Scientific Research

Physicochemical 2D-Qsar and 3D Molecular Docking Studies on N-Chlorosulfonyl Isocyanate Analogs as Sterol *O*-Acyl-Transferase-1 "Soat-1" Inhibitors

Khalid El Akri[1,2], Rokaya Mouhibi[1], Mohamed Zahouily[1*], Naîma Hanafi[2],
Moulay Abdellah Bahlaoui[1]

[1]Materials, Catalysis and Valorization of Natural Resources Laboratory (URAC 24), Faculty of Sciences and Techniques,
Hassan II University, Mohammedia, Morocco
[2]Organic Chemistry, Catalysis and Environment Laboratory, Faculty of Sciences,
Hassan II University, Casablanca, Marocco
Email: *mzahouily@yahoo.fr

ABSTRACT

A series of N-carbonyl-functionalized ureas, carbamates and thiocarbamates derivatives (or N-Chloro sulfonyl isocyanate "N-CSI") were involved in linear and nonlinear physicochemical quantitative structure-activity relationship "QSAR" analysis to find out the structural keys to control the inhibition against Sterol *O*-Acyl-Transferase-1 "SOAT-1". The results indicate the important effects of geometrical and chemical descriptors on the inhibitory activity of SOAT-1. The molecules were also screened for three-dimensional molecular docking on the crystal structure of ACAT-1 (1WL5 for ACAT-1, PDB). A comparison between 2D-QSAR and 3D molecular docking studies shows that the latter confirm the first results and represent a good prediction of the chemical and physical nature of interactions between our drug molecules and enzyme SOAT-1.

1. Introduction

Acyl-coenzyme A: Cholesterol *O*-Acyl-Transferase (ACAT) is an intracellular enzyme that catalyzes the formation of cholesterol ester from free cholesterol and long-chain fatty acyl-coenzyme A [1-3]. Recently this enzyme was named Sterol *O*-Acyl-Transferase (SOAT) to keep a difference between acyl-coenzyme A: Cholesterol *O*-Acyl-Transferase (ACAT) and acetyl-coenzyme A: acetyl-transferase or acetoacetyl-coenzyme A thiolase, also marked as ACAT [4]. Up to date, two SOAT isozymes have been identified [5,6]. SOAT-1 is an essential isozyme for intracellular storage of cholesteryl esters [7]. SOAT-2 is expressed exclusively in the small intestine and liver [8, 9]. In human liver, both of SOAT isozymes are expressed and may provide cholesteryl esters for very low density lipoprotein [10].

SOAT inhibitors have been developed over the last two decades as potential drugs for treatment of hyper-

cholesterolemia and atherosclerosis [11]. In general, SOAT inhibitors are thought to have two different pharmacological actions: suppression of cholesterol adsorption in the intestine that leads to reduction of plasma cholesterol level, and suppression of foam cell formation in the arterial walls [12]. Currently, cholesterol reduction can be achieved satisfactorily with so-called strong STATIN, such as atorvastatin and pritavastatin [13]. Thus, the direct suppression of foam cell formation in atherosclerotic lesions, which is independent of the cholesterol lowering effect, is a more attractive aspect of SOAT inhibitors. As SOAT-1 is the dominant isozyme expressed in macrophage, SOAT-1 is regarded as a target molecule of SOAT-1 inhibitors in the arterial walls. However, selective disruption of macrophage SOAT-1 by bone narrow transplantation from SOAT-1 null mice to LDL receptor-null mice resulted in exacerbation of diet-induced atherosclerosis. Therefore, it is still controversial whether pharmacological inhibition of SOAT-1 in macrophages contributes to exacerbation or amelioration

of atherosclerosis.

There is a large number of literature reports on the application of computational methods for describing the activity of biologically active compounds [14]. Quantitative structure-activity relationship studies are the most extensively used methods in computational chemistry. Appropriate representation of the structural and physicochemical features of chemical agents is an essential key to the successful application of QSAR models [15].

QSAR studies play a fundamental role in predicting the biological activity of new compounds and identifying ligand-receptor interaction. The first step in constructing the QSAR models is finding one or more molecular descriptors that represent variation in the structural property of the molecules by a number [16]. Structural descriptors have been classified into different categories according to different approaches including physiochemical, constitutional, geometrical, topological, and chemical descriptors. Currently, more than 1000 molecules descriptors can be easily calculated using available software such as Dragon [17].

There are different variable selection methods available including multiple linear regression (MLR) as linear method, and genetic algorithm, artificial neural network (ANN) as nonlinear methods. The relationship between molecular descriptors and activity is used to define the parameters affecting the biological activity and/or estimate the property of other molecules.

SOAT inhibitor compounds have been the subjects of QSAR studies to define statistical models describing the relationship between the structure and biological activity

into the two ways—linear and nonlinear. We used the structural invariants obtained from whole molecular structures of a series of 130 derivatives of the SOAT-1 inhibitors. We exploited two different chemometrics methods, *i.e.*, Multiple Linear Regression and Artificial Neural Network in order to make connections between structural parameters and SOAT inhibition. Finally, we performed the 3D molecular docking of these analogs on crystal of ACAT-1 (1WL5, PDB) to evaluate the binding energies as well as their mode of interaction.

2. Data and Methodology

2.1. Dataset Collection

An Intel® Core™ 2 Duo (CPU: T7300 2.00 GHz) with windows 7 and Linux Redhat operating systems were used. The two and three dimensional structures of molecules were drawn using ChemOffice v12 software package. The resulted geometry and Z-matrix were transferred into the Gaussian 98 program in order to perform the geometry and properties. The resulted geometry was performed with the semi-empirical AM1 method in MMP32 Pro program.

The biological data used in this study is SOAT inhibitory activity (in terms of $pIC_{50} = -Log_{10}IC_{50}$) of a set of 130 N-CSI derivatives. The data set was already synthesized and used by J. A. Picard *et al*. in 1996 for QSAR studies [18]. The structural features and biological activity of these compounds are listed in **Table 1 (annex)** and then used for subsequent QSAR analysis as dependent variable.

Table 1. Chemical structure representation of 130 N-CSI analogs (Family F) with the experimental value of IC_{50}.

Analog number	Reference	R_1	R_2	IC_{50}[*]	pIC_{50}[**]
1	CHEMBL277001			100	7,0000
2	CHEMBL281597			140	6,8539
3	CHEMBL22476			230	6,6383
4	CHEMBL284003			240	6,6198
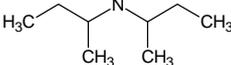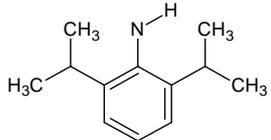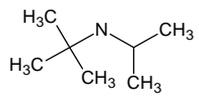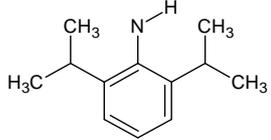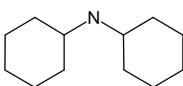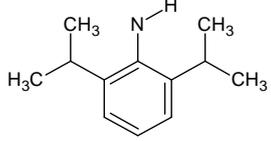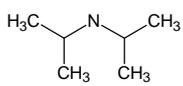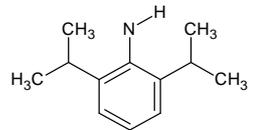
Continued
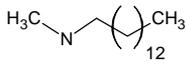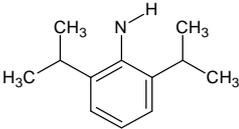
5	CHEMBL24887			330	6,4815
6	CHEMBL22673			360	6,4437
7	CHEMBL281391			360	6,4437
8	CHEMBL280349			400	6,3979
9	CHEMBL25655			420	6,3768
10	CHEMBL22415			420	6,3768
11	CHEMBL22834			420	6,3768
12	CHEMBL278131			510	6,2924
13	CHEMBL22774			720	6,1427
14	CHEMBL22733			1100	5,9586
15	CHEMBL279980			1200	5,9208
16	CHEMBL22896			1700	5,7696
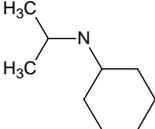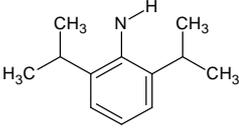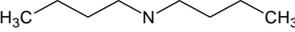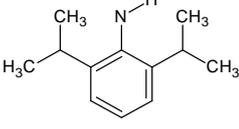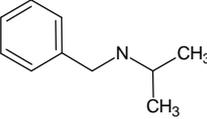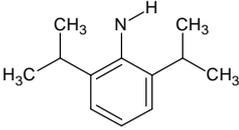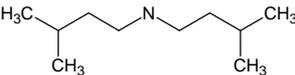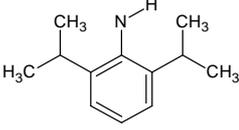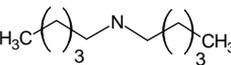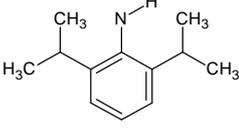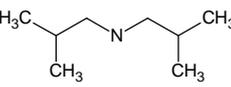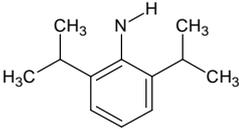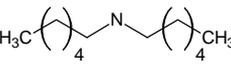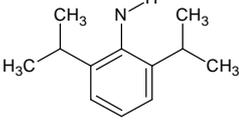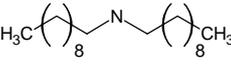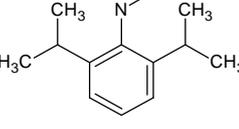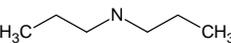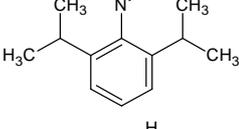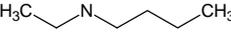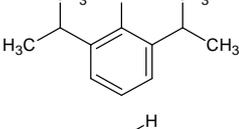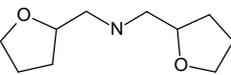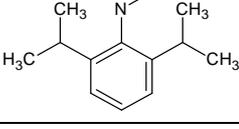
Continued
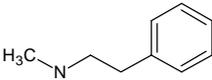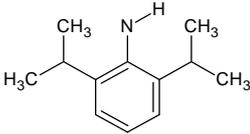
17	CHEMBL279727			1700	5,7696
18	CHEMBL22739			1900	5,7212
19	CHEMBL278132			1900	5,7212
20	CHEMBL21415			2100	5,6778
21	CHEMBL278808			2200	5,6576
22	CHEMBL22591			2300	5,6383
23	CHEMBL21586			2500	5,6021
24	CHEMBL22223			2700	5,5686
25	CHEMBL25705			3800	5,4202
26	CHEMBL21279			3900	5,4089
27	CHEMBL21583			4100	5,3872
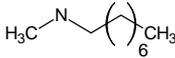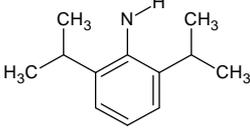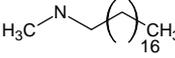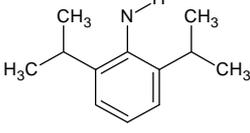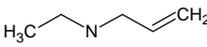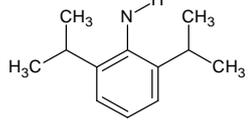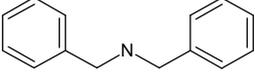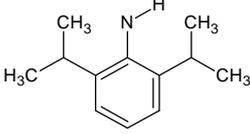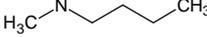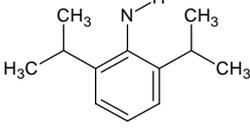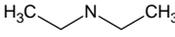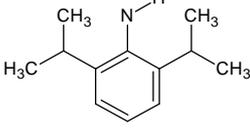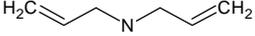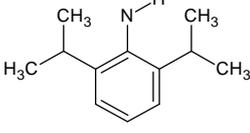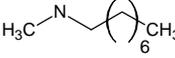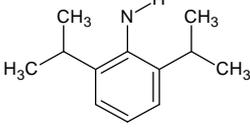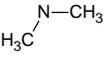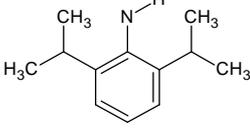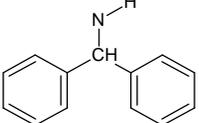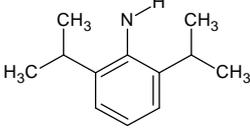
Continued

28	CHEMBL22891			5900	5,2291
29	CHEMBL22952			8100	5,0915
30	CHEMBL22135			10400	4,9830
31	CHEMBL22031			11000	4,9586
32	CHEMBL21863			15000	4,8239
33	CHEMBL22406			16000	4,7959
34	CHEMBL21976			16000	4,7959
35	CHEMBL22275			17000	4,7696
36	CHEMBL21981			18000	4,7447
37	CHEMBL21637			25000	4,6021
38	CHEMBL283531			30000	4,5229

Continued
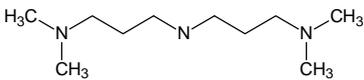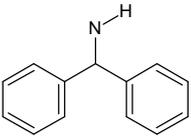
39	CHEMBL282718			55000	4,2596
40	CHEMBL277766			69000	4,1612
41	CHEMBL429228			1300	5,8861
42	CHEMBL279009			1900	5,7212
43	CHEMBL21535			2200	5,6576
44	CHEMBL277491			2700	5,5686
45	CHEMBL281188			3200	5,4949
46	CHEMBL22750			3200	5,4949
47	CHEMBL22889			4200	5,3768
48	CHEMBL422040			4300	5,3665
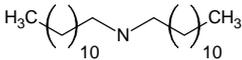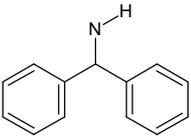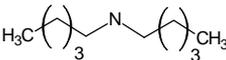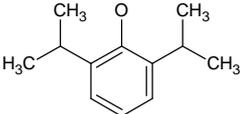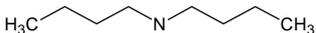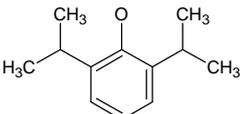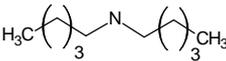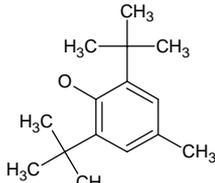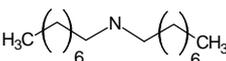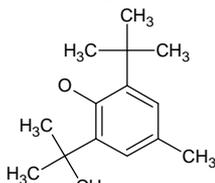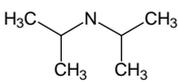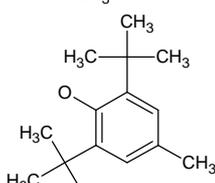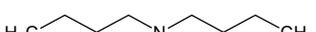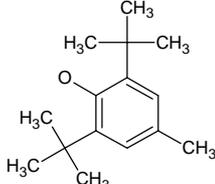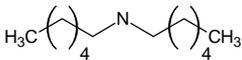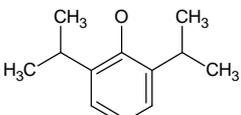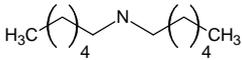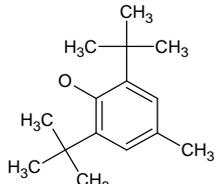
Continued
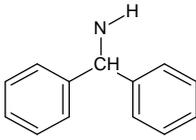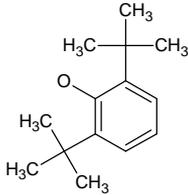
49	CHEMBL279928			4600	5,3372
50	CHEMBL22519			5100	5,2924
51	CHEMBL280187			10600	4,9747
52	CHEMBL22494			11000	4,9586
53	CHEMBL23043			11400	4,9431
54	CHEMBL22797			11500	4,9393
55	CHEMBL22335			12000	4,9208
56	CHEMBL21751			13500	4,8697
57	CHEMBL21794			15000	4,8239
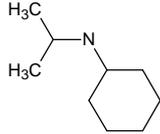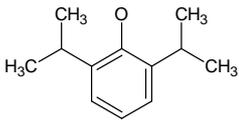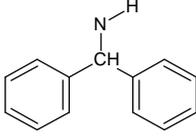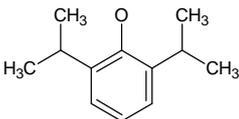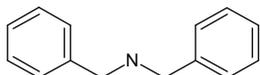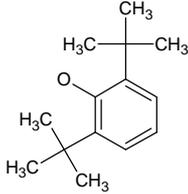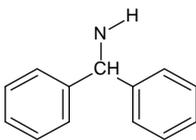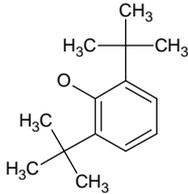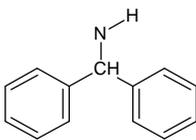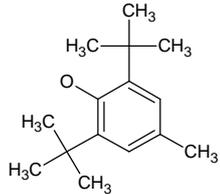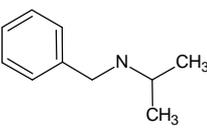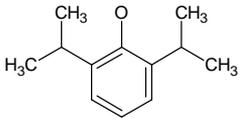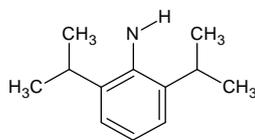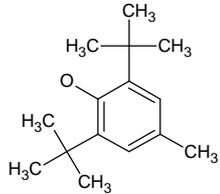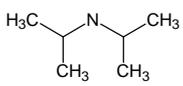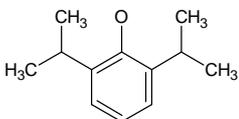
Continued
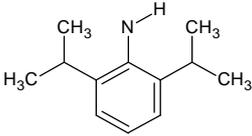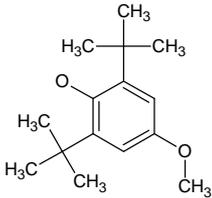
58	CHEMBL22552			15000	4,8239
59	CHEMBL21803			16800	4,7747
60	CHEMBL282437			18000	4,7447
61	CHEMBL277641			19000	4,7212
62	CHEMBL21687			19400	4,7122
63	CHEMBL277915			20000	4,6990
64	CHEMBL282051			21000	4,6778
65	CHEMBL22006			21000	4,6778
66	CHEMBL429222			21800	4,6615
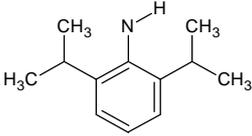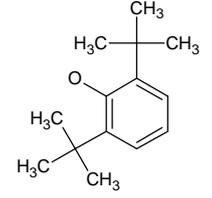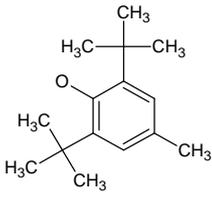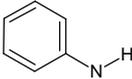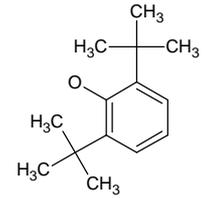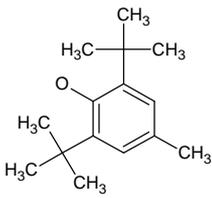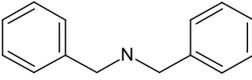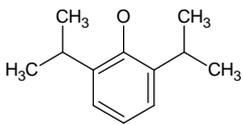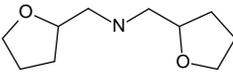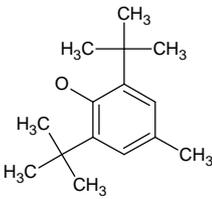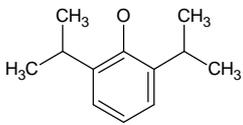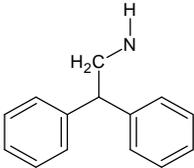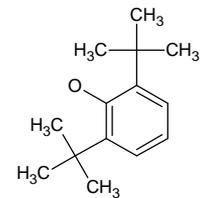
Continued
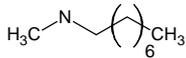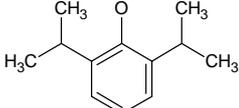
67	CHEMBL416183			22000	4,6576
68	CHEMBL22667			23400	4,6308
69	CHEMBL406139			24400	4,6126
70	CHEMBL277674			25800	4,5884
71	CHEMBL280732			30000	4,5229
72	CHEMBL22842			33000	4,4815
73	CHEMBL277914			38100	4,4191
74	CHEMBL279235			43000	4,3665
75	CHEMBL22217			49000	4,3098
76	CHEMBL22525			52000	4,2840
77	CHEMBL21797			53000	4,2757
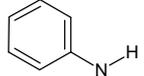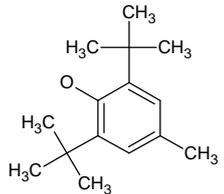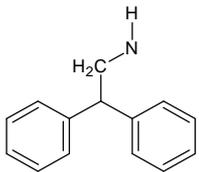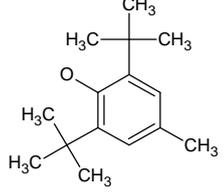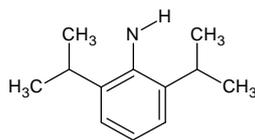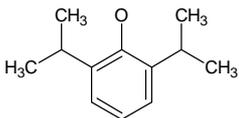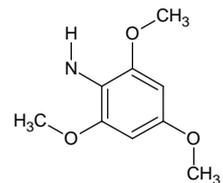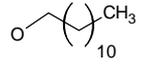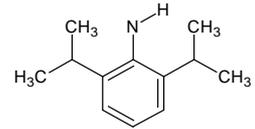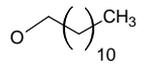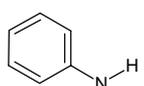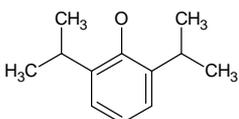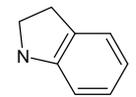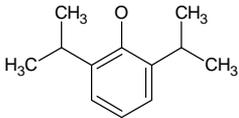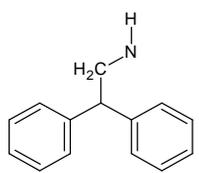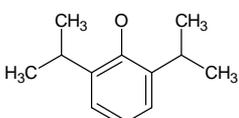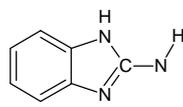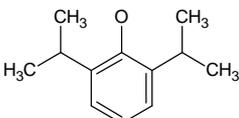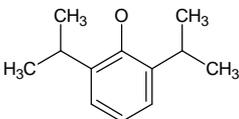
Continued
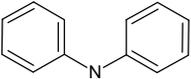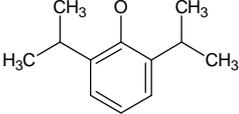
78	CHEMBL21870			58000	4,2366
79	CHEMBL22112			58000	4,2366
80	CHEMBL279920			82000	4,0862
81	CHEMBL22722			89000	4,0506
82	CHEMBL22496			164000	3,7852
83	CHEMBL22523			3600	5,4437
84	CHEMBL24891			4800	5,3188
85	CHEMBL416009			5700	5,2441
86	CHEMBL281332			7700	5,1135
87	CHEMBL22470			8800	5,0555
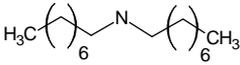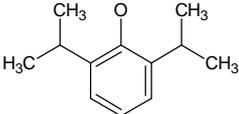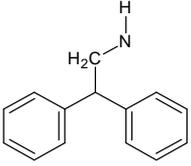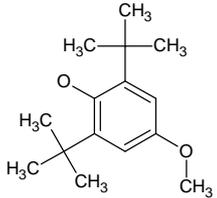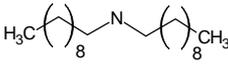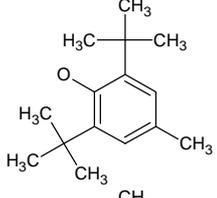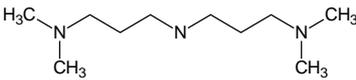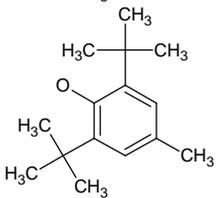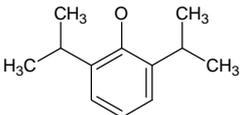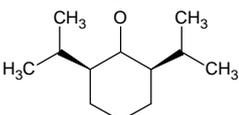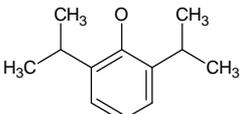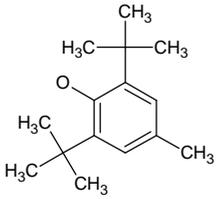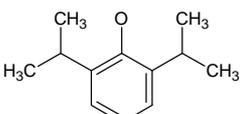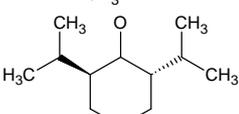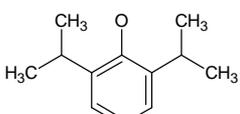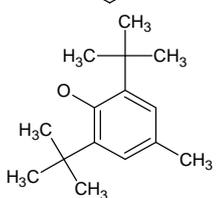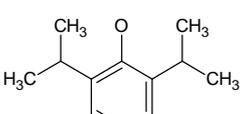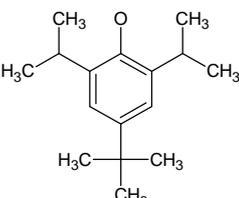
Continued
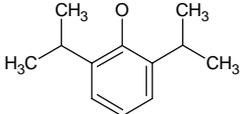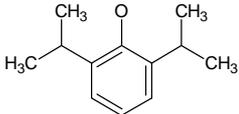
88	CHEMBL22279			9400	5,0269
89	CHEMBL22153			10300	4,9872
90	CHEMBL22839			12000	4,9208
91	CHEMBL25111			12000	4,9208
92	CHEMBL22836			13000	4,8861
93	CHEMBL24943			13000	4,8861
94	CHEMBL424218			13000	4,8861
95	CHEMBL25499			13000	4,8861
96	CHEMBL279927			14000	4,8539
97	CHEMBL22457			20000	4,6990
98	CHEMBL24992			21000	4,6778
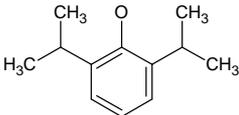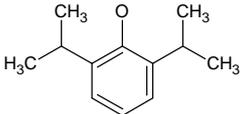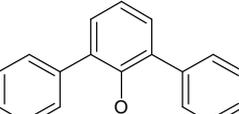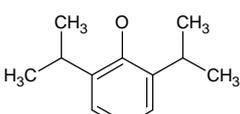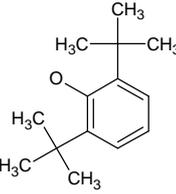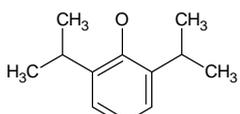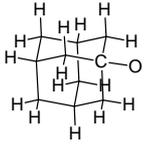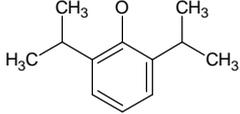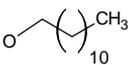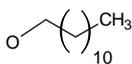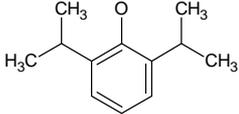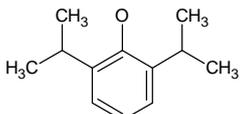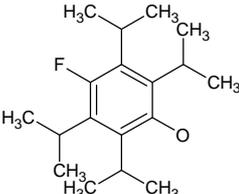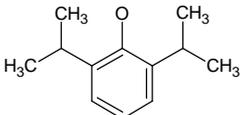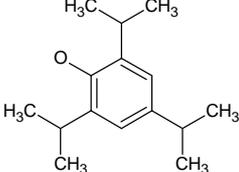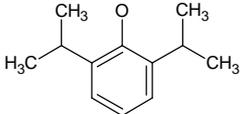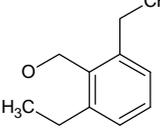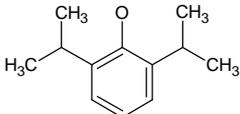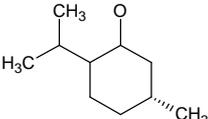
Continued
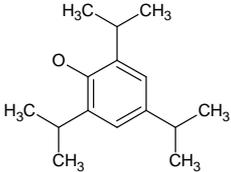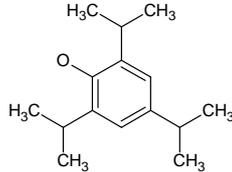
99	CHEMBL22804			22000	4,6576
100	CHEMBL423290			22000	4,6576
101	CHEMBL422929			23000	4,6383
102	CHEMBL282938			26000	4,5850
103	CHEMBL22409			26000	4,5850
104	CHEMBL280687			29000	4,5376
105	CHEMBL22725			30000	4,5229
106	CHEMBL21539			31000	4,5086
107	CHEMBL279513			38000	4,4202
108	CHEMBL21437			38000	4,4202
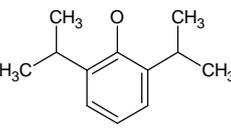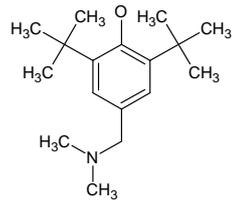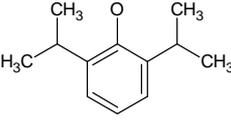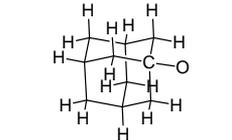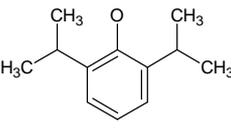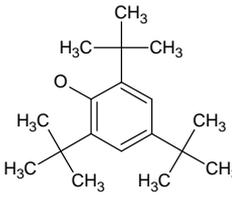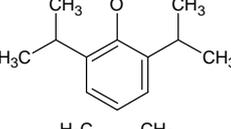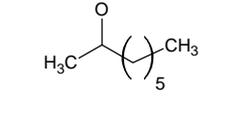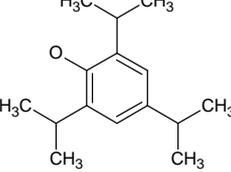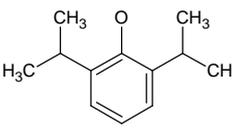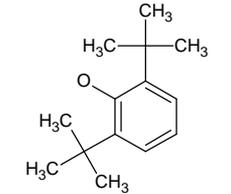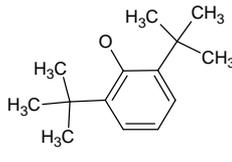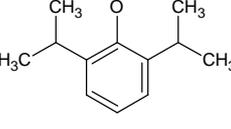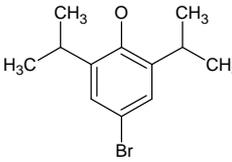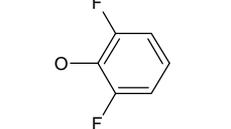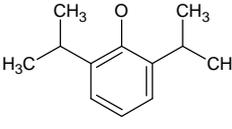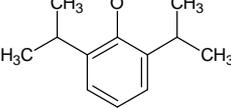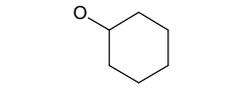
Continued
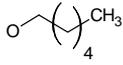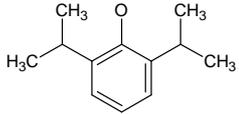
109	CHEMBL415036			40000	4,3979
110	CHEMBL22400			41800	4,3788
111	CHEMBL22297			42000	4,3768
112	CHEMBL278147			49000	4,3098
113	CHEMBL22407			50000	4,3010
114	CHEMBL22634			50000	4,3010
115	CHEMBL280686			52000	4,2840
116	CHEMBL22607			65000	4,1871
117	CHEMBL282671			66000	4,1805
118	CHEMBL22695			90000	4,0458
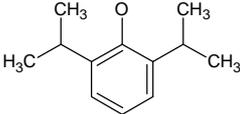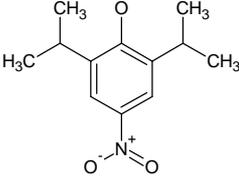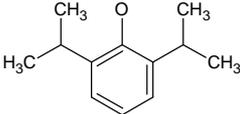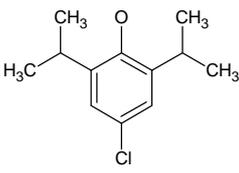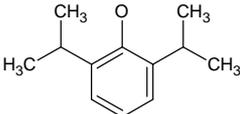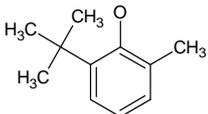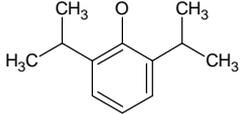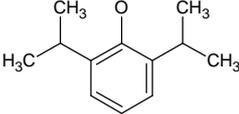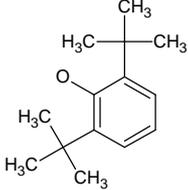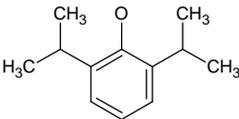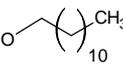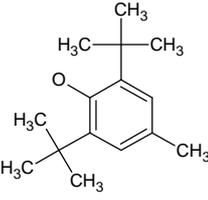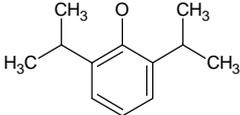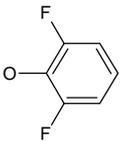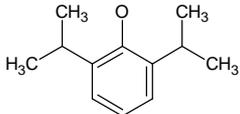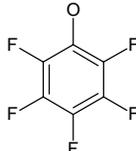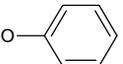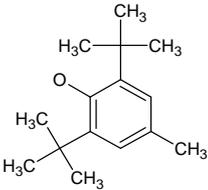
Continued
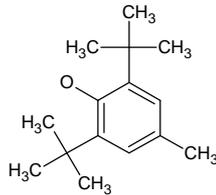
119	CHEMBL278161		95000	4,0223
120	CHEMBL22424		97000	4,0132
121	CHEMBL22985		6800	5,1675
122	CHEMBL283794		8700	5,0605
123	CHEMBL22631		12000	4,9208
124	CHEMBL281598		14000	4,8539
125	CHEMBL283532		16000	4,7959
126	CHEMBL281600		23000	4,6383
127	CHEMBL22014		27000	4,5686
128	CHEMBL22645		35000	4,4559
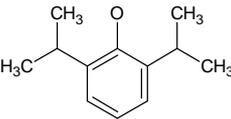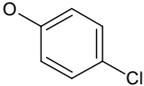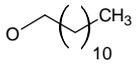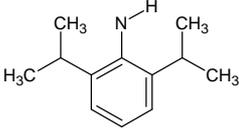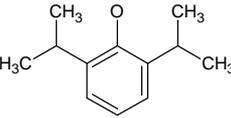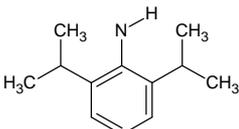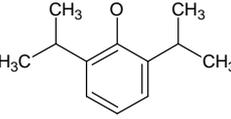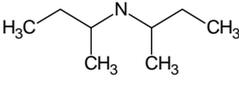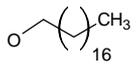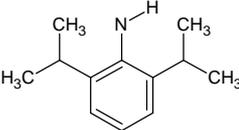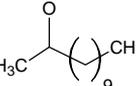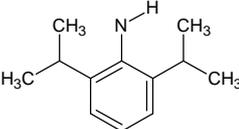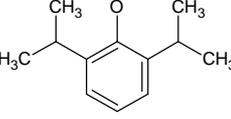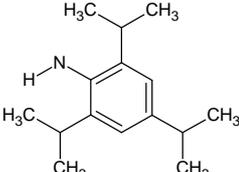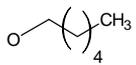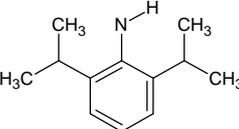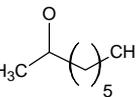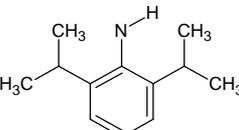
Continued
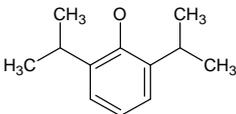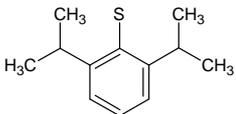
| 129 | CHEMBL22185 | | | 8700 | 5,0605 |
| 130 | CHEMBL283967 | | | 48000 | 4,3188 |
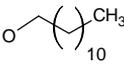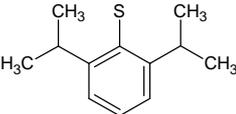
[*]Inhibitory concentration of 50% of the pathogen effect as (nM); [**]$pIC_{50} = -Log_{10}(IC_{50})$ with IC_{50} by Mole.

All of our molecular descriptors were calculated using the MMP32 Pro and Dragon packages. Some chemical parameters including molecular volume, molecular surface area, hydration energy and molecular polarizability were calculated using the ChemOffice v12 Software. The Gaussian program was employed for calculation of different quantum chemical descriptors including dipole moment, local charges, HOMO and LUMO energies.

2.2. Variable Selection and 2D-QSAR Model Generation

For each family of N-CSI derivatives (**Figure 1**), separate linear and nonlinear QSAR models were constructed. The development of QSAR equations has been done with two different methods: stepwise multiple linear regression and fitting function with artificial neural network. The selection of significant descriptors, which constructs a relationship between the biological activity data and the molecular structures, is an important step in QSAR modeling. Selection of significant descriptors was performed trough the following steps:

1) The calculated descriptors were collected in a data matrix, whose number of rows and columns were the number of molecules and descriptors, respectively. First the descriptors were checked for constant or near constant values and those detected were removed from the original data matrix. The correlation of descriptors with each other's and with the activity data was determined.

2) The input variable in MLR must not be highly correlated. Among the collinear descriptors detected ($r > 0.8$) one with the highest correlation with the activity was retained and the rest were omitted.

3) The selected descriptors from each class and the experimentally inhibition of SOAT data were analyzed by the stepwise regression with SATISTICA 8.0 Software.

Multiple linear regression was used to generate linear models between the inhibitory activity and the molecular descriptors. With consideration of large number of descriptors used in this study, we have used forward and backward stepwise algorithms to select the pertinent descriptors [19]. The predictive activity of the model is quantified in terms of r^2 which is defined as **Equation 1**.

Equation 1:

$$r^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}.$$

In this equation y_i and $\hat{y}_i$ are the predicted and the experimental values of the target property for the observation i respectively. The mean value of target property is noted as $\overline{y}$ and r^2 is the internal correlation coefficient.

To keep around 5% of the experimental error, we take away the molecules having d_i higher than $\frac{2s}{moy}$ as defined in **Equation 2**.

Equation 2:

$$d_i = \frac{|Obs_i - Cal_i|}{Obs_i} > \frac{2s}{moy_{(Obs)}},$$

with:

$moy_{(Obs)}$: The mean of the observed activity values.
s: Standard deviation.
OBs_i: Observed value of the activity for the i case.
Cal_i: Calculated value of the activity for the i case.

Leave-one-out (LOO) cross-validation was used to evaluate the predictive ability of the models. The cross-validated coefficient q^2 was calculated using **Equation 3** as follows:

Equation 3:

$$q^2 = 1 - \frac{PRESS}{Var} = 1 - \frac{\sum_{i=1}^{n}(Y_{predicted} - Y_{actual})^2}{\sum_{i=1}^{n}(Y_{observed} - Y_{mean})^2},$$

where $Y_{predicted}$, Y_{actual} and Y_{mean} are predicted, actual and mean values of the target property (pIC_{50}), respectively. $\sum_{i=1}^{n}(Y_{predicted} - Y_{actual})^2$ is the predictive sum of squares (PRESS). The optimum number of components used to derive the final regression models was the one that corresponds to the lowest PRESS value. In addition to the q^2, the corresponding PRESS, the conventional correlation coefficient r^2 and its standard errors s were also computed. To test the stability and robustness of the models, more rigorous statistical tests were performed by group cross-validation to eliminate the possibility of chance correlation.

The artificial neural network (ANN) consists of an input layer, an output layer and a number of hidden layers. At each node in a layer the information is received, stored, processed and communicated further to nodes in the next layer. All the weights are initialized to small random numeric values at the beginning of training. These weights are updated or modified iteratively using the generalized delta rule or steepest-gradient descent principle. The training process is stopped when no appreciable change is observed in the values associated with the connection links or some termination criterion is satisfied. Thus, the training of back-propagation network consists of two phases: a forward pass during which the processing of information occurs from the input layer to the output and a backward pass when the error from the output layer is propagated back to the input layer and the interconnections are modified. An example of a ANN topology is shown in **Figure 2**.

The contribution of descriptors $i : (i = \{1, n\})$ was estimated from the [n-m-1] neural network architecture
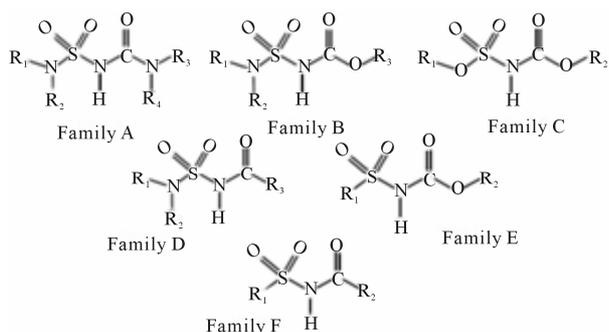
Figure 1. The family F represents the chemical structure of N-CSI analogs. The families A-E represent the sub-chemical structures with different radicals at each level.
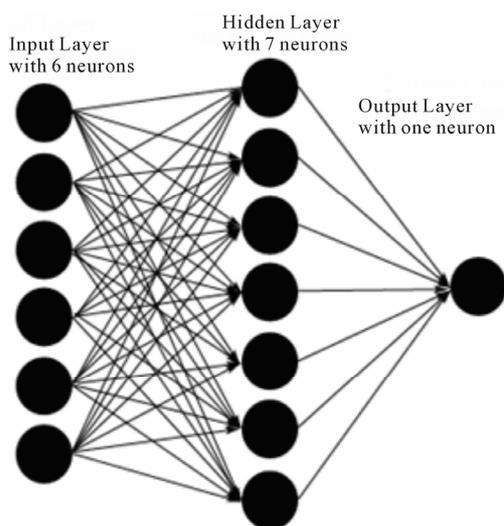
Example of ANN with [6-7-1] architecture

Figure 2. Example of artificial neural network with [6-7-1] architecture.

with n number of descriptors and m number of neurons in hidden layer. The descriptor under study was removed from the ANN calculated the output of each molecule as usual. The mean of the absolutes deviations value Δm_i between the observed activity and the predicted one for all compounds was determined. This process was reiterated for each descriptor. Finally, the contribution C_i of each descriptor i is given by **Equation 4**.

Equation 4:

$$C_i (\%) = \frac{\Delta m_i}{\sum_{i=1}^{n} \Delta m_i},$$

with:

Δm_i : Mean value of absolutes deviations between predicted and calculated activity.

$\sum_{i=1}^{n} \Delta m_i$: Sum of means values of absolutes deviations between predicted and calculated activity for n descriptors.

2.3. Alignment of Compounds and 3D Molecular Docking

All the dataset molecules were aligned (**Figure 3**) according to the binding models resulting from docking simulations. Each compound was docked within the ATP binding pocket of the active ACAT-1 site using the program Molegro [20]. The recently reported X-ray crystallographic structure (2.3 Å resolution; PDB accession code: 1WL4 and 1WL5) of ACAT-1 in complex with ATP (**Figure 4**) was chosen as template for SOAT-1
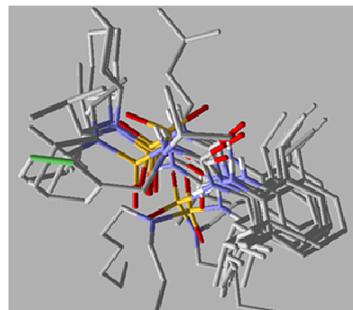
Figure 3. Alignment of the N-CSI analogs.
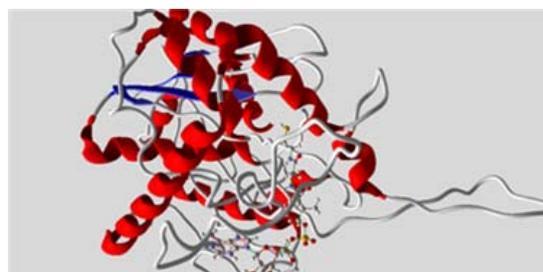
Figure 4. Representation of ACAT-1 from the X-ray data (PDB accession code: 1WL4 and 1WL5) in complex with ATP molecule.

target, based on resemblance between the co-crystallized ligand and the compounds under analysis. The X-ray structure was manipulated to prepare the input for Molegro calculations by removing both non-polar hydrogens and water molecules. The binding pocket was inserted into a grid box centred on the bound ligand and enclosing residues lying within about 11 Å from the ligand itself.

The Molegro program was used to add the hydrogens and establish the protonation state at the physiological pH for all the inhibitors, before adding the Gasteiger atomic charges. Finally, the rigid root and the rotatable bonds for each compound were defined using the module implemented on Molegro program. The MolDock algorithm was employed to explore the possible orientations and conformations of SOAT-1 inhibitors in the binding site. For each of the 100 independent runs, a maximum number of 1500,000 operations were performed with a population size of 200 individuals. Finally, the best ranked conformation of each molecule was selected. The scoring function and electronic, electrostatic and hydrophobicity interactions between ligand and analogs were deeply analyzed to extract the right and useful chemical information.

3. Results and Discussion

3.1. Physicochemical 2D-QSAR Studies of N-CSI Analogs

In order to study the effect of different levels of a radical analogues N-CSI, we divided the initial database (Family F) into two databases (Families D and E) while showing in each case a nitrogen atom and an oxygen atom in the basic skeleton. Then we divided the database (Families D and E) into three databases (Families A, B and C) to investigate further the effect and the nature of the atoms on the inhibitory activity of the enzyme SOAT-1.

For each database we have conducted a linear and nonlinear QSAR analysis to determine the nature of this relationship and the relevant physico-chemical descriptors. The specificity of this 2D-QSAR study is of using physicochemical descriptors to exploit chemical information for a useful design of new N-CSI generation analogs. The sets were devised into two groups using hierarchical ascending classification HAC to make groups with high similarity. **Table 2** summarizes the final linear models equations of the datasets (Families A to F) for training and validation sets.

Tables 3 and **4** include the linear and nonlinear statistical results, respectively, for each database while mentioning the number of validated molecules n, the correlation coefficient r, the standard deviation s, the Fischer test F and the coefficient of cross-validation q^2 with architecture of the neural network adopted.

The 2D-QSAR study of Family A, has allowed us to validate the hydrogen bond donor of radical R_4 and the molecular width of the radical R_1 as relevant descriptors with a contribution of 53% and 47% respectively. The Family B has allowed us to validate the lipophilicity of the radical R_1 as relevant descriptor with a contribution of 100%. Finally, for the Family D, we have validated the molecular volume, lipophilicity and hydrogen bond donor molecule as relevant descriptors with a contribution of 0.85%, 20.79% and 78.36% respectively.

The 2D-QSAR study of Families A, B, and D allowed

Table 2. Validated linear model equations of all datasets (Families A to F).

Family		*Linear model equations*
A	traning set	$pIC_{50} = (5.5483 \pm 0.2794) + (0.2354 \pm 0.05014)MW(R_1) - (7.6465 \pm 1.0913)HBD(R_4)$
	validation set	$pIC_{50} = (3.9694 \pm 0.6938) + (0.3934 \pm 1.77)MW(R_1) - (0.3927 \pm 1.9087)HBD(R_4)$
B	traning set	$pIC_{50} = (4.6586 \pm 0.0662) + (0.2190 \pm 0.0401)LogP(R_1)$
	validation set	$pIC_{50} = (4.5991 \pm 0.0683) + (0.2179 \pm 0.0281)LogP(R_1)$
C	traning set	$pIC_{50} = (4.3269 \pm 0.6291) - (0.0056 \pm 0.0029)PM - (0.0120 \pm 0.0048)VM$ $+ (0.0248 \pm 0.0079)VM(R_1) + (0.0238 \pm 0.0077)VM(R_2)$
	validation set	$pIC_{50} = (10.0674 \pm 1.4761) - (0.0268 \pm 0.0037)PM - (0.0238 \pm 0.0156)VM$ $+ (0.0622 \pm 0.0229)VM(R_1) + (0.0487 \pm 0.0076)VM(R_2)$
D	traning set	$pIC_{50} = (7.6408 \pm 0.3963) - (0.0165 \pm 0.0020)VM + (0.3192 \pm 0.0269)LogP - (1.9356 \pm 0.2336)HBE$
	validation set	$pIC_{50} = (8.4862 \pm 1.6241) - (0.0179 \pm 0.0066)VM + (0.3125 \pm 0.0831)LogP - (2.4708 \pm 0.8399)HBE$
E	traning set	$pIC_{50} = (3.4144 \pm 0.4032) - (0.0809 \pm 0.0224)ML + (0.0941 \pm 0.0234)ML(R_1) + (0.1244 \pm 0.0387)ML(R_2)$
	validation set	$pIC_{50} = (2.6575 \pm 0.4571) - (0.0793 \pm 0.0412)ML + (0.1763 \pm 0.0324)ML(R_1) + (0.0982 \pm 0.0479)ML(R_2)$
F	traning set	$pIC_{50} = (2.3137 \pm 0.3305) + (3.2623 \pm 0.3389)HBD - (3.7888 \pm 0.5095)HBD(R_1) + (0.1108 \pm 0.0430)LogP(R_2)$
	validation set	$pIC_{50} = (1.8908 \pm 0.7775) + (4.0882 \pm 0.6879)HBD - (3.4670 \pm 1.3233)HBD(R_1) + (0.0672 \pm 0.0925)LogP(R_2)$

Table 3. Linear statistical results and leave-one-out cross-validation coefficient of all datasets (Families A to F).

Family		n	r	s	F	q^2_{LOO}
A	*training set*	24	0.862	0.380	30.57	0.713
	validation set	10	0.864	0.389	10.34	
B	*training set*	27	0.737	0.263	29.88	0.71
	validation set	6	0.965	0.137	54.481	
C	*training set*	31	0.666	0.275	5.207	0.71
	validation set	7	0.984	0.133	15.889	
D	*training set*	57	0.863	0.383	51.709	0.693
	validation set	14	0.866	0.461	10.011	
E	*training set*	60	0.482	0.364	5.6418	0.531
	validation set	15	0.877	0.271	12.314	
F	*training set*	97	0.714	0.468	32.388	0.531
	validation set	24	0.808	0.490	12.61	

Table 4. Nonlinear statistical results and leave-20%-out cross-validation coefficient of all datasets (Families A to F).

Family	n	r	s	ANN	$q^2_{L20\%O}$
A	34	0.904	0.325	2-8-1	0.81
B	33	0.74	0.244	1-6-1	0.64
C	38	0.865	0.196	4-6-1	0.61
D	71	0.911	0.318	3-5-1	0.66
E	75	0.625	0.309	3-3-1	0.651
F	121	0.827	0.379	3-7-1	0.62

us to conclude that inhibitors of SOAT-1 must submit electronic properties of hydrogen bond donor and lipophilicity of the side of the sulfonyl group of the molecule, and the larger contribution confirms these descriptors.

The 2D-QSAR study of Families B, C and E, has allowed us to validate the molecular weight, molecular volume of the molecule, the radical R_1 and radical R_2 as the 4 relevants descriptors with a contribution of 18%, 19.77%, 27.55% and 34.22% respectively for Family C. Family E has allowed us to validate the molecular length of the molecule, the radical R_1 and the radical R_2 as relevant descriptors with a contribution of 36.54%, 40.25% and 23.21% respectively.

We concluded that SOAT-1 inhibitors must present steric properties such as molecular volume and molecular length of the side of the sulfonyl group and the side of the carbonyl group of the molecule. The high contributions of these descriptors confirm it. Finally, from the 2D-QSAR studies of Families D, E and F, we concluded that the hydrogen bond donor of the molecule and the

radical R_1 with the lipophilicity of the radical R_2 are the relevant descriptors with a contribution of 52.59%, 37.82% and 9.58% respectively.

We deduced that SOAT-1 inhibitors should exhibit properties of both electronic and steric hydrogen bond donor of the sulfonyl group side, namely the radical R_1, and lipophilicity of the carbonyl group side, namely the radical R_2 of the molecule. The high contribution of the hydrogen bond donor descriptor confirms that the inhibitory activity is mainly governed by the electronic properties.

On the other hand, comparison of linear (**Table 5**) and nonlinear results allowed us to conclude that the relationship between the chemical structure of N-CSI analogs and their inhibitory activity of the enzyme SOAT-1 is a non-linear relationship and the correlation coefficient values confirm it ($r_{\text{linear}} < r_{\text{non-linear}}$) share a Family B where the relationship is rather a linear relationship. By comparing the coefficients of cross-validation we find that nonlinear models are robust as $q^2 > 0.6$ (depending on the conditions of WOLD [21]) and the two linear models (Families C and E) are not.

3.2. Three Dimensional Molecular Docking Studies of N-CSI Analogs

In order to study the pharmacodynamics of each molecule derived from 130 N-CSI analogs and find the best relationship between the inhibitor and the protein SOAT-1, taking into account of interactions that exist before and after complex formation, we have performed a study of molecular docking with the program MOLE-GRO.

We conducted an analysis of molecular docking of ten

Table 5. Pertinent descriptors and their contributions in both linear and nonlinear analysis.

Family	Pertinent descriptors	Linear contribution	Non-linear contribution
A	Molecular width of radical R_1	47%	49.08%
	Hydrogen bond donor of radical R_4	53%	50.92%
B	Lipophily	100%	100%
C	Molecular weights	18.47%	13.6%
	Molecular volume	19.77%	12.17%
	Molecular volume of R_1	27.55%	32.53%
	Molecular volume of R_2	34.22%	41.70%
D	Molecular volume	0.85%	49.2%
	Lipophily	20.71%	32.16%
	Hydrogen bond donor	78.36%	18.64%
E	Molecular length	36.54%	43.7%
	Molecular length of radical R_1	40.25%	34.21%
	Molecular length of radical R_2	23.21%	22.09%
F	Hydrogen bond donor	52.595	13.94%
	Hydrogen bond donor of radical R_1	37.82%	81.73%
	Lipophily of radical R_2	9.58%	4.32%

best SOAT-1 inhibitors. The molecules 1 to 10 show the chemical structures of ten molecules that have been selected for this analysis after lead identification analysis using MOLEGRO. We followed the methodology described above (see Section 8.3) in the methodology section.

By analyzing the results we find that the pose No 01 of the ligand "Binding_DB_50050205" has both a better scoring function −123.383 and better interaction of hydrogen bonds with a total energy of −5.86. The ligand "Bindin_DB_50050205" corresponds to the molecule 9 with reference "CHEMBL22415".

By analyzing more closely the hydrogen interactions, we find that the ligand "CHEMBL22415" has three hydrogen bonds lengths 3.18, 3.02 and 2.67 with −1.174, −2.193 and −2.55 of energy respectively. The highlight of these hydrogen bonds in space (**Figure 5**) shows that the atoms of the urea function are the only responsible of this type of interaction.

In order to exhibit the effect of electrostatic and hydrophobic of radicals and the whole molecule on the pocket, we performed an analysis of these interactions with the program MOLEGRO. The results are shown in

Figures 6 and **7**. We conclude that the ligand has the electrostatic interactions of positive values on the left side of the cavity (shown in red) and negative electrostatic interactions values on the right side of the cavity (shown in blue).

Analysis of the hydrophobicity of the ligand in the cavity allowed us to conclude that the pocket has mainly hydrophobic regions (represented by the blue color **Figure 7**) and small areas rather hydrophilic outside the cavity (represented by the red color in the same figure). This high hydrophobicity of the cavity and the major scoring function of the ligand are vouched for the presence of aliphatic chains on the nitrogen atom of the urea function and isopropyl groups on the benzene ring.
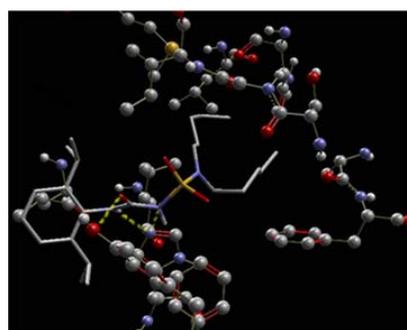
Figure 5. Highlighting three hydrogen bonds on the complex ligand-biological target (dashed lines in green).
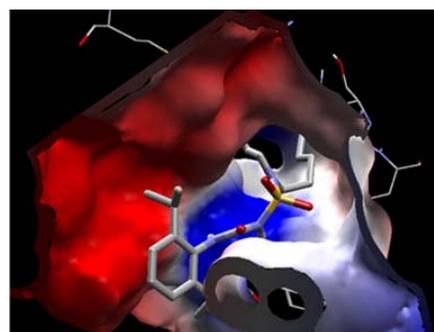
Figure 6. Highlighting the electrostatic interaction between the ligand and the biological target.
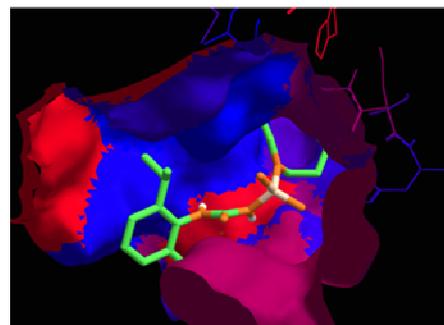
Figure 7. Highlighting the hydrophobicity of the ligand with the biological target.

4. Conclusions

In this paper, we have demonstrated that the ligand (N-CSI analogs) and the biological target (SOAT-1) have significant electronic interactions with the presence of three hydrogen bonds donor between the atoms of the urea function and the biological target. In addition, the presence of electrostatic and hydrophobic interactions is strongly justified by the presence of aliphatic chains on the nitrogen atom on urea function and the isopropyl radicals on the benzene ring. These results confirm the 2D-QSAR analysis.

Taken together, the 2D-QSAR model and the computational analysis of the pharmacodynamics properties described herein will constitute a valuable tool for the design of novel structurally related N-CSI inhibitors endowed with increased affinities toward SOAT-1 and improved ADME profiles.

REFERENCES

[1] N. Khelef, X. Buton, N. Beatini, H. Wang, V. Meiner, T. Y. Chang, R. V. Farese, F. R. Maxfield and I. Tabas, "Immunolocalization of Acyl-Coenzyme A: Cholesterol O-Acyltransferase in Macrophages," *The Journal of Bio-Logical Chemistry*, Vol. 273, No. 18, 1998, pp. 11218-11224. http://dx.doi.org/10.1074/jbc.273.18.11218

[2] V. Lather and A. K. Madan, "Predicting Acyl-Coenzyme A: Cholesterol O-Acyltransferase Inhibitory Activity: Computational Approach Using Topologicaldescriptors," *Drug Design and Discovery*, Vol. 18, No. 4, 2003, pp. 117-122.

[3] M. A. Diczfalusy, I. Björkhem, K. Einarsson and S. E. Alexson, "Acyl-Coenzyme A: Cholesterol O-Acyltransferase Is Not Identical to Liver Microsomal Carboxylesterase," *Arteriosclerosis, Thrombosis, and Vascular Biology*, Vol. 16, No. 4, 1996, pp. 606-610. http://dx.doi.org/10.1161/01.ATV.16.4.606

[4] P. Kursula, H. Sikkilä, T. Fukao, N. Kondo and R. K. Wierenga, "High Resolution Crystal Structures of Human Cytosolic Thiolase (CT): A Comparison of the Active Sites of Human CT, Bacterial Thiolase, and Bacterial KAS I," *Journal of Molecular Biology*, Vol. 347, No. 1, 2005, pp. 189-201. http://dx.doi.org/10.1016/j.jmb.2005.01.018

[5] B. Ferraz-de-Souza, R. E. Hudson-Davies, L. Lin, R. Parnaik, M. Hubank, M. T. Dattani and J. C. Achermann, "Sterol O-Acyltransferase 1 (SOAT1, ACAT) Is a Novel Target of Steroidogenic Factor-1 (SF-1, NR5A1, Ad4BP) in the Human Adrenal," *The Journal of Clinical Endocrinology & Metabolism*, Vol. 96, No. 4, 2011, pp. 663-668.

[6] M. R. Lafave, L. Katz, T. Donnon and D. J. Butterwick, "Initial Reliability of the Standardized Orthopedic Assessment Tool (SOAT)," *Journal of Athletic Training*, Vol. 43, No. 5, 2008, pp. 483-488. http://dx.doi.org/10.4085/1062-6050-43.5.483

[7] K. Hsieh, Y. K. Lee, C. Londos, B. M. Raaka, K. T. Dalen and A. R. Kimmel, "Perilipin Family Members Pref-

[8] W. S. Lee, K.-R. Im, Y.-D. Park, N.-D. Sung and T.-S. Jeong, "Human ACAT-1 and ACAT-2 Inhibitory Activities of Pentacyclictriterpenes from the Leaves of Lycopuslucidus TURCZ," *Biological & Pharmaceutical Bulletin*, Vol. 29, No. 2, 2006, pp. 382-384. http://dx.doi.org/10.1248/bpb.29.382

[9] A. Miyazaki, N. Sakashita, O. Lee, K. Takahashi, S. Horiuchi, H. Hakamata, P. M. Morganelli, C. C. Chang and T. Y. Chang, "Expression of ACAT-1 Protein in Human Atherosclerotic Lesions and Cultured Human Monocytes-Macrophages," *Arteriosclerosis, Thrombosis, and Vascular Biology*, Vol. 18, No. 10, 1998, pp. 1568-1574. http://dx.doi.org/10.1161/01.ATV.18.10.1568

[10] S. Saha, S. R. Bornstein, J. Graessler and S. Kopprasch, "Very-Low-Density Lipoprotein Mediates Transcriptional Regulation of Aldosterone Synthase in Humanadrenocortical Cells," *Cell and Tissue Research*, Vol. 348, No. 1, 2012, pp. 71-80. http://dx.doi.org/10.1007/s00441-012-1346-3

[11] Y. Asami, I. Yamagishi, S. Murakami, H. Araki, K. Tsuchida and S. Higuchi, "HL-004, the ACAT Inhibitor, Prevents the Progression of Atherosclerosis in Cholesterol-Fed Rabbits," *Life Sciences*, Vol. 62, No. 12, 1998, pp. 1055-1063. http://dx.doi.org/10.1016/S0024-3205(98)00028-9

[12] N. Sakashita, A. Miyazaki, M. Takeya, S. Horiuchi, C. C. Y. Chang, T.-Y. Chang and K. Takahashi, "Localization of Human Acyl-Coenzyme A: Cholesterol Acyltransferase-1 (ACAT-1) in Macrophages and in Various Tissues," *The American Journal of Pathology*, Vol. 156, No. 1, 2000, pp. 227-236. http://dx.doi.org/10.1016/S0002-9440(10)64723-2

[13] J. Liao and J. A. Farmer, "Aggressive Statin Therapy and the Risk of Malignancy," *Current Atherosclerosis Reports*, Vol. 15, No. 4, 2013, p. 316. http://dx.doi.org/10.1007/s11883-013-0316-x

[14] R. A. Kendall, E. Aprà, D. E. Bernholdt, E. J. Bylaska, M. Dupuis, G. I. Fann, R. J. Harrison, J. Ju, J. A. Nichols, J. Nieplocha, T. P. Straatsma, T. L. Windus and A. T. Wong, "High Performance Computational Chemistry: An Overview of NWChem a Distributed Parallel Application," *Computer Physics Communications*, Vol. 128, No. 1-2, 2000, pp. 260-283. http://dx.doi.org/10.1016/S0010-4655(00)00065-5

[15] H. Kubinyi, "QSAR and 3D QSAR in Drug Design. Part 1: Methodology," *Drug Discovery Today*, Vol. 2, No. 11, 1997, pp. 457-467. http://dx.doi.org/10.1016/S1359-6446(97)01079-9

[16] P. Kolb, R. S. Ferreira, J. J. Irwin and B. K. Shoichet, "Docking and Chemoinformatic Screens for New Ligands and Targets," *Current Opinion in Biotechnology*, Vol. 20, No. 4, 2009, pp. 429-436. http://dx.doi.org/10.1016/j.copbio.2009.08.003

[17] L. Xue and J. Bajorath, "Molecular Descriptors in Chemoinformatics, Computational Combinatorial Chemistry,

and Virtual Screening," *Combinatorial Chemistry & High Throughput Screening*, Vol. 3, No. 5, 2000, pp. 363-372. http://dx.doi.org/10.2174/1386207003331454

[18] J. A. Picard, P. M. O'Brien, D. R. Sliskovic, M. K. Anderson, R. F. Bousley, K. L. Hamelehle, B. R. Krause and R. L. Stanfield, "Inhibitors of Acyl-CoA: Cholesterol O-Acyltransferase. 17. Structure-Activity Relationships of Several Series of Compounds Derived from N-Chlorosulfonylisocyanate," *Journal of Medicinal Chemistry*, Vol. 39, No. 6, 1996, pp. 1243-1252. http://dx.doi.org/10.1021/jm9509455

[19] X. Li, "A Simulation Evaluation of Backward Elimination and Stepwise Variable Selection in Regression Analysis," M.S. Thesis, Shandong Polytechnic University, Shandong, 2012.

[20] P. A. Babu, V. T. S. S. Colluru and N. Anaparthy, "*In-Silico* Characterization of ECE-1 Inhibitors," *Computers in Biology and Medicine*, Vol. 42, No. 4, 2012, pp. 446-457. http://dx.doi.org/10.1016/j.compbiomed.2011.12.013

[21] D. J. Livingstone and D. T Manallack, "Statistics Using Neural Networks: Chance Effects," *Journal of Medicinal Chemistry*, Vol. 36, No. 9, 1993, pp. 1295-1297. http://dx.doi.org/10.1021/jm00061a023