, coumarin derivative, quantum descriptors, MCF-7, QSAR*

1. INTRODUCTION

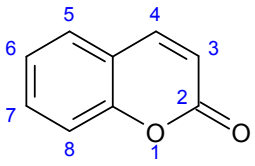
Cancer is the second biggest killer after cardiovascular diseases. Forty percent of cancer deaths can be avoided by prevention [1]. It is estimated that the global burden of cancer has risen to 18.1 million new cases and 9.6 million deaths in 2018. One out of five men and one out of six women worldwide will develop cancer in their lifetime. It should be noted that 1 out of 8 men and 1 out of 11 women die from this disease. Globally, an estimation of 43.8 million of people will live with a cancer diagnosis in the next five years [2]. Regarding women, breast cancer is placed in second position after uterus' one. In addition to medical treatments with products that have many undesirable and often destructive side effects, it has been wise to find more effective natural or synthetic molecules to fight against cancer and its side effects.

Coumarins have become unavoidable thanks to their various therapeutic properties such as anti-cancer, anti-tumor, anti-HIV and many others [3]. Coumarin was extracted for the first time in 1820 from the tonka bean (Dipteryx odorata Willd., Fabaceae). The word coumarin is the deformation of the name coumarou [4]. The coumarin molecule has 9 carbon atoms and 2 oxygen atoms. It is a member of the benzopyrone family. Benzopyrone molecules are formed by the junction of a benzene and a pyrone rings. This family is subdivided into two sub-families, namely benzo-α-pyrones and benzo-γ-pyrones. Coumarins come from the family of benzo-α-pyrones while flavonoids come from the family of benzo-γ-pyrones [5].

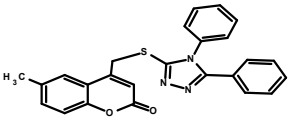
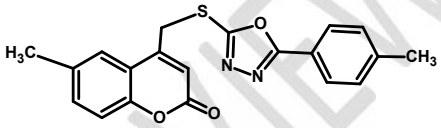
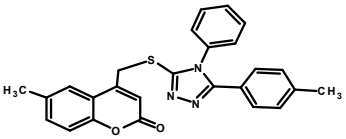
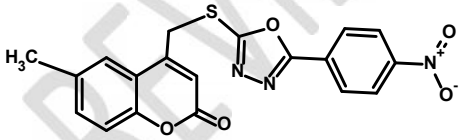
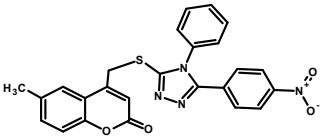
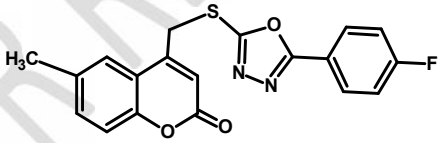
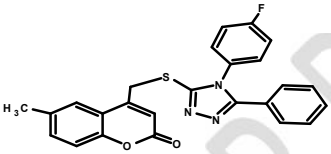
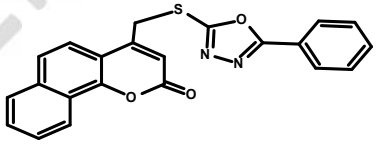
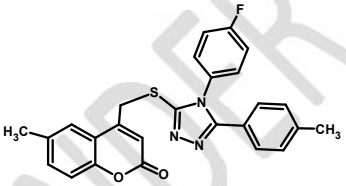
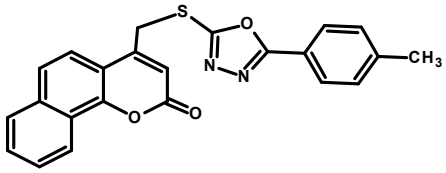
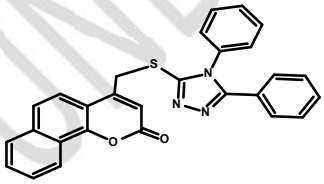
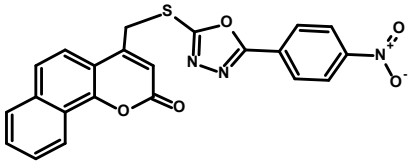
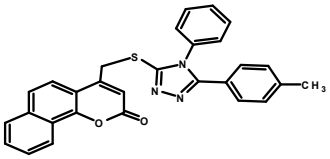
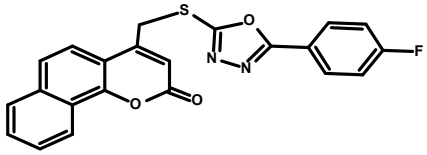
Studied molecules in this work are obtained from coumarin core and all the molecules are obtained by the substitution in C4 position through the thio-methyl linker. Coumarin core and the studied molecules are presented in Figure 1 below.

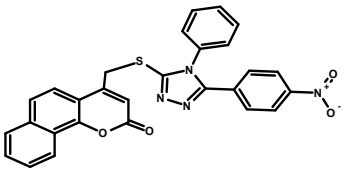
The study of Quantitative Structure Activity Relationship (QSAR) is one of the best methods to design new therapeutic agents. It allows to correlate quantitatively the mathematical models of the structures of the compounds and their biological activity. QSAR is a highly solicited technique because it helps to reduce the number of experiments that are sometimes long, dangerous and expensive in terms of time and finance [6-9]. The present manuscript is written in order to implement a QSAR model in the treatment of breast cancer in women. All this contributes to the reduction of medicines' production costs [10-11] and contributes to the protection of the environment. This QSAR approach has its origins in the studies carried out, on the one hand by HanschL [12] and on the other hand by Free and Wilson

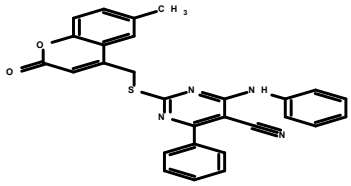
[13]. Indeed, Hansch has established models relating biological activity with the hydrophobic, electronic and steric properties of molecules. In general, the QSAR model is based on the fifth (1/5) of the initial database. The general aim of this work is to develop reliable model to explain and predict the anti-cancer activity MCF-7 from a series of twenty (20) coumarin derivatives (Figure 1). These compounds were synthesized and tested by Morsy et al. [14] for their biological activities.

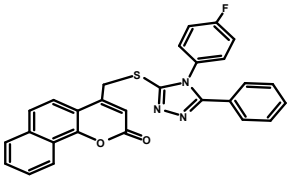


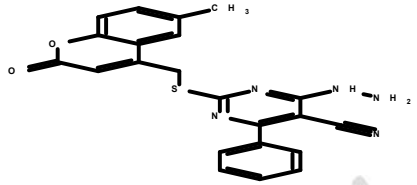
Coumarin core

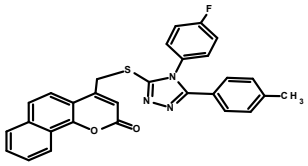
Code	2D structure of the molecules	code	2D structure of the molecules
1		11	
2		12	
3		13	
4		14	
5		15	
6		16	
7		17	

8 

18 

9 

19 

10 

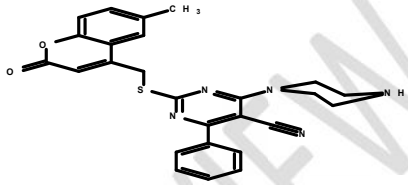
20 

Figure 1: Molecular structures of coumarin core and the studied coumarin derivatives

2. MATERIALS AND METHODS

2.1. Data Source

A test for cytotoxic activity against human breast cancer cell line (MCF-7) using MTT (Methyl Thiazolyl Tetrazolium) assay according to the method of Mosmann allows to obtain experimental values [15]. The twenty (20) studied molecules have median Inhibitory Concentration (IC50) ranging from 6.9 to 83.8 µg / mL. IC50 means the median concentration of molecules determined experimentally to inhibit 50% of cancer cells in a population of cancer cells [16]. Biological data are generally expressed as the opposite of the decimal logarithm of activity $(-\log_{10}(C))$. The purpose of this transformation is to obtain better mathematical values when structures are biologically active [17,18]. The anticancer activity will be expressed by the anticancer potential p_{IC50} defined by equation (1):

$$p_{IC50} = \log_{10}\left(\frac{IC50}{M} * 10^{-3}\right) \qquad (1)$$

Where M is the molecular molar mass of the compound expressed in g / mol and IC50, the median Inhibitory Concentration expressed in µg/mL.

2.2. Computational Methods

The correlations between the values of the biological activities of the studied molecules and their molecular structures were obtained by quantum chemistry calculations carried out using the software Gaussian 09 and its interface GaussView 5.0 [19]. The DFT method is generally known to generate a variety of molecular properties [20-23] [16] [24-25] in QSAR studies. These generated properties increase the predictability of QSAR models while reducing computational time and cost implications in the design of new drugs [26,27]. The theory level B3LYP / 6-31G (d, p) was used to determine the molecular descriptors. The modeling was done using the linear regression method implemented in Excel spread sheet [28] and XLSTAT version 2014 [29].

2.3. Quantum Descriptors

The dipolar moment μ_D, the energy gap ΔE_{Gap}, the distance d(c-c) (distance between the two carbon atoms that join the benzenic and pyronic rings) and the dihedral angle D (O=C-C-H) formed by the oxygen hybridized sp^2, carbons atoms C_2-C_3 and hydrogen bonded to C_3 carbon atom were determined in the framework of the development of a QSAR model. These descriptors are all obtained from the optimized structures of the different molecules. The dipole moment (μ_D) indicates the state of molecule's stability in water. That means a compound with high dipole moment will have a low solubility in organic solvents and high solubility in water [6]. The energy gap (ΔE_{Gap}) gives information about the reactivity of a molecule because the lower the energy Gap, the more reactive the molecule is and has very low kinetics [30]. The energy gap is established from the following relation:

$$\Delta E_{Gap} = E_{LUMO} - E_{HOMO} \quad (2)$$

The geometric descriptors such as d(C-C) bond length expressed in Armstrong (Å) and the dihedral angle D(O=C-C-H) expressed in degree are used (**Figure 2**). The dihedral angle reflects the coplanarity of the atoms involved in its formation. When it tends towards zero then the atoms that form the dihedral angle are coplanar.

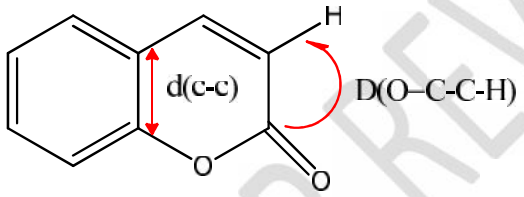


Figure 2: Used geometric descriptors of the coumarin ring: d (C-C) and D (O=C-C-H)

For all the studied descriptors, the analysis of the bivariate data, that is to say the calculation of the coefficient of the partial correlation between each pair of the set of descriptors, is less than 0.95 (the absolute value a_{ij}< 0.95), which means that these different descriptors are independent from each other [31-36].

2.4. Forecasting Power of a QSAR Model

The judgment of a model's quality is based on various statistical analysis criteria including the coefficient of determination R^2, the standard deviation S, the correlation coefficients of the cross validation Q^2_{CV} and Fischer test F. The quantities R^2, S and F relate to the adjustment of calculated and experimental values. They describe the ability to foretell in the boundaries of the model and permit to estimate the accuracy of the calculated values on the training set [37-39]. Concerning the cross-validation coefficient Q^2_{CV}, it gives information on the predictive power of the model. This forecasting power is qualified to be "internal" because it is calculated from the structures used to build this model.

The correlation coefficient R^2 tells how the theoretical values are spread around the experimental ones. The quality of the modeling is better when the points are close to the adjustment line **[40].** The adjustment of points to this line can be obtained by the coefficient of determination defined below

$$R^2 = 1 - \frac{\sum\left(y_{i,exp} - \hat{y}_{i,theo}\right)^2}{\sum\left(y_{i,exp} - \bar{y}_{i,exp}\right)^2} \quad (3)$$

Where:

$y_{i,exp}$: The experimental value of the anticancer activity

$\hat{y}_{i,theo}$: The theoretical value of the studied activity

$\bar{y}_{i,exp}$: The mean or average value of the experimental values of the anticancer activity

The more the value of **R^2** will be close to 1, the more theoretical and experimental values will be correlated (close to each other).

Moreover, the variance σ^2 is determined by the relation (4):

$$\sigma^2 = s^2 = \frac{\sum(y_{i,exp} - y_{i,theo})^2}{n-k-1} \quad (4)$$

Where *k* is the number of independent parameters (descriptors), *n* is the number of compounds (molecules) in the training set or learning set, and *n-k-1* is the degree of freedom.

The standard deviation **S** is another used statistical tool which indicates the reliability and the accuracy of a model:

$$s = \sqrt{\frac{\sum(y_{i,exp} - y_{i,theo})^2}{n-k-1}} \quad (5)$$

The level of statistical significance of the model is obtained through Fisher test **F coefficient** that is the appropriateness of the descriptors' choice constituting the model given in relation (6) below.

$$\mathbf{F} = \frac{\sum(y_{i,theo} - y_{i,exp})^2}{\sum(y_{i,exp} - y_{i,theo})^2} * \frac{n-k-1}{k} \quad (6)$$

The coefficient of determination of the cross validation $\mathbf{Q^2_{cv}}$ allows to evaluate the accuracy of the prediction on the training set. It is calculated using the following relation:

$$Q^2_{cv} = \frac{\sum(y_{i,theo} - \bar{y}_{i,exp})^2 - \sum(y_{i,theo} - y_{i,exp})^2}{\sum(y_{i,theo} - \bar{y}_{i,exp})^2} \quad (7)$$

The performance of a mathematical model, according to Eriksson et al. [41,42] is characterized by a value of $\mathbf{Q^2_{cv} > 0.5}$ for a satisfactory model when for the excellent model $\mathbf{Q^2_{cv} > 0.9}$. According to them, for a given training set, a model will perform well if the following acceptance criterion $\mathbf{R^2 - Q^2_{cv} < 0.3}$ is met. In addition, the foretelling power of a model can be obtained from the value of the $p_{MICtheo}$ / p_{MICexp} ratio for the validation set. The model is acceptable when the values of the ratio of theoretical and experimental activity are very close to unity. In addition, the predictive power of a model can be obtained from the five Tropsha criteria [43-44]. If at least three (3) of them are satisfied, then the model will be considered effective in predicting the studied activity. These criteria are determined from the validation set and are listed below:

1) $R^2_{Test} > 0.7$, **2**) $Q^2_{CvTest} > 0.6$, **3**) $|R^2_{Test} - R^2_0| \leq 0.3$,

4) $\frac{|R^2_{Test} - R^2_0|}{R^2_{Test}} < 0.1$ and $0.85 \leq k \leq 1.15$, **5**) $\frac{|R^2_{Test} - R'^2_0|}{R^2_{Test}} < 0{,}1$ *and* $0.85 \leq k' \leq 1.15$

3. RESULTS AND DISCUSSION

3.1. QSAR Models

Table 1 gathers the thirteen (13) molecules of the training set and the seven (07) molecules of the validation set. Moreover, the values of the partial correlation coefficients a_{ij} of descriptors are also indicated in Table 2.

Table 1: Descriptor values

	IC50(µg/mL)	pIC50	µ(D)	$\Delta_{E_{gap}(eV)}^{\mu\ Series}$	d(C-C) (°A)	D(O=C-C-H) (°)
			Training Series			
1	**47.7**	**1.321**	6.600	4.346	1.407	-0.380
2	**63.5**	**1.197**	6.563	4.317	1.407	-0.415
5	**29.7**	**1.527**	5.233	4.339	1.407	-0.374
7	**83.8**	**1.077**	8.494	3.944	1.391	1.353
9	**40.5**	**1.393**	7.184	3.935	1.391	1.252
10	**58.7**	**1.231**	7.130	3.938	1.390	1.364
11	**21.1**	**1.676**	4.508	4.252	1.407	0.415
12	**39.7**	**1.401**	8.193	3.654	1.407	0.121
13	**16**	**1.796**	5.232	4.310	1.407	-0.258
14	**33.6**	**1.474**	5.575	3.964	1.394	1.522
17	**19.8**	**1.703**	5.644	3.960	1.394	1.324
18	**6.9**	**2.161**	4.285	4.062	1.407	0.088
20	**10.9**	**1.963**	5.648	4.205	1.407	-1.421
			Validation series			
3	**22.1**	**1.656**	9.363	3.718	1.407	-0.970
4	**69.2**	**1.160**	5.200	4.334	1.407	-0.575
6	**13.2**	**1.879**	8.598	3.944	1.390	1.380
8	**9**	**2.046**	10.687	3.263	1.391	1.172
15	**51**	**1.292**	5.602	3.969	1.207	1.363
16	**81.5**	**1.089**	7.684	3.185	1.394	1.208
19	**49**	**1.310**	5.071	4.124	1.408	-1.719

Table 2: Descriptor Correlation Matrix

Variables	µ(Debye)	$\Delta_{E_{gap}(eV)}^{\mu\ Debye)}$	d(C-C) (A°)	D(O=C-C-H) (°)	pIC50
µ(Debye)	**1.000**				
$\Delta_{E_{gap}(eV)}^{\mu\ Debye)}$	-0.557	**1.000**			
d(C-C) (A°)	-0.466	0.547	**1.000**		
D(O=C-C-H) (°)	0.305	-0.610	-0.876	**1.000**	
pIC50	-0.788	0.193	0.470	-0.403	**1.000**

The partial correlation coefficients a_{ij} contained in Table 2 between the pairs of descriptors (μ, ΔE_{Gap}), (μ, d (C-C)), (μ, d (O = C-C-H)), (ΔE_{Gap}, d (C-C)) , (ΔE_{Gap}, D (O=C-C-H)) and (d (C-C), d (O=C-C-H)) are all less than 0.95 (a_{ij}<0.95), which means that the descriptors used in this model are independent from each other.

3.2. QSAR Model's validation

It should be noted that the negative or positive sign of a descriptor's coefficient in the model's equation reflects the effect of proportionality between the evolution of the biological activity and the concerned parameter. Thus, the negative sign indicates that the descriptor and the biological activity evolve inversely while the positive sign reflects the opposite effect. The statistical indicators obtained for QSAR model of the anticancer activity are given in Table 3.

The equation and the obtained statistical indicators are presented below in Table 3:

pIC_{50}^{exp} = 31.79-0.27*μD -1.09*ΔE_{Gap}- 17.14*d(C-C) – 0.30*D (O=C-C-H)

Table 3: Statistical indicators determined from the training set

Statistical Indicators of Multilinear Regression	Model
Number of compounds **n**	13
Coefficient of correlation of the $\mathbf{R^2}$ regression line	0.904
Standard deviation **S**	0.102
Validation of Fischer **F**	18.779
Coefficient of correlation of cross validation **Qcv**	0.894
Confidence level **α**	> 95%

Negative signs of dipole moment (μ_D), energy Gap (ΔE_{Gap}), distance $d_{(C-C)}$ and dihedral angle $D_{(O = CCH)}$ indicate that anticancer activity will be improved for low values of these quantum descriptors.

The significance of the model is expressed by the correlation coefficient of the cross validation $\mathbf{Q^2_{cv} = 0.894 > 0.5}$. Moreover, this model is acceptable with $\mathbf{R^2\text{-}Q^2_{cv} = 0.01 < 0.3}$.

The regression line between the experimental and theoretical anticancer activities of the training set (blue dots) and the validation set (red dots) is shown in **Figure 3**.

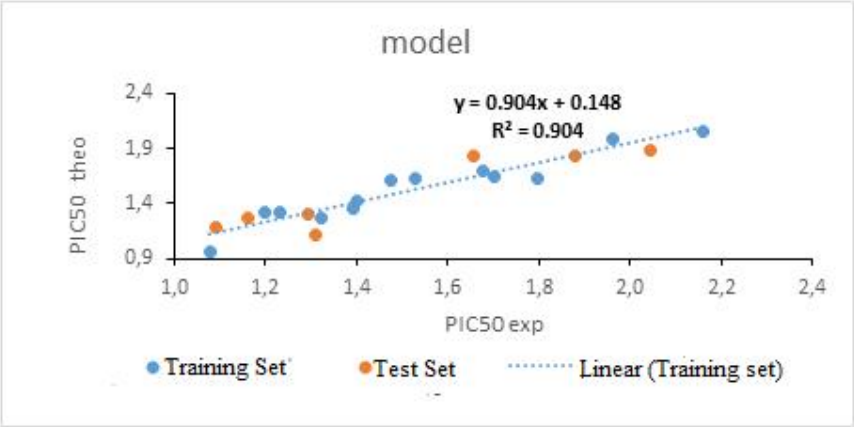


Figure 3: Model regression line

Furthermore, the external validation of the model was performed with the compounds of the validation set (3, 4, 6, 8, 15, 16, 19). These molecules which are used for external validation meet the five criteria of Tropsha.

Verification of Tropsha Criteria

R^2_{test}= 0.858> 0.7 Q^2_{CVtest}= 0.834> 0.6 $|R^2_{Test} - R^2_0| = 0.000 \leq 0.3$

$\frac{|R^2_{Test}-R^2_0|}{R^2_{Test}} = 0.000 < 0.1$ and $0.85 \leq k = 1.000 \leq 1.15$;

$\frac{|R^2_{Test}-R'^2_0|}{R^2_{Test}} = 0.000 < 0{,}1$ and $0.85 \leq k' = 0.858 \leq 1.15$

Besides, the low value of the standard error S = 0.102 attests that foretold values and experimental ones are very similar (Figure 4). This information is confirmed by the graph below which shows a similar evolution of these data given by the multilinear model p_{IC50} of anticancer activity of coumarin derivatives, despite some recorded differences.

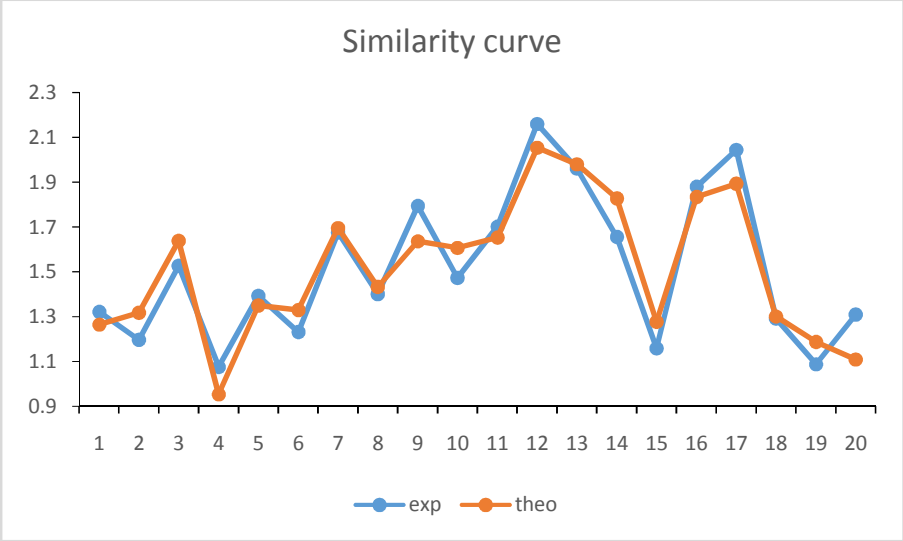


Figure 4: Relationship between predictive and experimental data.

This model is therefore acceptable for the prediction of the anticancer activity of the series of studied compounds. In addition, the fact that the five (5) Tropsha criteria are verified corroborates the predictive power of the model.

3.3. Analysis of Descriptors Contribution in the Model

The study of the relative contribution of the descriptors in the prediction of the anticancer activity of the compounds was carried out using the software XLSTAT version 2014 [28]. The different contributions are shown in Figure 5.

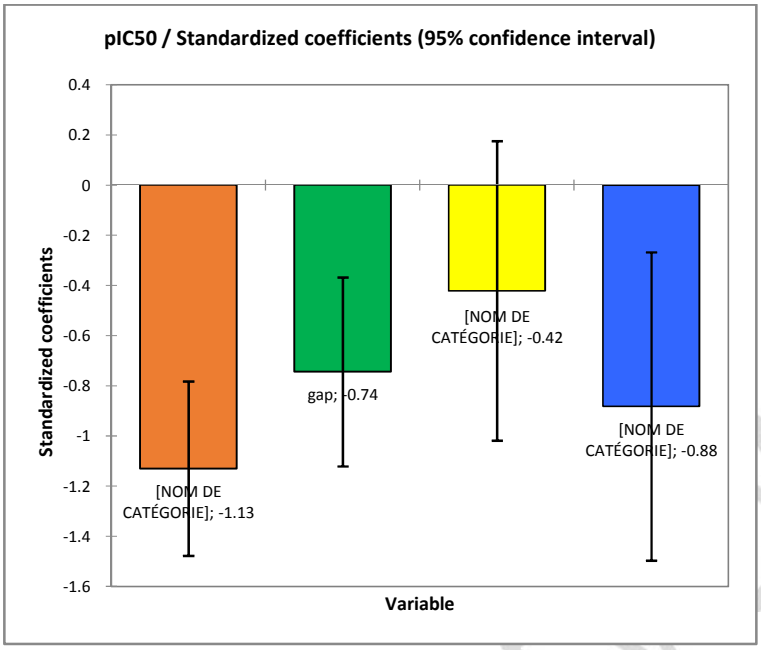


Figure 5: Contribution of the different constituents of the model

The descriptor priority order is given as follows:

(μD)> D (O = C-C-H)> (ΔE_Gap)> d (C-C). Therefore, the dipole moment μ_D is assumed to be the first descriptor in the prediction of anticancer activity because it has the highest contribution.

4. CONCLUSION

The dipole moment μ_D, the energy gap ΔE_{Gap}, the distance d(c-c) and the dihedral angle D(o=cch) allowed us to predict the anticancer activity of coumarin derivatives. This study showed the existence of a strong correlation between the calculated and experimental values of the anticancer potentials. The obtained QSAR model allows us to predict the activity of the best analogs called "lead". This proposed model reveals that the dipole moment is the paramount descriptor for improving anticancer activity. This work is a compass for designing new, more active molecules. The significance of this model was verified using a test set comprising seven (07) molecules. The work presented here will therefore play an important role in understanding the relationship between physico-chemical parameters of the structure and biological activity. The study of this QSAR model could help us to select the appropriate substituent to design new, more efficient compounds with improved biological activity.

REFERENCES

[1] World health organization's FIGHT AGAINST CANCER strategies that prevent, cure and care 1/02/2019

[2] World health organization press release n°263 12 September 2018 Latest global cancer data: Cancer burden rises to 18.1 million new cases and 9.6 million cancer deaths in 2018 1/02/2019

[3] Y. Garazd, M. Garazd, R.Lesyk, Synthesis and evaluation of anticancer activity of 6-pyrazolinylcoumarin derivatives, *Saudi Pharmaceutical Journal* , 2017*;* vol. 25(2), pp. 214-223.

[4] J Bruneton. Pharmacognosy, Phytochemistry, Medicinal Plants. Second Edition, *Hampshire UK, Intercept Ltd*. ,1999; pp 263-277.

[5] T Ojala. Biological Screening of Plant Coumarins. PhD Thesis, University of Helsinki, Helsinki, Finland, 2001.

[6] S. Doh, L. Ekou, M. G-R. Koné, T. Ekou, S. T. Affi, L. Ouattara and N. Ziao, Prediction of the Inhibitory Concentration of Hydroxamic Acids by DFT-QSAR Models on Histone Deacetylase 1, *International Research Journal of Pure & Applied Chemistry*, 2018; vol. 16(2), pp.1-13.

[7] M. Chhabria, B. Mahajan, P. Brahmkshatriya, QSAR study of a series of Acyl Coenzyme A (CoA): Cholesterol Acyltransferase inhibitors using genetic function approximation, *Medicinal Chemistry Research*., 2011; Vol. 20, pp. 1573-1580.

[8] V. Buha, D. Rana, M. Chhabria, K. Chikhalia, B. Mahajan, P. Brahmkshatriya, N. Shah, «Synthesis, Biological Evaluation and QSAR Study of a Series of Substituted Quinazolines as Antimicrobial Agents» *Medicinal Chemistry Research*, 2013*;* vol. 22, pp. 4096-4109.

[9]. A. Tropsha, «Best Practices for QSAR Model Development, Validation, and Exploitation» *Molecular Informatics*, 2010*;* vol. 29, pp. 476-488.

[10] T. I. Oprea, « Chemoinformatics in Drug Discovery» *Ed. WILEY-VCH Verlag. Allemagne*, 2005.

[11] E. A. Rekka, P. N. Kourounakis, « Chemistry and Molecular Aspects of Drug Design and Action» *Ed. Taylor & Francis Group, LLC. Etats Unies*, 2008.

[12] C. Hansch, T. Fujita, «ρ – σ – π,analysis: method for correlation of biological activity and chemical structure» *J. Am. Chem. Soc.* , 1964*,* vol. 86, pp. 1616-1626.

[13] S. M. Free, J. W. Wilson, «A Mathematical Contribution to Structure-Activity Studies» *J. Med. Chem.* , 1964*;* vol. 7, pp. 395-399.

[14] S. A. Morsy, A. A. Farahat, M. N.A. Nasr, A. S. Tantawy, Synthesis, molecular modeling and anticancer activity of new coumarin containing compounds, Saudi Pharmaceutical Journal, *2017*; Vol. 25 (6), pp. 873-883.

[15]. T Mosmann,. Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. J. Immunol. Methods. 1983, 65, 55–63.

[16] K. N. N'guessan, M. G. R. Koné, K. Bamba, O. W. Patrice, N. Ziao, Quantitative Structure Anti-Cancer Activity Relationship (QSAR) of a Series of Ruthenium Complex Azopyridine by the Density Functional Theory (DFT) Method, *Computational Molecular Bioscience*, 2017, 7, 19-31.

[17] S.Chaltterjee, A. Hadi, B. Price, *Regression Analysis by Examples*; Wiley *VCH: New York, USA, 2000.*

[18] HTNPhuong, « Synthèse et étude des relations structure/activité quantitatives (QSAR/2D) d'analoguesBenzo [c] phénanthridiniques» Doctorat Thesis, Angers University, (France), 2007.

[19] M. Frisch, G. Trucks, H. Schlegel, G. Scuseria, Gaussian 09, Revision C.01, Gaussian, Inc., Wallingford CT. , 2009.

[20] N. J.-B. Kangah, M. G.-R. Koné, C. G. Kodjo, B. R. N'guessan, A. L. C. Kablan, S. A. Yéo, N. Ziao, «Antibacterial Activity of Schiff Bases Derived from Ortho Diaminocyclohexane, Meta-Phenylenediamine and 1,6-Diaminohexane: Qsar Study with Quantum Descriptors» *International Journal of Pharmaceutical Science Invention*, 2017*;* vol. 6(13), pp. 38-43.

[21] J. S. N'Dri, M. G.-R. Koné, C. G. Kodjo, S. T. Affi, A. L. C. Kablan, O. Ouattara, D. Soro, N. Ziao, «Relation Quantitative Structure Activité(QSAR) d'une série d'azetidinones dérivés de Dapsone par la

méthode de Théorie de la fonctionnelle de la densité(DFT)» *IRA-International Journal of Applied Sciences;* , 2017 *;* vol. 8(12), pp. 55-62.

[22] P. Geerlings, F. De Proft, and J. M. L. Martin, In Theoretical and Computational Chemistry; Seminario, J., Ed.; Elsevier; Amsterdam. Vol-4 (Recent Developments in Density Functional Theory), 1996; vol. 4, pp. 773-809.

[23] Y. Traoré, M. G-R. Koné, O. Ouattara, N. Ziao, «Qsar Approach to Estimating the Analgesic Activity of a Serie of Tri-Substituted Pyrimidine Derivatives», *Journal of Computational Chemistry & Molecular Modelling*, 2018; vol. 2(4), pp. 1-14.

[24] P. Geerlings, F. De Proft, and W. Langenaeker, Adv. *Quantum Chem.*, 1999; vol. 33, pp. 303-328.

[25] R. G. Parr, R. A. Donnelly, M. Levy, and W. E. Palke, *J. Chem. Phys.*, 1978*;* vol. *68*, pp. *3801-*3807.

[26] C. Hansch, P. G. Sammes, and J. B. Taylor, Computers and the medicinal chemist; in: Comprehensive Medicinal Chemistry, *Eds. Pergamon Press, Oxford*, 1990; vol. 4, pp. 33-58.

[27] R. Franke, Theoretical Drug Design Methods, *Elsevier, Amsterdam,* 1984; pp 184-195.

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

[29] XLSTAT Version 2014.5.03 Copyright Addinsoft 1995-2014 XLSTAT and Addinsoft are Registered Trademarks of Addinsoft. 2014, https://www.xlstat.com

[30] J. S. N'dri , M. G-R Koné , C. G. Kodjo , A. L. C. kablan, S. T. Affi , L. Ouattara and N. Ziao, Theoretical Study of the Chemical Reactivity of Five Schiff Bases Derived From Dapsone by the DFT Method, Chemical Science International Journal, 2018; vol. 22(4), pp. 1-11.

[31]. A. R. Katritzky, V. S. Lobanov, M. Karelson, CODESSA Comprehensive Descriptors for Structural and Statistical Analysis, Reference Manual, version 2.0, *University of Florida, FL Gainesville*, 1994.

[32] V Y. Nalimov, The Application of Mathematical Statistics to Chemical Analysis, *Addison-Wesley, Reading*, M A, 1962.

[33] R. Calcutt, R Body, Statistics for Analytical Chemists, Champman & Hall, New York, (1983).

[34] J. C. Miller, J. N Miller, Statistics for Analytical Chemistry, 2nd ed. Ellis Horwood, New York, 1988; pp. 53–59.

[35] P. C. Meier, R E Zund, Statistical Methods in Analytical Chemistry, John Wiley & Sons, New York, 1993; pp 84-120.

[36] P. Dagnélie, Statistique Theorique et Appliquée : Statistique descriptive et bases de l'inférence statistique, *Tomes 1, De Boeck et S. Larcier, Bruxelle*, 1998 ; pp.508 (french).

[37] G. W. Snedecor, W. G. Cochran, Statistical Methods; Oxford and IBH: New Delhi, India, 1967; pp. 381-418.

[38] M. V. Diudea, QSPR/QSAR Studies for Molecular Descriptors; Nova Science: Huntingdon, New York, USA, 2000.

[39] Y. H. Kpidi, O. B. Yapo, M. G-R. Koné, G. A. Gadji, A. E. J. E. Y. Gnagne, J. S. N'dri, and N. Ziao, "Monitoring and Modeling of Chlorophyll-a Dynamics in a Eutrophic Lake: M'koa Lake (Jacqueville, Ivory Coast)." American Journal of Environmental Protection, 2018; vol. 6(1), pp, 1-9.

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

[41] L. Eriksson, J. Jaworska, A. Worth, M.T. D. Cronin, R. M. Mc Dowell, P. Gramatica, Methods for Reliability and Uncertainty Assessment and for Applicability Evaluations of Classification- and Regression-Based QSARs, Environmental Health Perspectives, 2003; vol. 111(10), pp. 1361-1375.

[42] O. Ouattara, M. G-R Koné, T. S. Affi, K. Bamba, Y. Traore, N. Ziao, Contribution to The Molecular Lipophilicity Scale By QSPR Models Of Lipophilicity Prediction, Journal of Engineering Research and Application, 2018; vol. 8(7), pp. 55-61.

[43] A. Golbraikh, A. Tropsha, Beware of qsar, J. Mol. Graph. Model, 2002; vol. 20, pp. 269-276.

[44] (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., 2003; vol. 22, pp. 69-77.

(b) A. Tropsha, Best practices for QSAR model development, validation, and exploitation. Molecular Informatics, 2010; vol. 29, pp. 476-488.