Journal of Fundamental and Applied Sciences

Research Article

ISSN 1112-9867

Available online at http://www.jfas.info

COMPUTATIONAL SCREENING AND QSAR STUDY ON A SERIES THEOPHYLLINE DERIVATIVES AS ALDH1A1 INHIBITORS

F Z. Fadel[1], N. Tchouar[1], S. Belaidi[2,*], F. Soualmia[1], O. Oukil[1], K. Ouadah[1]

[1]Laboratoire d'Ingénierie des Procédés et de l'Environnement (LIPE), Faculté de Chimie, Université des Sciences et de la Technologie d'Oran Mohamed BOUDIAF, USTO-MB, BP 1503, El M'naouer, 31000 Oran, Algérie

[2]Group of Computational and medicinal Chemistry, LMCE Laboratory, University of Biskra, BP 145 Biskra 07000, Algeria

Received: 26 March 2021 / Accepted: 28 April 2021 / Published online: 01 May 2021

ABSTRACT

In the present study, we explored a series of molecules with anticancer activity, so that qualitative and quantitative studies of the structure-activity relationship (SAR/QSAR) were performed on seventeen theophylline derivatives. These are inhibitors of ALDH1A1. The present study shows the importance of quantum chemical descriptors, constitutional descriptors and hydrophobicity to develop a better QSAR model, whose studied descriptors are Log*P*, MW, Pol, MR, S, V, HE, DM, E_{HOMO} and E_{LUMO}.

A multiple linear regression (MLR) and artificial neural networks (ANN) procedure was used to design the relationships between molecular descriptors and the inhibition of ALDH1A1 by theophylline derivatives. The validation and good quality of the QSAR model are confirmed by a strong correlation between experimental and predicted activity.

Keywords: Theophylline; ALDH1A1 inhibitor; SAR; QSAR; ANN; MLR.

Author Correspondence, e-mail : prof.belaidi@mail.com

doi: http://dx.doi.org/10.43 14/jfas.v13i2.17

1. INTRODUCTION

Aldehyde dehydrogenases (ALDHs) metabolize reactive aldehydes and have important physiological and toxicological functions in areas such as CNS, metabolic disorders, and cancers. Increased ALDH (e.g., ALDH1A1) gene expression and catalytic activity are vital biomarkers in a number of malignancies and cancer stem cells, highlighting the need for the identification and development of small molecule ALDH inhibitors. So, a new series of theophylline-based analogs as potent ALDH1A1 inhibitors is described [1].

Theophylline, one of xanthines, is a naturally occurring alkaloid. It is habitually used as a respiratory drug in the treatment of asthma and obstructive pulmonary disease [2].

The QSAR method is based on defining mathematical dependencies between the variance in molecular structures (encoded by so-called molecular descriptors), and the variance in a given physico-chemical or biological property (so-called endpoint) in a set of compounds. In practice, this imply that if one has experimentally measured substituent constants, other physico-chemical properties or calculated some molecular parameters for a group of similar chemicals and toxicological data are available only for a part of this group, one is able to interpolate the lacking data from the molecular descriptors and a suitable mathematical model [3].

Such predictive computational models could help to decrease the number and cost of synthesis and further requirements of characterization and testing as well as to design nanoparticles having the properties required for their future applications that are simultaneously safe for human health and the environment [4].

The QSAR analysis can be used for two types of purposes:[5]

(1) Qualitative QSAR: To identify the structural/pharmacophoric features, which are responsible for the activity/toxicity profile of a con-generic series of molecules.

(2) Quantitative QSAR: To estimate the activity/toxicity of a molecule before its synthesis and/or biological screening [6]. In this study, the focus is on deriving qualitative and quantitative QSAR models.

The multi-parameter optimization (MPO) methods used to predire drug-likeness and identify bioactive compounds with a good balance ofthe many physicochemical and biological properties essentially to become a successful, efficacious and safe drug [7]. In the MPO

methods, we realised some of rules of thumb including Lipinski and Veber rules and calculated metrics [8-10].

Therefore, discovering drugs is a process, which realizes a sustained balanced search for molecules that have structural features that produce: 1) strong target binding using structure-activity relationship (SAR) and 2) high performance at in vivo barriers, using structure property relationship (SPR) [11].

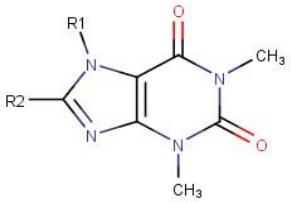
Multiple linear regression (MLR) as well as artificial neural network (ANN) analysis with backward elimination of variables was used to model the structure–activity relationship. A mathematical technique minimizes the difference between the actual and predicted values [12].

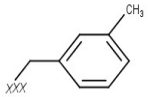
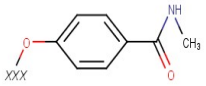
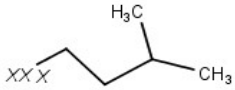
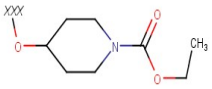
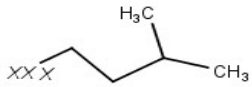
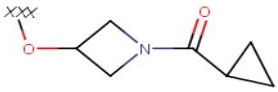
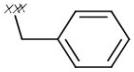
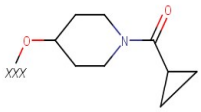
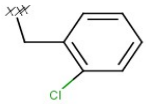
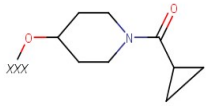
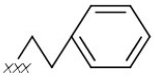
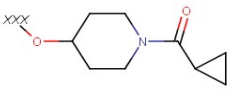
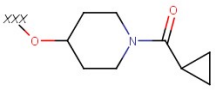
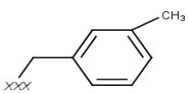
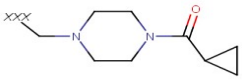
In this contribution, we interest at a series of 17 theophylline derivatives reported by yang and al [1]. Our research aims to describe the qualitative and quantitative structure-activity relationship study on theophylline derivatives and to develop QSAR model for these compounds with regard to their activity sited above.

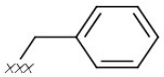
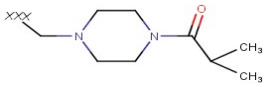
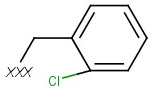
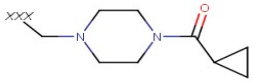
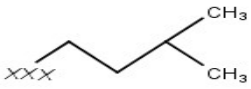
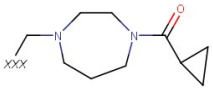
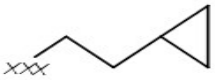
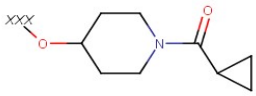
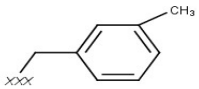
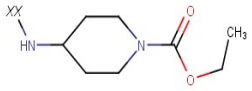
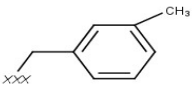
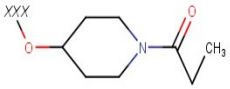
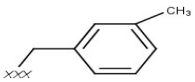
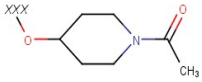
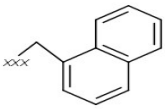
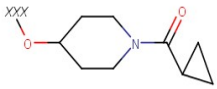
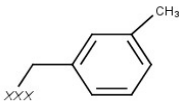
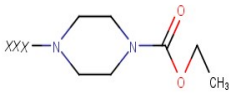
The IC_{50} values (concentration of a compound required to inhibit 50% of theophylline inhibitors activity) were adopted as reported by yang and al (Tables **I**). Then for the used as a dependent variable for the QSAR model were converted to the logarithmic scale [pIC_{50}], ($pIC_{50}=-\log_{10} IC_{50}$).

The molecular modeling calculations for all the theophylline derivatives to describe the QSAR the following software's performs properties: HyperChem 8.08 [13], Gaussian 09 program package [14] and Molinspiration online database [15].

Table I. Chemical structures and experimental activities of the theophylline derivatives.



N°	R1	R2	IC_{50}(nM)
1			57
2			138
3			537
4			214
5			177
6			629
7			81
8			33

9			**256**
10			**69**
11			**320**
12			**555**
13			**562**
14			**225**
15			**245**
16			**91**
17			**270**

2. EXPERIMENTAL DETAILS

2.1. Data Set

A data set of theophylline derivatives as ALDH1A1 inhibitors is described. Seventeen molecules presented in (table **I**), were adopted as reported by Yang and al [1], the reported IC_{50} values (nM) have been converted to the logarithmic scale [pIC_{50}], for QSAR study.

2.2. Descriptors Generation

Seventeen investigated molecules were pre-optimized by means of the Molecular Mechanics Force Field (MM+) included in HyperChem version 8.0.8 package. So the resulting minimized structures were refined by HyperChem using the PM3 semi-empirical Hamiltonian. This approach allowed us to identify a number of physico-chemical descriptors: surface area grid (S), molar volume (V), hydration energy (HE), partition coefficient octanol/water (Log*P*), the molar refractivity (MR), molar polarizability (Pol) and molecular weight (MW).

Then we use Gaussian 09 program package, at the density functional theory (DFT) level using Becke's three-parameter LeeYang-Parr (B3LYP) [16], with the 6-311G (d, p) basis set to re-optimized the group of theophylline derivatives and identify other electronic descriptors: dipole moment (DM) and energy of frontier orbital's (E_{HOMO} and E_{LUMO}).

In addition, Veber's and Lipinski rules suggest that the polar surface area (PSA), number of rotatable bonds (NRB), hydrogen bond donors (HBD) and hydrogen bond acceptors (HBA) are important to determine the oral bioavailability [8]. These descriptors were calculated by using Molinspiration.

The calculation of the two parameters Log*P* and molar refractivity (MR) were performed using atomic parameters derived by Viswanadhan and co-workers [17].

Refractivity was calculated using atomic contributions to refractivity by Ghose and Crippen [18]. Solvent-accessible surface bounded molecular volume and van der Waals surface-bounded molecular volume were calculated using the atomic radii of Gavezotti [19], and basing on a grid method derived by Bodor and al [20].

Based on exposed surface area [21], hydration energy (HE) was considered an essential factor in determining the stability of various molecular conformations [22, 23].

The additivity scheme makes it possible to estimate the polarizability with a precision of 3%, which has been proposed by Miller [24] where different increments are associated with different

atom types; the polarizability of a molecule characterizes the capability of its electronic system to be distorted by the external field [25].

The molecular weight (MW) of a system calculation is based on a general applicability method [13].

2.3. Regression Analysis

Multiple linear regression (MLR) analysis of molecular descriptors and artificial neural networks (ANNs) are used. The reliability of such models is mainly evaluated by the correlation coefficient R^2 [26]. The MLR and ANN models were generated using the software JMP 8.0.2 [27].

3. RESULTS AND DISCUSSION

3. 1. Computational screening for theophylline derivatives

In this part, we have applied rules of thumb and metrics methods on seventeen derivatives of theophylline (Table **I**) with respect to their anticancer activity (pIC_{50}) against ALDH1A1 [1]. The properties involved are: partition coefficient octanol/water (Log*P*), molecular weight (MW), hydrogen bond donors (HBD), hydrogen bond acceptors (HBA), number of rotatable bonds (NRB), polar surface area (PSA), Ligand efficiency (LE) and Lipophilic efficiency (LipE).The results using HyperChem 8.0.8 [13] and Molinspiration online data base [15] are shown in (Table **II**).

At first, we have studied Lipinski and Veber rules to identify “drug-like” compounds. Rich absorption or permeability is more likely when: [7,8]

(1) H-bond donors, nitrogen or oxygen atoms with one or more hydrogen atoms (HBD) ≤5 (expressed as the sum of OHs and NHs).

(2) The molecular weight (MW) ≤500 Da.

(3) Octanol water partition coefficient log*P* ≤5.

(4) H-bond acceptors, nitrogen or oxygen atoms (HBA) ≤10 (expressed as the sum of Ns and Os).

(5) Rotatable bonds ≤10.

(6) Polar surface area ≤140 $Å^2$.

We used the Lipinski's rules to identify compounds with problems of absorption and permeability if these compounds do not validate at least two of its rules [28].

In addition, the last two descriptors mentioned are identified by Veber and al [8] concerning the oral bioavailability of the drug. Lipinski and Veber rules are based on a strong physicochemical rationale. Hydrogen bonds increase solubility in water and must be broken allowing the compound to permeate into and through the lipid bilayer membrane [29]. Thus, an increasing number of hydrogen bonds reduce partitioning from the aqueous phase into the lipid bilayer membrane for permeation by passive diffusion [30]. (Table **II**) shows that all the studied derivatives are compatible with rules number (1) and (4). Therefore, it is possible to say that they are less polar and more absorbed.

Table II. Pharmacological activities and properties involved in MPO method for theophylline derivatives

	Lipinski rules					Veber rules		Ligand efficiency and Lipophilicity efficiency		
N°	**Log P**	**MW (uma)**	**HBA**	**HBD**	**Lipinski score of 4**	**NRB**	**PSA Å²**	**pIC_{50}**	**LE**	**LipE**
1	0.19	433.47	9	1	4	5	100.17	7.244	0.3169	7.054
2	1.55	421.50	10	0	4	7	100.61	6.860	0.3201	5.310
3	1.17	389.46	9	0	4	6	91.38	6.270	0.3135	5.100
4	0.75	437.50	9	0	4	5	91.38	6.669	0.2917	5.919
5	0.53	471.94	9	0	4	5	91.38	6.752	0.2860	6.222
6	1.00	451.53	9	0	4	6	91.38	6.201	0.2630	5.201
7*	1.67	431.53	9	0	4	7	91.38	7.015	0.3168	5.345
8*	-0.03	450.54	9	0	4	5	85.38	7.481	0.3173	7.511
9	0.32	438.53	9	0	4	5	85.38	6.591	0.2883	6.271
10	-0.41	470.96	9	0	4	5	85.38	7.161	0.3038	7.571
11*	0.39	430.55	9	0	4	6	85.38	6.494	0.2932	6.104
12	0.77	415.49	9	0	4	6	91.38	6.255	0.2905	5.485
13	0.87	454.53	10	1	4	6	103.41	6.250	0.2651	5.380
14	0.84	439.51	9	0	4	5	91.38	6.647	0.2908	5.807

15	0.22	425.49	9	0	4	4	91.38	6.610	0.2985	6.390
16	0.82	487.56	9	0	4	5	91.38	7.040	0.2737	6.220
17	1.05	440.50	10	0	4	5	94.62	6.244	0.2731	5.194

* corresponds to test molecules.

Molecular weight (MW) is related to the size of the molecule, with its increasing, a larger cavity should be formed in water to solubilize the compound [31]. There is an inverse relationship between (MW) and the concentration of the compound on the surface of the intestinal epithelium and its absorption. If the size increases will create barriers such as the prevention of passive diffusion through the tight aliphatic side chains of the bilayer membrane. We find that the molecular weight of all the compounds of the theophylline derivatives series is less than 500 Da (rule number 2), so we can consider them soluble and easily cross cell membranes.

The oral solubility of the drug is determined by Log*P*, this parameter is give by partitioning the molecule between water and the hydrophobic solvent n-octanol, and determining the P value as the ratio of the concentration of the compound in n-octanol and that in water. However the increasing Log*P* decreases aqueous solubility, which minimizes absorption. If the values of Log*P* are negative indicates that the compound is too hydrophilic. So it has good aqueous-solubility, better gastric tolerance and efficient elimination through the kidneys. But if the values of Log*P* are positive indicates that the compound is too lipophilic. So it has a good permeability through biological membrane, a better binding to plasma proteins, elimination by metabolism but a poor solubility and gastric tolerance [32]. All studied molecules have almost optimal (Log*P*) values; for good oral bioavailability, the Log*P* must be greater than zero and less than 3 (0 <Log*P*<3). If Log*P* is too high (>3), the drug has low solubility. Where as for too low Log*P* (<0), the drug has difficulty penetrating the lipid membranes [33,34].

In this study, it is noted that the compound **10** has the lowest value of Log*P*, so is expected to have the highest hydrophilicity, this implies that this compound will have good aqueous-solubility, better gastric tolerance and efficient elimination through the kidneys. This during compound **7** which has the highest Log*P* value will be the most lipophilic; this implies that this

compound will have good permeability across cell membrane. Note that all compounds of the chosen series have Log*P* values less than 5.

It is well known that high oral bioavailability is a significant factor for the progress of bioactive molecules as therapeutic agents. Reduced molecular flexibility (measured by the number of rotatable bonds) and low polar surface areas are found to be important predictors of good oral bioavailability [35,36].

The number of hydrogen bond acceptors (O and N atoms) and the number of hydrogen bond donors (NH and OH) have shown to be critical in a drug development setting as they influence absorption and permeation [37]. These are found to be within Lipinski's limit i.e., less than 10 and 5 respectively, in the tested compounds. Molecules violating more than one of these parameters may have problems with bioavailability and high probability of failure to display drug likeness [38,39].

Whereas, rotatable bonds and polar surface area tend to increase with molecular weight may in part explain the success of these two parameters in predicting the oral bioavailability and the transport across membranes.

The number of rotatable bonds (NRB) was defined as any single bond, not in a ring, bound to a no terminal heavy (i.e., non-hydrogen) atom. Excluded from the count were amide C–N bonds because of their high rotational energy barrier [8]. The low number of rotatable bonds (reduced flexibility) in the studied series indicates that these Ligands upon binding to a protein change their conformation only slightly [40].

The number of rotatable bonds (NRB) is a simple topological parameter that measures molecular flexibility and is considered to be a good descriptor of oral bioavailability of drugs. The low number of rotatable bonds (reduced flexibility) in the studied series indicates that these ligands upon binding to a protein change their conformation only slightly [41]. Rotatable bonds are under 10 so all the screened compounds were flexible.

Polar surface area (PSA) is a very useful parameter for prediction of drug transport properties. PSA is defined as a sum of surfaces of polar atoms (usually oxygens, nitrogens and attached hydrogens) in a molecule [42]. This parameter has been shown to correlate very well with the human intestinal absorption, Caco-2 monolayer's permeability, and

blood-brain barrier penetration [43]. Molecules with PSA values of 140 **Å²** or more are expected to exhibit poor intestinal absorption [44].

Indeed, all the **17** molecules with PSA values between **85.38** and **103.41 Å²**, belong to the compounds with reduced absorption (Table **II**).

We have studied Lipophilic Efficiency LipE, which is considered important for normalizing potency over lipophilicity.

LipE is used to compare compounds of different potencies (pIC_{50}) and lipophilicities (Log*P*). For a given compound lipophilic efficiency is defined as the pIC_{50} of interest minus the Log*P* of the compound [45,46]. Although in vitro potency and lipophilicity of compounds are important parameters to evaluate, the concept of Lipophilic Efficiency (LipE) aids in establishing a more balanced relationship between the potency observed in vitro and lipophilicity properties of evaluated chemical compounds [47]. Ryckmans et al [48] reported that high quality lead compounds possess higher LipE values.

Lipophilicity efficiency (LipE) is defined as follows:

$$LipE = pIC50 - Log P \quad (1)$$

The lipophilicity is the major factor for the promiscuity of compounds, LipE optimized compounds should be more selective. It is suggested to target a LipE in a range of 5–7 or even higher [32]. In the series studied, for the **17** compounds, the LipE value is in the suggested range of **5** to **7** or even a little above, indicating that these compounds have been successfully optimized.

Ligand Efficiency (LE) is a particularly important parameter in fragment drug design as it gives priorities to small molecules with relatively lower potency rather than larger, higher potency molecules [49,50].

Ligand efficiency is defined by the following equation:

$$LE = 1.4pIC_{50}/NH \quad (2)$$

Where: NH is the number of heavy atoms. So if the number of heavy atoms increases, the value of LE decreases [51].

From the results obtained in (Table **II**), all the derivatives containing a pIC_{50} between **6.201** and **7.481** and we can penalize the compounds **6** and **13** with the lowest values of LE respectively

0.2630 and **0.2651**.

3.2. Structure Activity Relationship for Theophylline Derivatives

In the first step of our studies, we have studied seven physicochemical properties of theophylline derivatives shown in (table **III**) taken from the literature with their IC_{50} against ALDH1A1 [1]. The properties involved are: surface area grid (**S**), molar volume (**V**), hydration energy (**HE**), partition coefficient octanol/water (**Log*P***), the molar refractivity (**MR**), polarizability (**Pol**) and molecular weight (**MW**) using HyperChem 8.0.8; Also, we have studied three quantum properties of theophylline derivatives (Table **III**). The properties involved are: dipole moment (**DM**), Energy of frontier orbital's $\mathbf{E_{HOMO}}$ and $\mathbf{E_{LUMO}}$ using Guaussian 09.

The attractive part of the Van der Waals interaction is a good measure of the polarizability [52]. Molecular polarizability of a molecule characterizes the capability of its electronic system, and it plays an important role in modeling many molecular properties and biological activities [37], also, the molar refractivity is important criterion to measure the steric factor and designated as a simple measure of the volume occupied either by individual atom or cluster (group) of atoms [53]. Molar refractivity and polarizability relatively increase with the size and the molecular weight of the studied theophylline derivatives (Table **III**). This result is in agreement with the formula of Lorentz-Lorenz [42] which gives a relationship between polarizability, molar refractivity and the molecular size [54].

For example, the compound **16** has great values of polarizability (**51.76**Å³), molar refractivity (**139.77**Å³) and volume (**1335.13**Å³). In contrast, the compound **3** is the small molecule in this studied series, which has a small value of polarizability (**39.58**Å³), of molar refractivity (**101.52**Å³) and volume (**1153.65**Å³). The decreasing order of polarizability for these studied for examples **16**, **10, 8, 5** and **3** (Table **III**).

Table III. Values of some descriptors used in the regression analysis.

N°	Log P	Pol Å³	MW (uma)	V Å³	MR Å³	S Å²	HE kcal/mol	DM (D)	E_{HOMO} (a.u)	E_{LUMO} (a.u)
1	0.19	45.7700	433.47	1217.6100	126.0400	708.0800	-5.1700	2.5236	-0.2434	-0.0724
2	1.55	42.8300	421.50	1244.2800	110.2200	724.2600	0.2900	2.9090	-0.2148	-0.0332
3	1.17	39.5800	389.46	1153.6500	101.5200	680.3400	0.0200	4.4100	-0.2464	-0.0674

4	0.75	45.5700	437.50	1225.9600	121.5800	701.5700	-2.4600	4.5937	-0.2448	-0.0668
5	0.53	47.5000	471.94	1262.9400	126.2900	721.7400	-2.2100	5.5317	-0.2492	-0.0677
6	1.00	47.4100	451.53	1249.3600	126.3300	693.7700	-2.0700	0.6801	-0.2323	-0.0742
7*	1.67	45.0900	431.53	1284.0900	115.8500	742.6100	0.6000	4.7468	-0.2429	-0.0637
8*	-0.03	48.1200	450.54	1275.0700	128.2600	721.4800	-0.4200	1.8830	-0.1602	-0.0959
9	0.32	47.0600	438.53	1246.3200	125.7800	697.8200	-0.9700	6.1760	-0.1769	-0.0746
10	-0.41	48.2200	470.96	1248.7800	128.7000	698.5700	-1.3100	2.8554	-0.2150	-0.0708
11*	0.39	45.8000	430.55	1221.8700	118.5200	678.6700	0.7700	4.8421	-0.2412	-0.0677
12	0.77	42.4800	415.49	1181.1900	109.4400	670.0500	-0.3100	1.8993	-0.2433	-0.0671
13	0.87	47.7000	454.53	1249.8200	126.7800	703.4000	-1.6800	4.2402	-0.2351	-0.0692
14	0.84	46.3500	439.51	1252.5800	123.0900	702.0100	-1.0200	5.6455	-0.2423	-0.0681
15	0.22	44.5100	425.49	1199.5400	118.4600	675.5200	-1.3100	5.6278	-0.2423	-0.0679
16	0.82	51.7600	487.56	1335.1300	139.7700	755.3000	-2.8300	1.6417	-0.2331	-0.0674
17	1.05	45.8600	440.50	1234.8300	122.6700	700.2900	-1.3000	3.0396	-0.2291	-0.0741

* corresponds to test molecules.

Surface and distribution volume of these molecules are definitely higher than those of more polar molecules like the lipopeptides or beta-lactams [55]. We found that surfaces vary from (**670.05** to **755.30** Å²). The most important hydration energy in the absolute value, is that of compound **1** (**5.17** kcal/mol) and the weakest is that of compound **3** (**0.02** kcal/mol) (Table **III**). As seen in (Tables **III**) the compound **3** have the smallest value of hydration energy (**0.02** kcal/mol) whereas compound **1** correspond to very high value of absolute hydration energy given by (**5.17** Kcal/mol). The results obtained by calculating Log*P* of theophylline derivatives show that the compounds **10** present small coefficient of lipophilicity (**-0.41**).

Although the inhibition effect produced by all molecules seems to be the same pharmacological point of view, an additional element of answer provided by the theoretical study is that each theophylline derivative has negative, different and lower energy. Compound **16 (-6989.1755** a.u) is more stable in comparison with compound **3 (-5563.1607** a.u). This may explain the inhibition behaviour. Compound **9** indicates the value of the maximum dipole moment (**6.1760** D). It comes from a resonance effect, involving a donor effect of the nucleus towards the electro-attractive.

For more precision, in this work we have studied other very important quantum chemical descriptors such as the energies of the **HOMO** (the highest molecular orbital occupied) and the **LUMO** (the lowest molecular orbital occupied).

The $\mathbf{E_{HOMO}}$ is directly related to the ionization potential and characterizes the susceptibility of the molecule toward attack by electrophiles, where the $\mathbf{E_{LUMO}}$ is directly related to the electron affinity and characterizes the susceptibility of the molecule toward attack by nucleophiles. Both the $\mathbf{E_{HOMO}}$ and the $\mathbf{E_{LUMO}}$ are important in radical reactions [56-57].

3.3. Quantitative Structure-Activity Relationships Studies (QSAR) of theophylline derivatives

In the second step, we conducted this study in order to develop the best QSAR model and explain the correlations between physicochemical parameters and biological activities pIC_{50} values of theophylline derivatives.Various statistical parameters allowed us to select the best QSAR model, among which we can mention: squared correlation coefficient (**R^2> 0.6**) which is relative measure of quality of fit, Fischer's value (**F**), F is the Fisher ratio, reflects the ratio of the variance explained by the model and the variance due to the error in the regression. High values of the F-test indicate that the model is statistically significant [58].

The selection of a set of appropriate descriptors that encode various structural features of the molecules among many of them for the development of a QSAR model requires a method capable of distinguishing the parameters.

Pearson's correlation matrix has been performed on all descriptors by using the software JMP 8.0.2. [27] The analysis of the matrix revealed physico-chemical descriptors and quantum descriptors for the development of MLR model. MLR is one of the earliest and still one of the most commonly used methods for constructing QSAR mathematical models [59-61] because of its simplicity, transparency, reproducibility, and easy interpretability [62].

The values of descriptors used in MLR analysis are presented in (Table**III)**. The data set was randomly divided into two sets: a training set (fourteen compounds) and a testing set (three compounds: 7, **8** and **11**)

After multiple regression analysis, we have revealed four independents descriptors for the development of the model. The resulting MLR QSAR model is represented by the following equation:

$$\mathbf{pIC_{50}} = -0.455 + 0.017\mathbf{S} + 10.354\,\mathbf{E_{LUMO}} - 0.762\,\mathbf{LogP} - 0.008\,\mathbf{MW} \qquad (3)$$

$$R^2 = 0.93 \qquad F = 30.36 \qquad RMSE = 0.11$$

Where, R^2 is the coefficient of determination, F is the Fischer statistics and **RMSE** is the root-mean-square error

Squared correlation coefficient $\mathbf{R^2}$ is ***0.93***, explains **93%** variance in biological activity. The $\mathbf{R^2}$ value is more **0.6**, which suggest that a good percentage of the total variance in biological activity is accounted by the model.

In the Eq***(3)***. The negative coefficient of l**og*P*** explains that any increase in the lipophilicity of the molecules causes a decrease in biological activity. From this parameter it may be concluded that hydrophilic molecules are more important for anti-cancerous activity against ALDH1A1.

For validation of the model, we plot in **Fig.1** the experimental activities against the predicted values as determined by equation ***(3)***. We can observe that the predicted pIC50 values are in an acceptable agreement and regular distribution with experimental ones with $\mathbf{R^2 = 0.93.}$

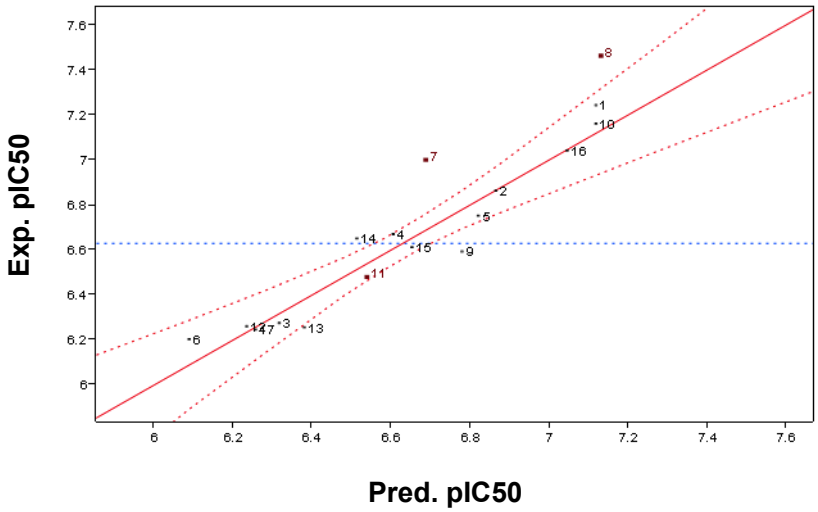


Fig.1. Correlation of experimental and predicted pIC50 as derived using MLR

3.4. Artificial neural networks

Artificial neural networks (ANNs) models are non-linear models useful to predict the biological activity of large data sets of molecules [63]. For instance, ANN was successfully used for the prediction and synthesis of new organic chemical compounds [64].

For our work, ANN contained four inputs corresponding to the four descriptors selected from the correlation matrix, two hidden neurons, and one output neuron which is pIC50 **(Fig.2)**. The number of artificial neurons in the hidden layer was adjusted experimentally [65], two neurons in the hidden layer permitted to attain the best correlation between experimental and predicted data.

A good correlation of experimental and predicted pIC50 by ANN is found. This is shown in (**Fig**.**3)**, and illustrated by $\mathbf{R^2 = 0.97.}$

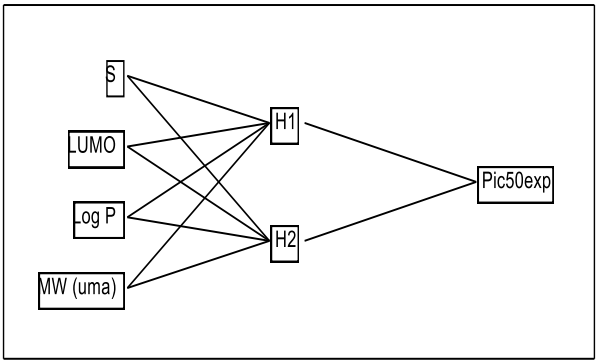


Fig.2. Structure of ANN.

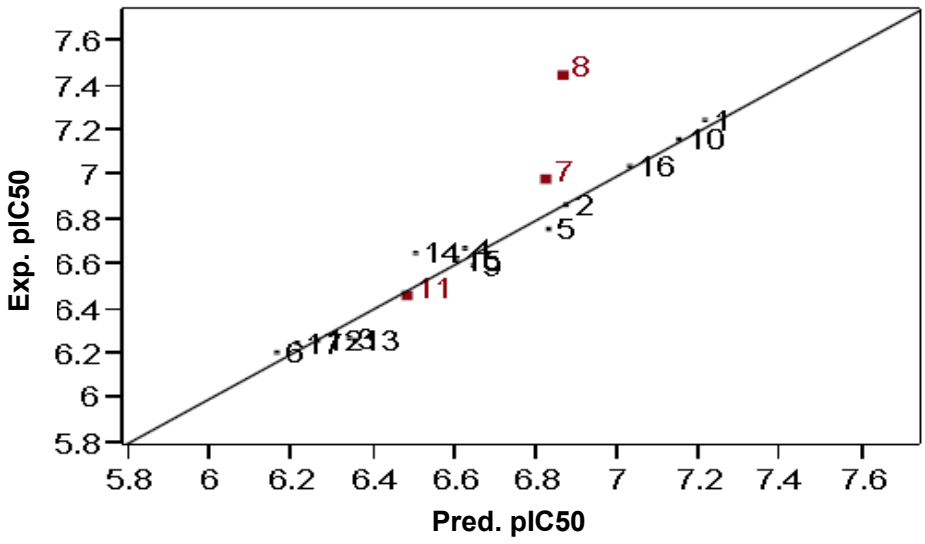


Fig.**3**. Correlation of experimental and predicted pIC50 as calculated by ANN

The model was used to predict the activity values for both training and testing sets. Table **IV** reports the experimental and predicted **pIC50** activities, as well as their differences. The plot of the predicted versus experimental activity (**Fig**.**1** and **3**) shows a linear relationship, indicating a satisfactory internal predictability of the generated model independently of the method used (MLR or ANN).

Moreover, the plot of the calculated residuals against the experimental activity values shows that the residuals are evenly distributed around the zero line, thus confirming the absence of systematic errors in the model.

Table IV. Experimental and predicted activities pIC_{50} of the molecule studied using MLR and ANN.

N°	PIC 50 exp	Pic50Pred. (MLR)	Pic50Pred. (ANN)
1	7.244	7.119	7.211
2	6.860	6.863	6.870
3	6.270	6.317	6.340
4	6.669	6.605	6.627
5	6.752	6.821	6.829
6	6.201	6.088	6.162
7*	7.015	6.686	6.822
8*	7.481	7.130	6.865
9	6.591	6.780	6.644
10	7.161	7.119	7.150
11*	6.494	6.538	6.479
12	6.255	6.233	6.260
13	6.250	6.379	6.353
14	6.647	6.514	6.500
15	6.610	6.653	6.606
16	7.040	7.045	7.027
17	6.244	6.254	6.214

* corresponds to test molecules.

4. CONCLUSION

In this paper, Computational screening and SAR/QSAR analysis was carry out to find the qualitative and quantitative effects of molecular structure of the compounds on their anti-cancerous activity. Our study shows that this series of molecules obey the Lipinski's and Veber's rules. Various physicochemical parameters, particularly partition coefficient Log*P*, MW, S and E_{LUMO} can be used successfully for modelling anti-cancerous activity of theophylline derivatives. Two different methodologies: MLR and ANN were used to identify QSAR models. The comparison shows that ANN has better predictive abilities than MLR. This superiority suggests a nonlinear relation between the selected molecular descriptors and the inhibition activity. The validation and predictive ability of the models were examined by data separation into independent training and testing sets, leave one-out cross-validation and Y-randomization; we notice that all test molecules (**7**, **8** and **11**) are in a good agreement with the two models. The results of which indicate the accuracy and robustness of the proposed QSAR model.

As we considered the predictive capability of the QSAR model developed as well as the low residual activity and cross-tabulation obtained. It indicates the validation of this model and the success of its application to predict the anticancer activity of this series of molecules.

6. REFERENCES

[1] Yang S , Yasgar A, Miller B, Lal-Nag M, Brimacombe K, Hu X, Sun H, Wang A, Xu X, Nguyen K, Oppermann U, Ferrer M, Vasiliou V, Simeonov V, Jadhav A, and Maloney D. J. Med. Chem. 58(15), 2015, 5967–5978, doi: 10.1021/acs.jmedchem.5b00577

[2] Barnes P J. J. Respir. Crit. Care Med.167, 2003, 813–818, doi: 10.1164/rccm.200210-1142PP

[3] Ambure P, Aher R B, Gajewicz A, Puzyn T, and Roy K. J. Chemom. Intell. Lab. Syst. 147, 2015, 1-13, doi: 10.1016/j.chemolab.2015.07.007

[4] Pindelska E, Sokal A, Szeleszczuk L, Pisklak D M, and Kolodziejski W. J. pba. 100, 2014, 322-328, doi: 10.1016/j.jpba.2014.07.011

[5] Fujita T, Winkler D A. J. Mol. Struct. 56(269), 2016, 1243–1252, doi:10.1021/acs.jcim.6b00129

[6] Masand V H, El-Sayed N N E, Bambole M U, and Patil V R. J. Mol. Struct. 1175, 2019, 481-487, doi:10.1016/j.molstruc.2018.07.080

[7] Segall D M, J. Curr. Pharm. Des. 18 (9), 2012, 1292-1310(19), doi: 10.2174/138920012799362864

[8] Veber D F, Johnson S R, Cheng H Y, Smith B R, Ward K W and Kopple K D. J. Med. Chem. 45 (12), 2002, 2615-23, doi: 10.1021/jm020017n

[9] Lipinski C A, Lombardo F, Dominy B W, and Feency P J. J. Adv. Drug Delivery Rev. 64(1–3), 2012, 4-17, doi: 10.1016/j.addr.2012.09.019

[10] Zerroug A, Belaidi S, Benbrahim I, Sinha L, Chtita S. J. JKSUS. 31(4), 2019, 595-601, doi: 10.1016/j.jksus.2018.03.024

[11] Oukil O, Tchouar N, Belaidi S, Salah T, and Cinar M. STRUCTURAL investigation, drug likeness scoring and structure activity/property relationships applied on 1,2,3-thiadiazole derivatives, with kinase inhibitors activity. Rev. Roum. Chim. 2017**,** 62(1): 81-92

[12] Ghamali M, Chtita S, Aouidate A, Ghaleb A, Bouachrine M, Lakhlifi T. J. TUSCI. 11(3), 2017, 422-433, doi: 10.1016/j.jtusci.2016.06.005

[13] HyperChem (Molecular Modeling System) Hypercube, Inc., 1115 NW, 4th Street, Gainesville, FL 32601, USA, 2008.

[14] Gaussian 09, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian Inc., Wallingford, CT, 2010.

[15] Database, [http://www. Molinspiration.com]

[16] Khamouli S, Belaidi S, Almi Z, Medjahed S, and Belaidi H. J. Bionanosci. 11(4), 2017, 301-309, doi: 10.1166/jbns.2017.1445.

[17] Viswanadhan V N, Ghose A K, Revankar G R, and Robins R K. J. Chem. Inform. Comput. Sci. 29(3), 1989, 163–172, doi: 10.1021/ci00063a006

[18] Ghose A K and Crippen G M. J. Chem. Inf. Comput. Sci. 27(1), 1987, 21–35, doi : 10.1021/ci00053a005.

[19] Gavezzotti A. J. Am. Chem. Soc. 105(16), 1983, 5220–5225, doi: 10.1021/ja00354a007.

[20] Bodor N, Gabanyi Z, and Wong C K. J. Am. Chem. Soc. 111(11), 1989, 3783–3786, doi: 10.1021/ja00193a003

[21] Ooi T, Oobatake M, Nemeth G, and Scheraga H A. Proc. Natl. Acad. Sci. USA. 84(10), 1987, 3086–3090, doi: 10.1073/pnas.84.10.3086

[22] Jiang C W, Zhou X, Xie R H, and Li F L. Quantum Matter. 2(5), 2013, 353-363, doi: 10.1166/qm.2013.1067

[23] Chang S K, Cho Y, and Choe J I. J. Comput. Theor. Nanosci. 10(8), 2013, 1701-1705, doi: 10.1166/jctn.2013.3112

[24] Miller K J. J. Am. Chem. Soc. 112(23), 1990, 8533–8542, doi: 10.1021/ja00179a044

[25] Wang J, Xie X Q, Hou T, and Xu X. J. Phys. Chem. A. 111(20), 2007, 4443–4448, doi : 10.1021/jp068423w

[26] Darnag R, Minaoui B, Fakir M. Arab. J. Chem, 10(1), 2017, 600-608, doi: 10.1016/j.arabjc.2012.10.021

[27] JMP 8.0.2, SAS Institute Inc, 2009

[28] R. Aurélien, “ Détermination et Caractérisation Des Cibles De DSP1 Chez DrosophilaMelanogaster ”, Ph.D. Dissertation, *OrleansUniversity*, 2006

[29] Aoumeur N, Tchouar N, Belaidi S, Medjahed S, Oukil O, and Lanez T. J. Bionanosci. 12(3), 2018, 428-436, doi: 10.1166/jbns.2018.1536

[30] Sebaa Z, Tchouar N, Salah T, Belaidi H, Almi Z, and Belaidi S. J. Bionanosci. 12(2), 2018, 278-283, doi: 10.1166/jbns.2018.1511

[31] Kerassa A, Belaidi S and Lanez T. Quantum Matter. 5(1), 2016, 45-52, doi: 10.1166/qm.2016.1253

[32] Schultes S, Graaf C, Haaksma E, Iwan J P and Kramer O. J. Drug Discov. Today: Technologies. 7(3), 2010, e157-e162, doi: 10.1016/j.ddtec.2010.11.003

[33] Zhokhova N I, Baskin I I, Palyulin V A, Zefirov A N, and Zefirov N S. Russ. Chem. B, 52, 2003, 1885–1892, doi: 10.1023/B:RUCB.0000009629.38661.4c

[34] Belaidi S, Kerassa A, Lanez T, and Cinar M. J. Comput. Theor. Nanosci. 12(9), 2015, 2127-2133, doi: 10.1166/jctn.2015.3999

[35] Vistoli G, Pedretti A and Testa B. J. Drug Discov. Today : Technologies. 13(7-8), 2008, 285-94, doi : 10.1016/j.drudis.2007.11.007

[36] Remko M, Swart M and Bickelhaupt F M. J. Bioorg. Med. Chem. 14(6), 2006, 1715-28, doi: 10.1016/j.bmc.2005.10.020

[37] Bemis G W and Murcko M A. J. Med. Chem. 39(15), 1996, 2887–2893, doi: 10.1021/jm9602928

[38] Lipinski C A, Lombardo F, Dominy B W and Feeny P J. J. Adv. Drug. Deliv. Rev. 23(1-3), 1997, 3-25, doi: 10.1016/S0169-409X (96)00423-1

[39] Singh T, Adekoya O A, and Jayaram B. Mol. Bio. Syst. 11, 2015, 1041-1051, doi: 10.1039/C5MB00003C

[40] Medjahed S, Belaidi S, Djekhaba S, Tchouar N, and Kerassa A. J. Bionanosci. 10(2), 2016, 118-126, doi: 10.1166/jbns.2016.1358

[41] Ouadah K, Tchouar N, Belaidi S, Oukil O, and Cinar M. J. Bionanosci. 12(2), 2018, 250-259, doi: 10.1166/jbns.2018.1513

[42] Kerassa A, Belaidi S, Harkati D, Lanez T, Prasad O and Sinha L. Rev. Theor. Sci. 4(1), 2016, 85-96, doi: 10.1166/rits.2016.1050

[43] Ertl P, Rohde B and Selzer P. J. Med.Chem, 43(20), 2000, 3714-7, doi: 10.1021/jm000942e

[44] Viswanadhan V N, Ghose A K, Revankar G R and Robins R K. J. Chem. Inf. Comput. sci. 29(3), 1989, 163–172, doi : 10.1021/ci00063a006

[45] Edwards M P and Price D A. Annu. Rep. Med. Chem. 45, 2010, 380 -391, doi: 10.1016/S0065-7743(10)45023-X

[46] Leeson P D, Springthorpe B. Nat. Rev. Drug. Discov. 6, 2007, 881–890, doi: 10.1038/nrd2445

[47] Tarcsay A, Nyíri K and Keseru G M. J. Med. Chem. 55(3), 2012, 1252-60, doi: 10.1021/jm201388p

[48] Ryckmans T, Edwards M P, Horne, Correia A M, Owen D R, Thompson L R, Tran I, Tutt M F and Young T .Bioorg. Med. Chem. Lett. 19(15), 2009, 4406-4409, doi: 10.1016/j.bmcl.2009.05.062

[49] Carr R A, Congreve M, Murray C W, and Rees D C. Drug. Discov. Today. 10(14), 2005, 987-92, doi: 10.1016/S1359-6446(05)03511-7

[50] Mortenson P N and Murray C W. J. Comput. Aided. Mol. Des. 25, 2011, 663–667, doi: 10.1007/s10822-011-9435-z

[51] Hopkins A L, Groom C R, Alex A. Drug Discov. Today. 9(10), 2004, 430-1, doi: 10.1016/S1359-6446(04)03069-7

[52] Soualmia F, Belaidi S, Belaidi H, Tchouar N, and Almi Z. J. Bionanosci. 11(6), 2017, 584-591, doi: 10.1166/jbns.2017.1476

[53] Wang J, Xie X Q, Hou T and Xu X. Fast. J. Phys. Chem. A. 111(20), 2007, 4443–4448, doi: 10.1021/jp068423w

[54] Almi Z, Belaidi S and Segueni L. Rev. Theor. Sci. 3(3), 2015, 264-272, doi: 10.1166/rits.2015.1038

[55] Belaidi S, Almi Z, and Bouzidi D. J. Comput. Theor. Nanosci. 11(12), 2014, 2481-2488, doi: 10.1166/jctn.2014.3665

[56] Sklenar H and Jager J. J. Quantum. Chem. 16(3), 1979, 467-484, doi: 10.1002/qua.560160306

[57] Tuppurainen K, Lotjonen S, Laatikainen R, Vartiainen T, Maran U, Strandberg M and Tamm T. Mutat. Res. 247(1), 1991, 97-102, doi: 10.1016/0027-5107(91)90037-o

[58] Sanmati K J, Rahul J, Lokesh S and Arvind K Y. QSAR Analysis on 3,5-disubstituted-4,5-dihydropyrazole-1-carbothioamides as epidermal growth factor receptor (EGFR) kinase inhibitors. J. Chem. Pharm. Res. 2012, 4(6):3215-3223

[59] Liu P, Long W. Int. J. Mol. Sci. 10(5), 2009, 1978-1998, doi: 10.3390/ijms10051978

[60] Boudergua S, Alloui M, Belaidi S, Mogren Al Mogren M, AbdEllatif Ibrahim U A, Hochlaf M. J. Mol. Struct. 1189, 2019, 307-314, doi: 10.1016/j.molstruc.2019.04.004

[61] Almi I, Belaidi S, Zerroug E, Alloui M, Ben Said R, Linguerri R, Hochlaf M. J. Mol. Struct.1211, 2020, 128015, doi : 10.1016/j.molstruc.2020.128015

[62] Roy K, Kar S, Das R N. A Primer on QSAR/QSPR Modeling, Springer Briefs in Molecular Science. USA, 2015, pp.1-36

[63] Díaz-González J M, Aguirre-Crespo F J, García-Mera X, Prado-Prado F J. MOL2NET. 1, 2015, pp.2-6

[64] Lazar M, Mouzdahir A, Badia M, Zahouily M. Artemisinin analogues selection in linear and non-linear QSAR. J. Pharm. Res. 2014, 8:118-122

[65] Larif M, Adad A, Hmammouchi R, Taghki A I, Soulaymani A, Elmidaoui A, Bouachrine M, Lakhlifi T. Arab. J. Chem. 10, 2017, S946-S955, doi: 10.1016/j.arabjc.2012.12.033

Press, 1981, pp. 221–247.

How to cite this article:
F Z. Fadel, N. Tchouar, S. Belaidi, F. Soualmia, O. Oukil, and K. Ouadah. Computational screening and QSAR study on a series of theophylline derivatives as aldh1a1 inhibitors. J. Fundam. Appl. Sci., 2021, 13(2), *942-964*.