BABEŞ-BOLYAI UNIVERSITY
FACULTY OF CHEMISTRY AND CHEMICAL ENGINEERING

Harsa Teodora Elena

Abatract

QSAR Study of nitrogen heterocycles

SCIENTIFIC COORDINATOR
Prof.Dr. MIRCEA V. DIUDEA

PHD STUDENT
HARŞA TEODORA ELENA

Cluj-Napoca

2015

Contents

Investeşte în oameni !
FONDUL SOCIAL EUROPEAN
Programul Operaţional Sectorial Dezvoltarea Resurselor Umane 2007 – 2013
Axa prioritară 1. „Educaţia şi formarea profesională în sprijinul creşterii economice şi dezvoltării societăţii bazate pe cunoaştere”
Domeniul major de intervenţie 1.5. „Sprijin pentru doctoranzi şi cercetători postdoctorat”
Titlul proiectului: Programe doctorale şi postdoctorale - suport pentru creşterea competitivităţii cercetării în domeniul Ştiinţelor exacte
Cod Contract: POSDRU/159/1.5/S/ 137750
Beneficiar- Universitatea „Alexandru Ioan Cuza” din Iași, Partener 2-Universitatea Babeș-Bolyai

Acknowledgment

I thank the doctoral supervisor, prof. Dr. Mircea V. Diudea for help and guidance given during my work for this objective.

I express my full consideration and gratitude to the members of the doctoral committee for the courtesy with which they agreed to be part of the doctoral committee and assessment of the thesis.

I would like to thank the advisory committee: Assoc. Prof. dr. Majdik Cornelia, Conf. univ. dr. KATONA Gabriel, Prof. univ. dr. ing. DĂRĂBANŢU Mircea.

I express special gratitude also my brother Aurel and my sisters Alexandra whose support, patience and understanding is reflected in each line of paper. Thanks also to colleagues where I held research and faculty colleagues for their presence, advice and support provided when necessary. I would like to express my gratitude to my parents for the support and assistance provided in these years.

This paper is a result of a doctoral research made possible by the financialsupport of the Sectoral Operational Programme for Human ResourcesDevelopment 2007-2013, co-financed by the European Social Fund, under theproject POSDRU/159/1.5/S/137750 - “Doctoral and postoctoral programs -support for increasing research competitiveness in the field of exact Sciences”.

INTRODUCTION

Thesis has 128 pages, grouped in two main chapters. The first part is presented QSAR algorithm. The aim of this thesis is to achieve theoretical studies on molecules containing nitrogen atoms substituted studies based on graph theory to achieve advanced models for predicting physicochemical and biological activities. The introduction is followed by Chapter I, "QSAR algorithm Diudea", focusing on QSAR methods in order to achieve connections between the molecular structure of substances and their properties in order to propose new molecular structures still nesintetizate with potential biological activity. There are general information about: methods for prediction of physico-chemical and biological matrices and using topological indices; Statistical elements, ie regression models and advanced models in QSAR; similarity of molecular structures and molecular docking.

The object of the thesis is to correlate biological activity log P and IC50 LD50 derivatives: flavonoids caffeine, serotonin (1-methyl-piperidin-4-yl) propane -indolizine and cloropropaminei with descriptors / global indicators / local. This correlation - QSAR has been performed in order to obtain information on how molecules interact with active sites when using different proteins, namely dividing the dataset into school set and test set.

2. PERSONAL CONTRIBUTIONS

2.1. QSAR study of derivatives of flavonoids containing a nitrogen atom

A set of 40 compounds of the flavonoid has been downloaded from the database PubChem [1, 2] (Table 1), together with log P. Set was divided into school set and the test set (25 15 molecules). The structures were optimized with HyperChem program at the level of theory MM + molecular mechanics and semi-empirical PM3 [3].

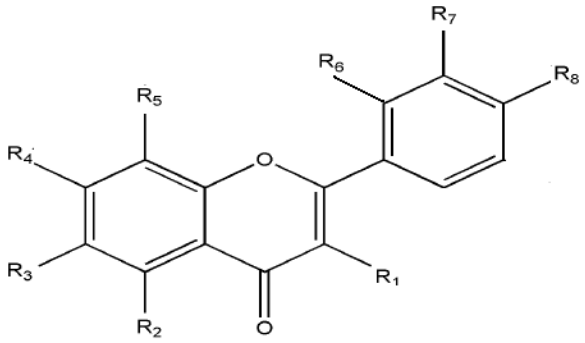

Table 1. The set of flavonoids, downloaded from the database PubChem [4].

	Substituent	log P		Substituent	log P

1	R_2-NH_2	3.4	2	R_1-H_2, R_3-OH, R_8-NH_2	2.9
3	R_1-H_2, R_8-NH_2	2.5	4	R_2-NH_2, R_6-NH_2	2.4
5	R_2-NH_2, R_8-$N(Me)_2$	2.2	6	R_2-NH_2, R_8-$N(Et)_2$	3
7	R_3-NH_2	3.3	8	R_2-NH_2, R_8-NH(Et)	2.5
9	R_4-NH_2	3.3	10	R_1-NH_2	3.1
11	R_1-NH_2, R_3-NH_2	2.4	12	R_2-OH, R_5-NH_2	2.7
13	R_3-NH_2, R_4-OH, R_5-NH_2	1.5	14	R_1-NH_2, R_6-NH_2	2.4
15	R_6-NH_2, R_7-Me	3.2	16	R_2-OH, R_4-OH, R_7-NH_2 , R_7-OMe	1.4
17	R_2-OH, R_3-NH_2, R_4-OH	1.4	18	R_2-OH, R_5-NH_2, R_8-NH_2	1.1
19	R_2-OH, R_4-OH, R_7-NH_2 , R_8-OH	1.1	20	R_4-OH, R_5-NH_2 , R_8-NH_2	1.8
21	R_3-OH, R_8-NH_2	2.9	22	R_3-NH_2, R_8-$N(Me)_2$	3
23	R_3-OMe, R_8-NH_2	2.8	24	R_7-OMe, R_8-NH_2	2.8
25	R_3-Me, R_4-Me, R_8-$N(Me)_2$	4.4	26	R_2-NH_2, R_8-OH	1.8
27	R_7-NH_2, R_8-NH_2	2.2	28	R_2-NH_2, R_8-NH(Me)	2.1
29	R_2-NH_2, R_6-OH, R_8-OH	2.3	30	R_3-Me, R_8-$N(Me)_2$	4
31	R_2-NH_2, R_6-$N(Me)_2$	3.2	32	R_2-NH_2, R_6-NH(Me)	3
33	R_2-NH_2, R_6-OH	2.7	34	R_2-NH_2, R_6-OMe	3
35	R_2-NH_2, R_5-OH, R_6-NH_2	2	36	R_2-NH(Me), R_6-$N(Me)_2$	3.8
37	R_2-NH(Me), R_8-$N(Me)_2$	2.9	38	R_2-NH_2, R_8-NH_2	1.4
39	R_2-NH_2, R_7-OH	2.7	40	R_2-NH_2, R_6-OH, R_8-NH_2	2

B. Model validation

(a) External validation

Values log P of flavonoids test set (Table 1, last 15 structures) were calculated using the equation

in Table 5, entry 17. Data is shown in Figure 3, n=15; R^2=0.768; s=0.366; F=42.956 [5].

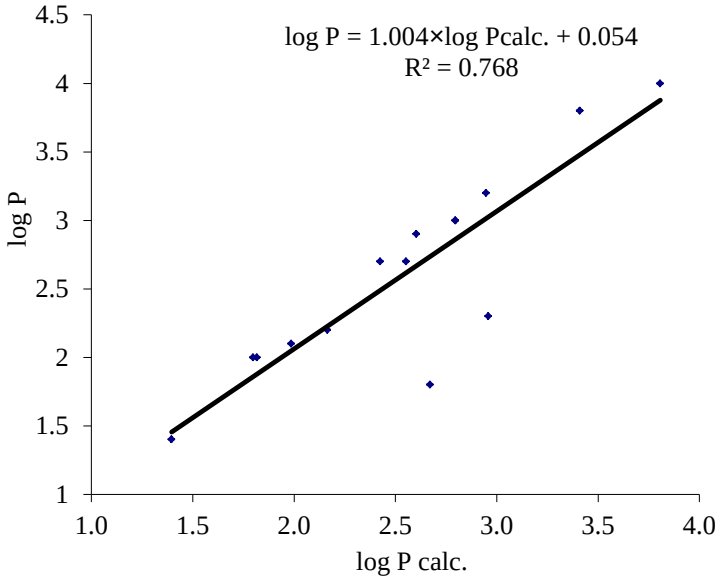

Figure 3. log P vs. log P calc. (External validation)

(b) Validation by cluster similarity

Validation can be done by calculating log P molecules in the set test equations learning clusters of similarity: each of the 15 molecule is leading his own band, chosen similarity (2D) [6] 25 structures school set. Values Log P calc. for each of the 15 molecules in the test set have been calculated by the equation 15 we (the leader being omitted) with the same descriptors as in equation 17, Table 5. The data are presented in Figure 4, n = 15; R2 = 0.951; s = 0.168; F = 252.005.

It may be noted that the log P prediction on clusters of similarity is far better than under the external validation (R^2 = 0,951 vs 0,950; s = 0,168 vs 0,195 și F = 252.005 vs 96.555).

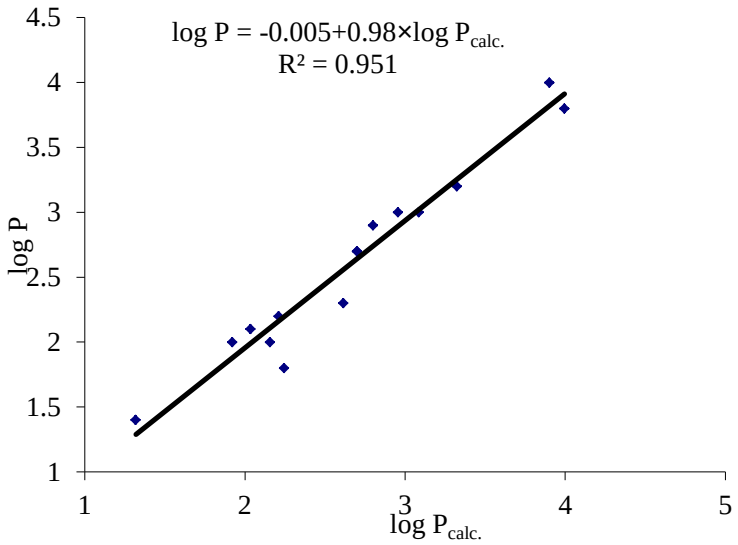

Figura 4. log P vs. log P calc. (prin clusteri de similaritate)

Explicația acestor rezultate excepționale este faptul că, prin grupuri de similaritate, se obține un set de cvasi-congeneri, făcând astfel posibilă paradigma de bază a QSAR: structurile similare au proprietăți similare. Numărul structurilor din cluster pot fi variată pentru a obține o mai bună estimare în cadrul fiecărui cluster și astfel, o mai bună predicție. Aceasta reprezintă o nouă procedură de corelarea, numită " predicție directă".

2.2. QSAR study of caffeine derivatives

A set of 40 structures belonging to the class of derivatives of caffeine [10.7] have been downloaded from the database PubChem (Table 6), together with log P values of LD50.

Table 6. Caffeine derivatives (SMILES cod) and log P (from PubChem)

	Canonical SMILES	log P	LD50
	Training Set		
1	CN1C=NC2=C1C(=O)N(C(=O)N2C)C	-0.1	127
2	CCCCC1=NC2=C(N1)C(=O)N(C(=O)N2CC(C)C)C	3.6	340
3	CC1=NC2=C(N1)C(=O)N(C(=O)N2CC(C)C)C	2.3	25
4	CC=CC1=NC2=C(N1C)C(=O)N(C(=O)N2C)C	1	100
5	CN1C=NC2=C1C(=O)N(C(=O)N2C)CC=C	0.6	191
7	CC=CCN1C(=O)C2=C(N=CN2C)N(C1=O)C	0.8	667
8	CCCN1C(=O)C2=C(N=CN2C)N(C1=O)C	0.8	126
10	CN1C=NC2=C1C(=O)N(C(=O)N2C)CC(CO)O	-1.8	1954
13	CC(CN1C(=O)C2=C(N=CN2C)N(C1=O)C)O	-0.3	580
14	C1=NC2=C(N1)C(=O)NC(=O)N2	-0.7	500
16	CN1C2=C(C(=O)NC1=O)NC=N2	-0.7	894

18	CCCN1C(=O)C2=C(N=CN2CCCCC(C)O)N(C1=O)C	1.2	1345
21	CN1C=NC2=C1C(=O)NC(=O)N2C	-0.8	837
22	CN1C2=C(C(=O)N(C1=O)C)N(C=N2)CCO	-1.2	400
23	CN1C2=C(C(=O)N(C1=O)C)NC=N2	-0.02	235
24	CCC1=NC2=C(N1)C(=O)N(C(=O)N2C)C	0.8	175
25	CCCCCCC1=NC2=C(N1)C(=O)N(C(=O)N2C)C	2.8	500
26	CCN(CC)CCN1C=NC2=C1C(=O)N(C(=O)N2C)C	0.3	1237
27	CCCC1=NC2=C(N1)C(=O)N(C(=O)N2C)C	1.2	250
28	CC(C)CC1=NC2=C(N1)C(=O)N(C(=O)N2C)C	1.6	250
33	CC(C)CN1C2=C(C(=O)N(C1=O)CC(C)C)NC=N2	2.8	796
35	CCCCN1C2=C(C(=O)N(C1=O)C)NC=N2	1.3	237
38	CCC1=NC2=C(N1)C(=O)N(C(=O)N2CC(C)C)CC(C)C	4	322
39	CC1(N=C2C(=N1)N(C(=O)N(C2=O)C)C)C	-0.2	265
40	CCCCC1=NC2=C(N1)C(=O)N(C(=O)N2C)C	1.7	250
	Set Test		
6	CCN1C(=O)C2=C(N=CN2C)N(C1=O)C	0.3	61
9	CCOC1=NC2=C(N1C)C(=O)N(C(=O)N2C)C	0.6	56
11	CN1C=NC2=C1C(=O)N(C(=O)N2C)CC(CO)O	-1.4	1920
12	CC(C)CN1C2=C(C(=O)N(C1=O)C)N(C=N2)CC(CO)O	-0.5	784
15	C1=NC2=C(N1)C(=O)NC(=O)N2O	-1.2	100
17	C1=NC2=C(N1)C(=O)NC(=O)N2CCO	-1.2	490
19	CCCCN1C2=C(C(=O)N(C1=O)CCCC)N(C=N2)CC(=O)C	1.9	1000
20	CC1=NC2=C(N1)C(=O)N(C(=O)N2C)C	0.4	130
29	CN1C(=O)C2=C(NC1=O)N=CN2	-0.3	510
30	CCCN1C2=C(C(=O)N(C1=O)C)NC=N2	1	79

B. Validation of the model (for log P mass fragments)

(a) external validation

Value log P of the caffeine test set (Table 1) was calculated using the best equation of Table 9, entry 17. The data is shown in Figure 6,n=15; R^2= 0.805; s=0.476; F=53.515.

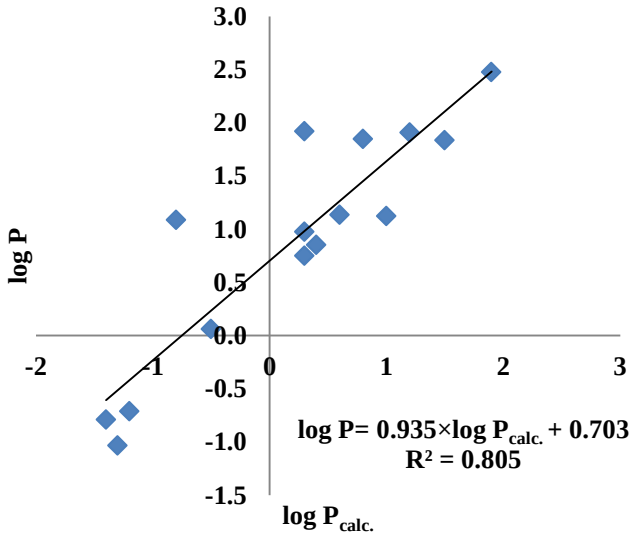

Figure 6. log P vs. log P $_{calc.}$ (external validation)

(b) Validation by clusters of similarity

Values Log P calc. for each of the 15 molecules in the test set have been calculated by the equation 15 we (the leader being omitted) with the same descriptors as in equation 17, tab.9. The data are presented in Figure 7, n = 15; R2 = 0.974; s = 0.173; F = 488.584.

Also, the value of log P calc. to 15 molecules in the test set was designed with the same descriptors as in equation 10 Table 9. The data are presented in Figure 8, n=15; R^2=0.980; s=0.153; F=629.628.

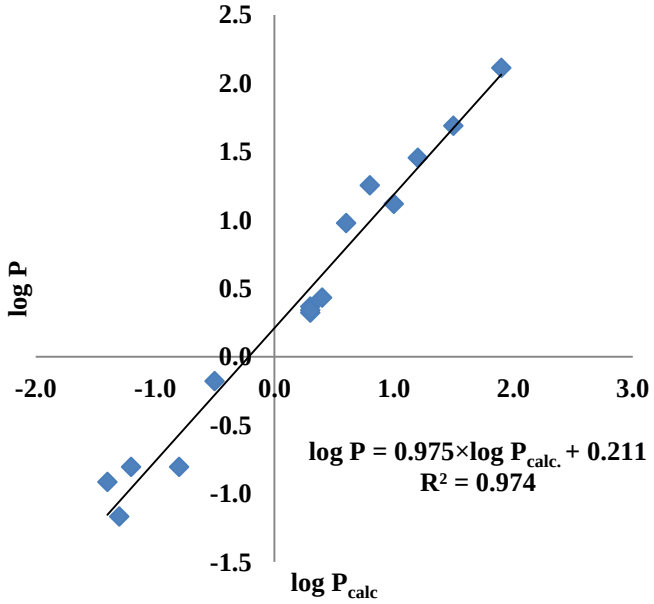

Figura 7. log P vs. log P $_{calc.}$ prin clusteri de similaritate

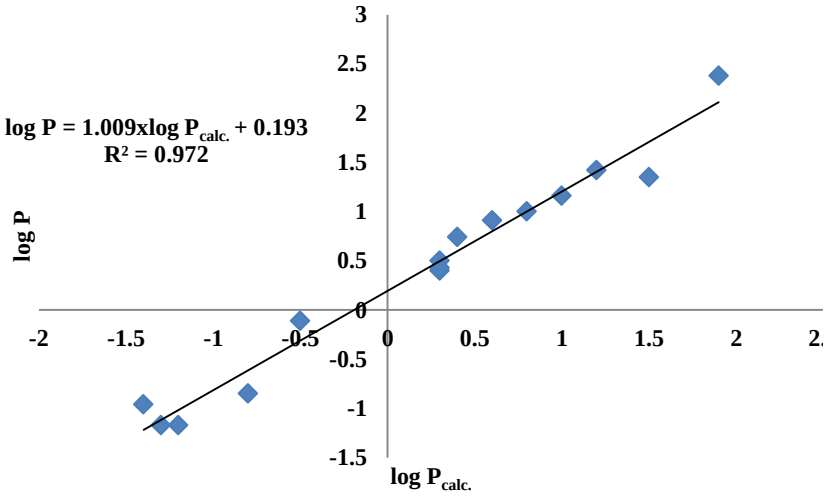

Figure 8. log P vs. log P $_{calc.}$ by similarity clusters

B. Model validation (for LD50)

(a) External validation

LD50 values for the set of caffeine test (Table 6, the last 15 structures) was calculated using the best equation of Table 10, entry 17. The data is shown in Figure 9, n=15; R^2=0.761; s=283.996; F=41.43.

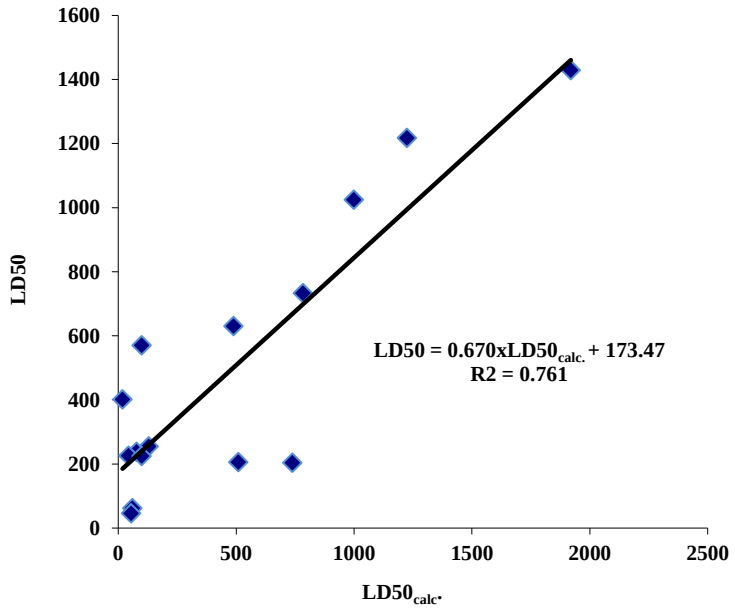

Figure 9. LD50 vs. LD50$_{calc.}$ (external validation)

(b) Validation by clusters of similarity

Validation was performed by calculating the LD50 test molecules from the set value LD50calc. was calculated by the same equation as in the descriptor 11 Table 10. Data are presented in Figure 10,n=15; R^2=0.893; s=190.43; F= 108.06.

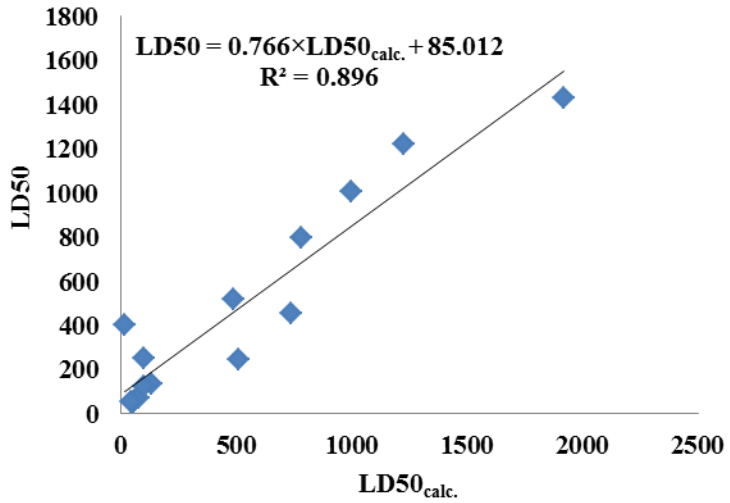

Figure 10. LD50 vs. $LD50_{calc.}$ by similarity clusters

Partial charges (case 2)

B. Model validation (for log P)

(a) external Validation

Log P values of caffeine test set (Table 1), were calculated using the best equation in Table 11entry 17. The data is shown in Figure 11,n=15; R^2=0.742; s=0.547; F=37.302.

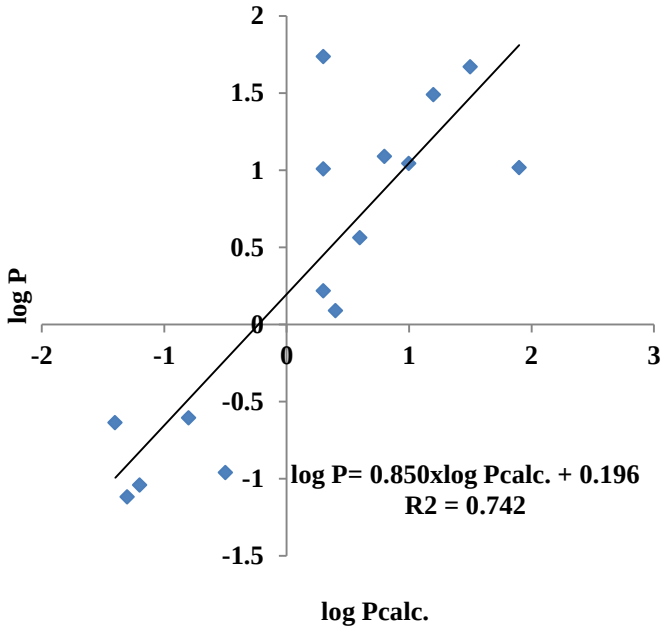

Figure 11. log P vs. log P calc. (external validation)

(a) Validation by clusters of similarity

Validation was done by calculating the log P of the set test molecules, Pcalc log values. Those were calculated with the equation 17, Table 11. Data are presented in Figure 12,n=15; R^2=0.963; s=0.208; F= 334.794.

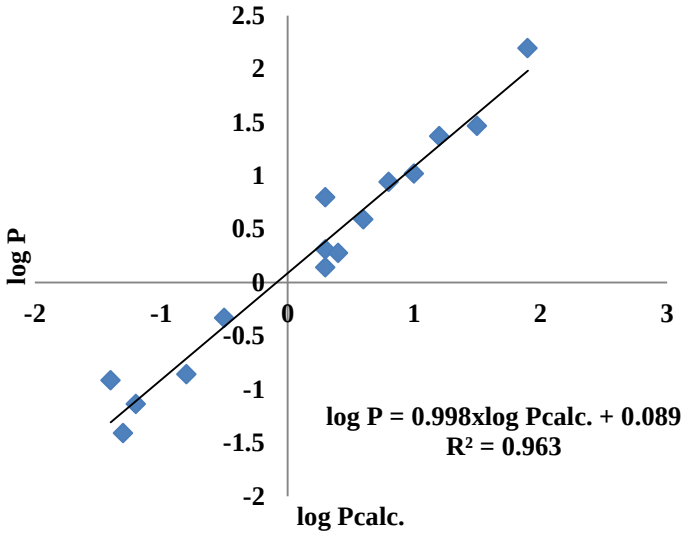

Figure 12. log P vs. log $P_{calc.}$ by similarity clusters

(b) Similarity Cluster Validation

Values log P were calculated with the same entry 10, Table 11. The data are shown in Figure 13, n=15; R^2=0.969; s=0.191; F= 401.555.

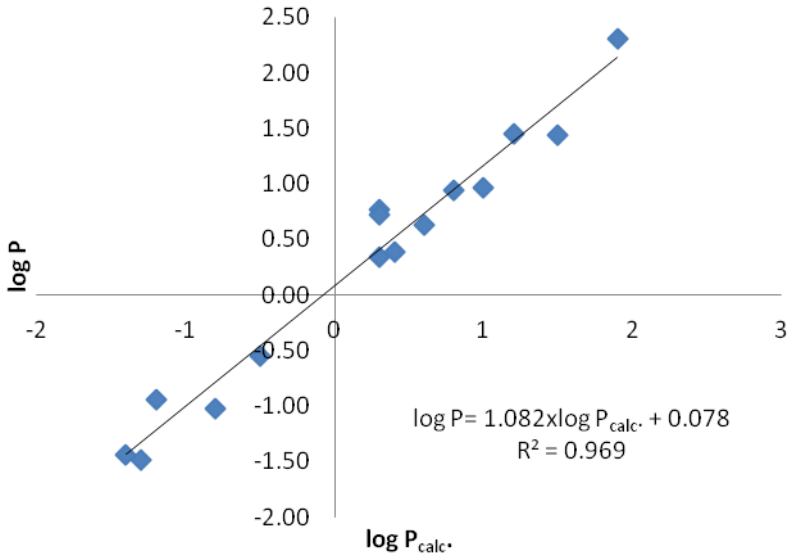

Figure 14. log P vs. log $P_{calc.}$ by similarity clusters

2.3. QSAR Study on indole

A set of derivatives of indole 40 [11-15] have been downloaded from the database PubChem (Table 12) and was divided into school set (25 molecules) and test set (15 molecules), chosen randomly. Modeling of properties was selected for log P and LD50 (rat, intraperitoneal), see Table 12.

Table 12. Indols (cod SMILES) and log P, LD50

Nr. Crt.	**Canonical SMILES**	Log P	LD50
1	C1=CC=C2C(=C1)C(=CN2)CCN	1.6	100
2	CNCCC1=CNC2=CC=CC=C21	2.1	158
3	CC1=CC2=C(C=C1)NC=C2CCN	1.9	50
4	C1=CC=C2C(=C1)C(=CN2)CC(C(=O)O)N	-1.1	4800
5	C1=CC=C2C(=C1)C(=CN2)CC(=O)O	1.4	150
6	C1=CC=C2C(=C1)C(=CN2)C=O	1.7	600
7	C1=CC2=C(C=C1O)C(=CN2)CCN	0.2	160
8	CC(=O)C1=CNC2=CC=CC=C21	2.1	300
9	CC(=O)OC1=CNC2=CC=CC=C21	2	600
10	C1=CC=C2C(=C1)C(=CN2)CCO	1.8	351
11	CN(C)CCC1=CNC2=C1C=C(C=C2)O	1.2	290
12	C1=CC=C2C(=C1)C=CN2	2.1	117

13	CN(C)CCC1=CNC2=C1C(=CC=C2)O	2.1	196
14	CN1C(=O)CC2=CC=CC=C21	1	1177
15	CCC(=O)NCCC1=CNC2=CC=CC=C21	2.3	900
16	C1=CC=C2C(=C1)C(=CN2)CCC(=O)O	1.8	100
17	CCNCCC1=CNC2=CC=CC=C21	2.4	562
18	CCOC(=O)C1=CC2=C(N1)C=CC(=C2)OC	3.2	350
19	C1=CC=C2C(=C1)C=C(N2)O	2.4	400
20	CC(=O)C1=CC2=C(C=C1)NC=C2	2	450
21	C1=CC2=C(C=CN2)C=C1C(=O)CO	1.3	600
22	CC(=O)CC1(C2=CC=CC=C2NC1=O)O	0.2	800
23	CN(C)CCC(C1=CNC2=CC=CC=C21)O	1.8	767
24	COC1=CC2=C(C=C1)NC=C2CC(=O)O	1.4	98
25	C1=CC2=C(C=C1O)C(=CN2)CC(=O)O	1.1	1125
26	C1=CC2=C(C=CN2)C=C1O	2	1000
27	COC1=CC2=C(C=C1)NC=C2	2.1	370
28	CC(CC1=CNC2=CC=CC=C21)N	2	20
29	CCC(CC1=CNC2=CC=CC=C21)N	2.5	400
30	CN(C)CC1=CNC2=CC=CC=C21	1.8	122
31	C1=CC2=C(C=C1O)C(=CN2)CC(=O)O	1.1	1125
32	C1=CC=C2C(=C1)C(=CN2)C(=O)CO	1.4	700
33	CCC(=O)NCCC1=CNC2=C1C=C(C=C2)OC	1.3	850
34	CC1=C(C2=C(N1)C=CC(=C2)OC)CCN(C)C	2.7	100
35	COC1=CC2=C(C=C1)NC=C2CCN	0.5	176
36	CC(=O)C1=CNC2=C1C(=O)CCC2	0.5	233
37	CC1=C(NC2=C1C(=O)CCC2)C(=O)C	1.2	533
38	CC1=CC2=C(C=C1)NC=C2C(=O)CO	1.8	600
39	COC1=CC2=C(C=C1)NC=C2C(=O)CO	1.1	600
40	COC1=CC2=C(C=C1)C(=CN2)C(=O)CO	1.1	600

Mass fragment (case 1)

Model validation (for log P)

(a) Leave-one-out

Leave-one-out analysis on the best models in Table 18 are presented in Table 19.

Table 19. Leave-one-out for LD50.

	Descriptors	Q^2	R^2-Q^2	St, $Error_{loo}$	F_{loo}
1	**SD_3**	**0.812**	0.04	98.356	77.926
5	**SD_3, De**	**0.802**	0.052	100.923	73.108
11	**SD_3, CfDi, CjDe**	**0.794**	0.065	103.057	69.374
17	**SD_3, De, CjDi, CfDi**	**0.805**	0.095	100.373	74.108

(b) External validation

LD50 values for the test set (Table 12, 10 structures) were calculated using the best equation in Table 18entry 11. The data is shown in Figure 18,n=10; R^2=0.869; s=72.596; F= 78.415.

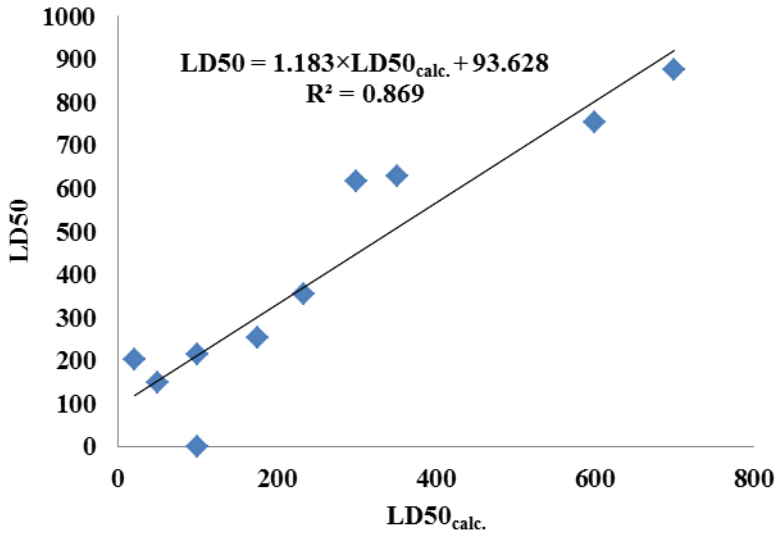

Figure 18. LD50 vs. $LD50_{calc.}$

(c) Validation by clusters similarity

Validation was done by calculating the LD50 for the test molecule from the set values LD50calc. Descriptors were calculated by the same equation as 11 Table 18. Data are presented in Figure 19, n=10; R^2=0.909; s=73.729; F=80.241 [87].

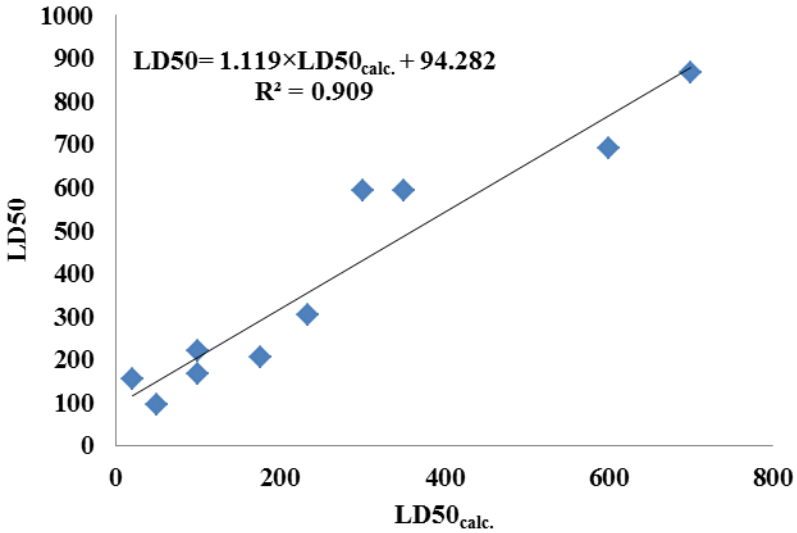

Figure 19. LD50 vs. $LD50_{calc.}$

Partial charges (case 2)

B. Model Validation (for log P)

(a) Leave-one-out

Analiza leave-one-out pentru cele mai bune modele din tabelul 20 sunt prezentate în tabelul 21.

Table 21. Leave-one-out for log P.

	Descriptors	Q^2	R^2-Q^2	St. Error $_{loo}$	F_{loo}
1	**SD_2**	**0.85**	0.038	0.256	130.397
5	**SD_2, HOMO**	**0.879**	0.029	0.23	168.056
10	**SD_2, HOMO, De**	**0.884**	0.032	0.225	175.399
16	**SD_2, HOMO, De, Adj**	**0.878**	0.041	0.232	165.098

(b) External validation

Log P values for the test set (Table 12) was calculated using the best equation from Table 20entry 10. The data is presented in Figure 20, n=15; R^2=0.736; s=0.403; F=30.605.

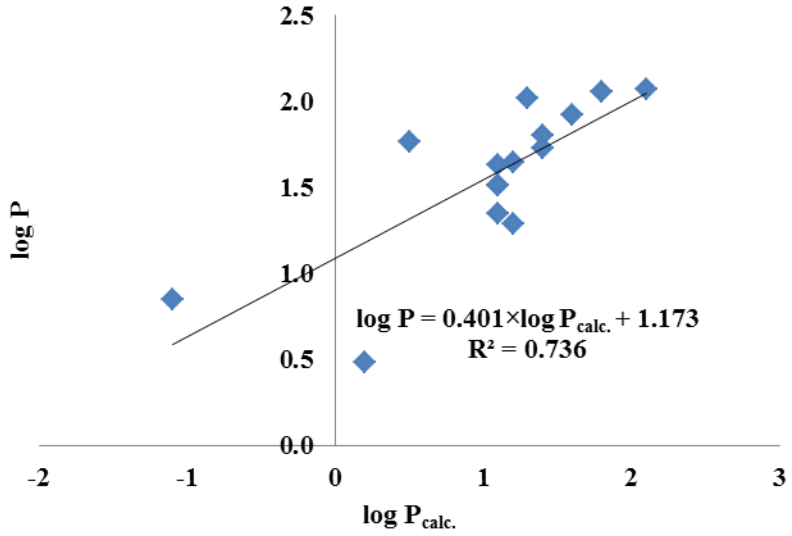

Figure 20. log P vs. log P $_{calc.}$

(c) Similarity Cluster Validation

Validation was done by calculating the log P of the set test molecules, log Pcalc values. Descriptors were calculated by the same equation as 10 Table 20. Data are presented in Figure 21, n=15; R^2=0.922; s=0.217; F=155.571.

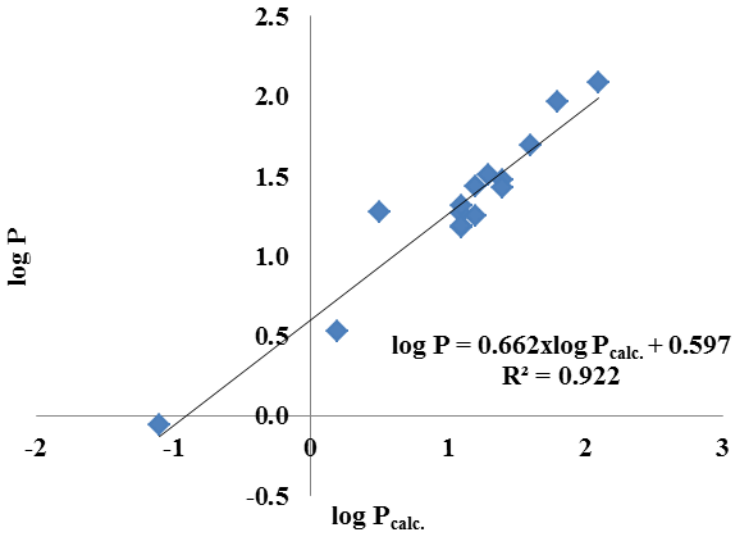

Figure 21. log P vs. log $P_{calc.}$

by similarity clusters

B. Model validation (for LD50)

(a) Leave-one-out

Leave-one-out analysis [16] the best models of Table 22 are presented in Table 23.

Table 23. Leave-one-out for LD50.

	Descriptors	Q^2	R^2-Q^2	St. $Error_{loo}$	F_{loo}
1	**SD_4**	**0.868**	0.028	80.594	118.112
5	**SD_4, HOMO**	**0.865**	0.031	81.381	115.492
10	**SD_4, Di, De**	**0.864**	0.041	81.522	115.033
17	**SD_4, De, CjDi, CfDi**	**0.821**	0.1	93.722	82.653

(b) External Validation

The values LD50 for the test set of serotonin were calculated by using equation 11 in Table 22. Data is plotted in Figure 22, n=10; R^2=0.811; s=143.121; F=29.989.

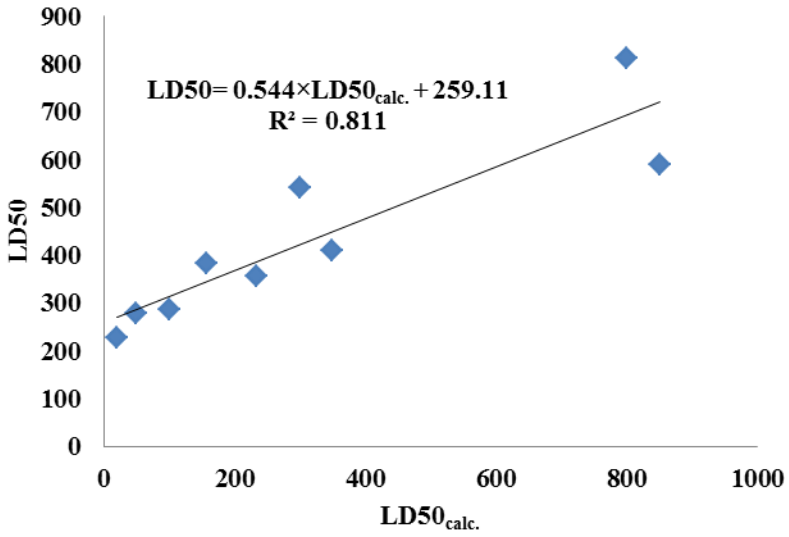

Figure 22. LD50 vs. $LD50_{calc.}$

(c) Similarity Cluster Validation

Validation was done by calculating the LD50 for the test molecule from the set values LD50calc. descriptors were calculated by the same equation as 10 Table 22. Data are presented in Figure 23, n=10; R^2=0.934; s=82.445; F= 114.006.

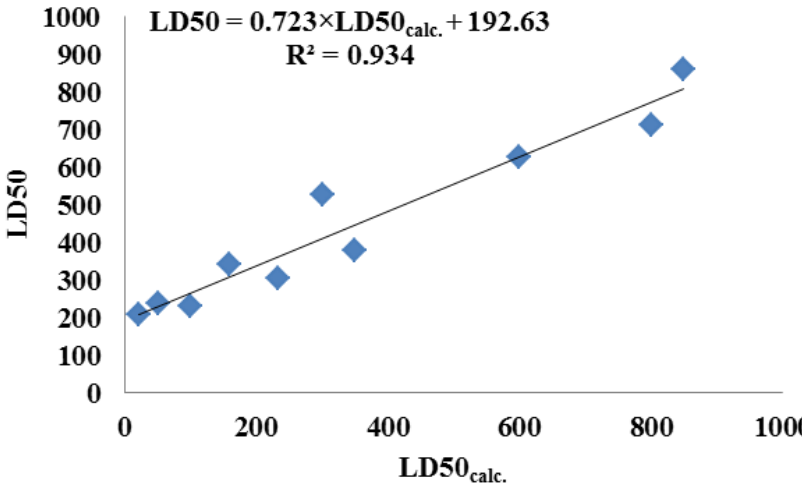

Figure 23. LD50 vs. $LD50_{calc}$ by similarity clusters

2.4. QSAR study on derivatives (1-methylpiperidin-4-yl) propanoate

A set of 40 molecules structures derivatives (1-methylpiperidin-4-yl) propanoate [17 to 20] have a valuable Pubchem downloaded from the database (Table 23), together with log P set was divided into set school and test set (25 and 15 molecules randomly chosen). The structures were optimized at Hartree-Fock level of theory HF (6-31G (d, p)). Calculations were performed with Gaussian 09 program.

Table 23(1-methylpiperidin-4-yl) propanoate (cod SMILES) and log P PubChem).

Mol.	Canonical SMILES	log P	CID
1	CCC(=O)OC1CCN(CC1)[11CH3]	1.2	6540307
2	CC(C)(C)C(=O)OC1CCN(CC1)C(C)(C)C	3.1	57613500
3	CC(C)N1CCC(CC1)OC(=O)C(C)(C)C	2.9	58873543
4	CC(CC(C)(C)N1CCCCC1)OC(=O)C(C)(C)C	3.8	40500608
5	CCC(=O)OC1CCN(CC1)CC	1.5	24843023
6	CCC(=O)OC1CCN(CC1)C	1.2	133349
7	CCOC1CCN(CC1)CCC(=O)OCC	1.2	61220923
8	CCN1CCC(CC1)OC(=O)C	1.1	11480704
9	CCOC1CCN(CC1)CCC(=O)OC	0.9	61218612
10	CCCC(C)OC(=O)CCN1CCCCC1	2.6	71028323
11	CCC(C)(C)C(=O)OC(C)CN1CCCCC1	3.2	58545331
12	CCC(C)(C)C(=O)NCCN1CCCCC1	2.2	58795427
13	CC(C)(C)C(C(=O)N1CCCCC1)NC	1.9	58700497
14	CC(C)(C)C(CC1=CC=CC=C1)N(C)C	4	58172339
15	CC(C)(C)C(=O)C(CC1=CC=CC=C1)N(C)C	3.4	3637447
16	CCNC(=O)N1CCC(CC1)C(=O)NC(C)C	0.5	53548901
17	CC(C)(C)C1=CC=C(C=C1)C(=O)NC(C)(C)C	4.6	349124
18	CC(=O)NC1=CC=C(C=C1)C(=O)NC(C)(C)C	2	151118
19	CC1=CC(=C(C=C1)C)C(=O)NC(C)(C)C	2.9	925429
20	CC(C)(C)OC(=O)N1CCCC(C1)C=O	1.1	42325667
21	CC1CCN(CC1C=O)C(=O)OC(C)(C)C	1.5	58010030
22	CC(C)(C)NC(=O)N1CCC(CC1)C(=O)N	-0.1	894347
23	CCC(C)(C)NC(=O)C1CCN(CC1)C(=O)N(C)C	1	47205727
24	CC(C)(C)NC(=O)C1CCN(CC1)C(=O)N(C)C	0.5	60779224
25	CCCC(=O)NCC1CCN(CC1)C(=O)NC(C)(C)C	1.4	49687908
26	CC(=O)C1CCN(CC1)C(=O)NC(C)(C)C	0.7	58171886
27	CC(C)CC(=O)NCC1CCN(CC1)C(=O)NC(C)(C)C	1.9	49687914
28	CC(C)C(=O)NCC1CCN(CC1)C(=O)NC(C)(C)C	1.6	49687909
29	CCC(=O)NCC1CCN(CC1)C(=O)NC(C)(C)C	1.1	49687905
30	CC(C)(C)CC(C)(C)NC(=O)N1CCC(CC1)C(=O)N	1.6	24159137
31	CCC(=O)OC1CC[N]CC1	0.7	57426704
32	COC1CCN(CC1)CCC(=O)OC	0.5	43216573
33	CC(C)(C)CC(=O)NCC1CCN(CC1)C(=O)NC(C)(C)C	2.3	49687921
34	CC(C)CC(C(C)(C)C)NC(=O)C1CCN(CC1)C(=O)N(C)C	3	56793859
35	CC(C)C(=O)N1CCC(CC1)C(=O)NC(C)(C)C	1.5	60726650
36	CCCC(=O)N1CCC(CC1)C(=O)NC(C)(C)C	1.3	45596615
37	CCC(C)(C)NC(=O)C1CCN(CC1)C(=O)C	1	39959127
38	CC(=O)N1CCC(CC1)C(=O)NC(C)(C)C	0.5	17148671
39	CC(C)(C)C(=O)NCC1CCN(CC1)C(=O)NC(C)(C)C	2	49687916
40	CC(=O)CCN1CCC(CC1)OC(=O)C	0.4	58811219

Mass fragment (case 1)

B. Model validation (for log P)

(a) Leave-one-out

The performances in leave-one-out analysis related to the models listed as best in Table 24 are presented in Table 25.

Table 25. LOO for log P

	Descriptors	Q^2	R^2-Q^2	St. Error$_{loo}$	F_{loo}
1	**SD_1**	**0.919**	0.011	0.334	260.613
5	**SD_1, D3D**	**0.913**	0.017	0.347	240.626
11	**SD_1, De, CjDi**	**0.916**	0.028	0.34	250.326
19	**SD_1, HOMO, Adj., C**	**0.922**	0.03	0.328	272.132

(b) External validation

Values log P for the test set derivatives (1-methylpiperidin-4-yl) propanoate was calculated using the equation cf. entry 11, Table 24. Data are presented in Figure 25, R^2=0.869, s=0.331, F=86.123.

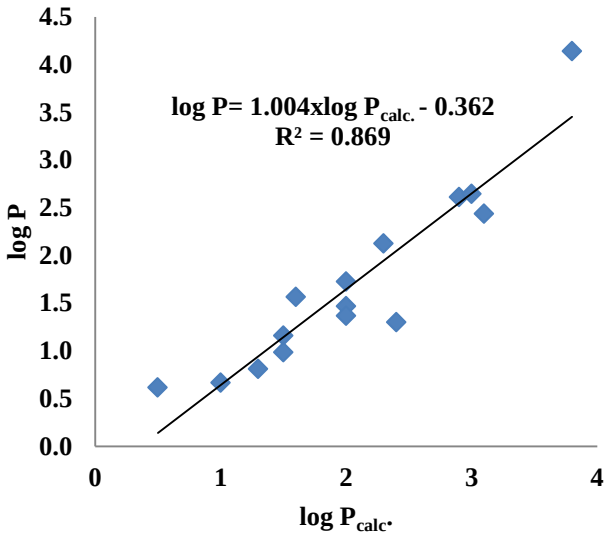

Figure 25. log P vs. log P_{calc}.

(external validation)

(c) Similarity Cluster Validation

The values LD50 for the test set of serotonin were calculated by using equation 11 in Table 24. Data is plotted in Figure 26, R^2=0.953, s=0.197, F=265.759.

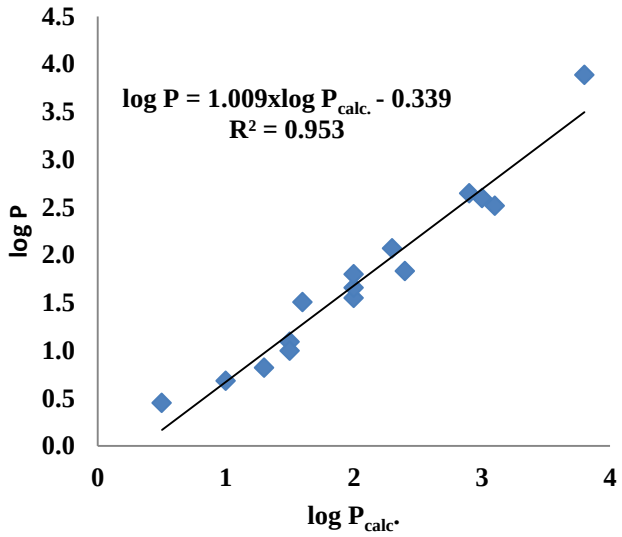

Figure 26. log P vs. log P calc. (by similarity clusters)

Partial charges (case 2)

B. Model validation (for log P)

(a) Leave-one-out

The performances in leave-one-out analysis related to the models listed as best in Table 26 are presented in Table 27.

Table 27. LOO for log P

	Descriptors	Q^2	R^2-Q^2	**St. Error$_{loo}$**	F_{loo}
1	SD_1	0.9	0.017	0.396	206.94
5	SD_1, HOMO	0.884	0.034	0.426	175.824
11	SD_1, HOMO, De	0.881	0.038	0.432	170.297
19	SD_1, HOMO, Adj., C	0.868	0.052	0.455	171.485

(b) External Validation

Value log P for the test set of the compounds (1-methylpiperidin-4-yl) propanoate was calculated using the equation 10 Table 26. Data are presented in Figure 27, R^2=0.833, s=0.286, F=64.662.

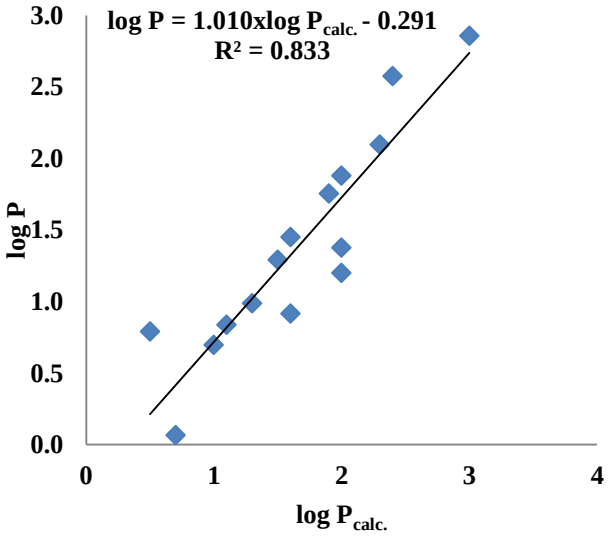

Figure 27. log P vs. log P calc. (external validation)

(c) Similarity Cluster Validation

Value log P for the test set of the compounds (1-methylpiperidin-4-yl) propanoate was calculated using the equation 10 Table 26. Data are presented in Figure 28, R^2=0.945, s=0.164, F=222.647.

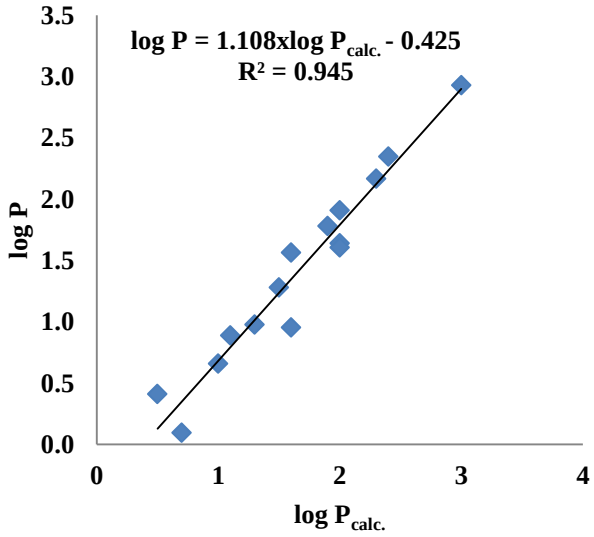

Figure 28. log P vs. log P $_{calc.}$ (by similarity clusters)

2.5. QSAR and Docking Study on Dopamine

Dopamine is one of the main neurotransmitters, a substance that transmit information between brain neurons [21]. It plays a key role in the central nervous system in the body and its deficiency can lead to neurological disorders such as Parkinsonism [22-26].

A set of 40 dopamine have been downloaded from the database PubChem (Table 28) and is divided into two sets: set school (25 molecules) and set test (15 molecules italic) taken energy docking lowest (see lower). The property is chosen for modeling LD50 (mouse, intraperitonial).

Table 28. Dopamine, LD50 (cod Smiles, from PubChem) and topological indices

Mol.	Canonical SMILES	LD50	SD	CjDe	CfDi	Di
1	C1=CC(=C(C=C1CCN)O)O	950	418.549	139	226	160
2	*C1=CC(=C(C=C1C(CN)O)O)O*	*6*	*828.993*	*174*	*277*	*198*
3	*C1=CC=C(C=C1)CC(C(=O)O)N*	*1322*	*1187.315*	*197*	*269*	*94*
4	C1=CC=C(C=C1)CCN	175	435.814	82	127	133
5	CNCCC1=CC=CC=C1	190	348.992	120	174	321
6	*C1=CC(=CC=C1CC(C(=O)O)N)O*	*1450*	*1479.283*	*239*	*345*	*211*
7	CNCCC1=CC=C(C=C1)O	780	549.527	151	231	256
8	*C1=CC(=C(C=C1CC(C(=O)O)N)O)O*	*588*	*970.196*	*291*	*426*	*126*
9	CNCC(C1=CC=C(C=C1)O)O	1000	773.346	185	278	164
10	*CNCC(C1=CC(=C(C=C1)O)O)O*	*4*	*390.303*	*229*	*348*	*170*
11	C1=CC(=CC=C1C(CN)O)O	600	691.229	137	217	126
12	CC(CC1=CC=CC=C1)N	4.4	304.705	113	167	168
13	COC1=CC=CC(=C1)CCN	230	367.010	148	230	160
14	COC1=CC=C(C=C1)CCN	100	237.751	143	231	206
15	CC(CC1=CC=CC=C1)N	5.5	292.395	113	167	162
16	*CC(CC1=CC=CC=C1)NC*	*7*	*495.254*	*154*	*217*	*114*
17	CC(C)(CC1=CC=CC=C1)N	71	175.741	146	209	88
18	C1=CC(=C(C=C1O)O)O	25	-36.423	69	124	332
19	*CC(CC1=CC(=C(C=C1)O)O)(C(=O)O)N*	*150*	*458.007*	*341*	*492*	*121*
20	*CC(C)NCC(C1=CC(=C(C=C1)O)O)O*	*440*	*497.723*	*367*	*518*	*162*
21	CNCC(C1=CC(=CC=C1)O)O	60	233.726	190	278	270
22	C1=CC(=CC=C1CCN)O	800	590.743	107	174	223
23	CC(CC1=CC=CC=C1)NC	82	204.239	154	217	166
24	*CC(C)(CC1=CC=CC=C1)NC*	*89*	*72.473*	*190*	*262*	*156*
25	*CC(=O)NC1=CC(=CC=C1)O*	*1025*	*785.809*	*147*	*223*	*126*
26	C1=CC(=CC=C1N)N	50	319.911	48	89	212
27	*C1=CC=C(C(=C1)C(=O)N)O*	*180*	*304.477*	*104*	*164*	*268*
28	C1=CC=C(C=C1)C(=O)N	1282	120.697	76	121	174
29	*CC(C)NCC(C1=CC(=CC=C1)O)O*	*320*	*341.982*	*314*	*426*	*160*
30	C1=CC(=C(C=C1C(CN)O)O)O	15.6	138.430	183	289	84
31	COC1=C(C=C(C=C1)CCN)OC	181	290.725	231	360	374
32	C1=CC(=CC=C1C(CN)O)O	600	691.211	137	217	400
33	C1=CN=CC=C1C(=O)NN	100	300.468	108	162	127
34	CN(C)C1=CC=C(C=C1)C=O	200	285.728	135	223	168
35	*CC(=O)NC1=CC=C(C=C1)O*	*367*	*909.929*	*143*	*223*	*205*
36	CN(C)CCC1=CC(=C(C=C1)O)O	240	297.644	243	362	62
37	*CN(C)CCC1=CC=C(C=C1)O*	*299*	*678.261*	*197*	*290*	*204*
38	CC(CC1=CC=C(C=C1)O)N	300	460.891	143	223	256
39	*C1=CC(=CC(=C1)O)C(CN)O*	*198*	*153.779*	*141*	*217*	*160*
40	CC(CC1=CC=CC=C1)N	5.5	304.693	113	167	166

B. . Model Validation

(a) Leave-one-out

The performances in leave-one-out analysis related to the models listed as best in Table 30 are presented in Table 31.

Table 31. Leave-one-out for LD50

	Descriptors	Q^2	R^2-Q^2	St. Error $_{loo}$	F_{loo}
1	**SD**	**0.812**	0.03	157.749	94.219
5	**SD, D3D**	**0.800**	0.056	162.036	88.151
10	**SD, CjDe, CfDi**	**0.794**	0.07	164.598	84.748

(b) External Validation

The values LD50 for the test set of 15 dopamines (Table 28) were calculated by using the best equation in Table 30, entry 10. Data is plotted in Figure 31, n=15; R^2=0.683; s=277.017; F=25.885.

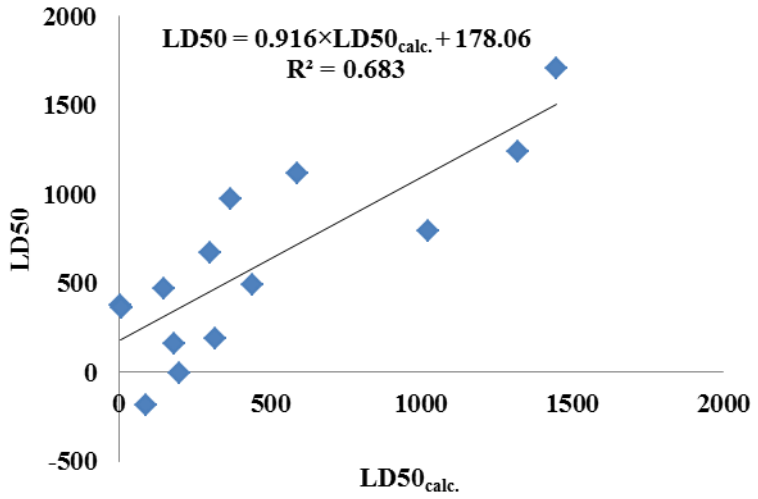

Figure 31. LD50 vs. $LD50_{calc.}$ by external validation

(c) Similarity Cluster Validation

The values LD50 for the test set of 15 dopamines (Table 28) were calculated by using the best equation in Table 30, entry 10. Data is plotted in Figure 32, n=15; R^2=0.905; s=150.862; F= 123.189.

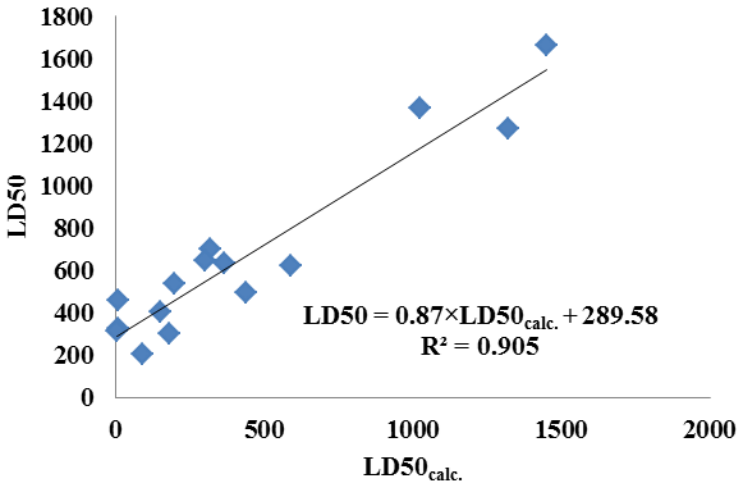

Figure 32. LD50 vs. $LD50_{calc.}$ similarity clusters

Compare the results in Figures 32 and 33 to see: (i) a rather low prediction (R^2=0.683) by the external test set and (ii) a better prediction (R^2=0.905) by the same set predicted by the similarity clusters (approaching to the congeneric status), in comparison with the model (R^2=0.864), even the test set has been chosen the one with the lowest docking energies. This result put our approach in a favorable light and demonstrates its utility in QSAR studies. Modeling the lethal dose LD50 of dopamines, by integrating their database information within the hypermolecule complex structural information, provided stable and performing QSAR models, excellent in prediction by similarity clustering. 1CJM (protein (aryl sulfotransferase) has been investigated for the binding affinity with the selected dopamine derivatives.

2.6. QSAR and Docking Study on Caffeine

In this study, clusters of similar structures (aimed to be quasi-congeneric subsets, in a better prediction of the toxicological activity) were chosen, with the leaders the best scored in the docking on the target protein cid1.

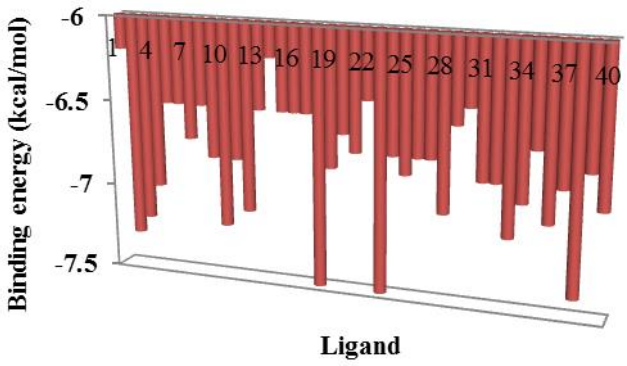

Figure 44: Binding energy

To obtain a pharmacophore model that fits at the receptor Poly(A) RNA polymerase protein cid1, conformers with the most favorable interactions with the receptor resulting from docking, were chosen. Ligands 2, 18, 23 and 38 have the lowest binding energy between -7.5 and -7.3. The resulting pharmacophore is shown in Figure 45.

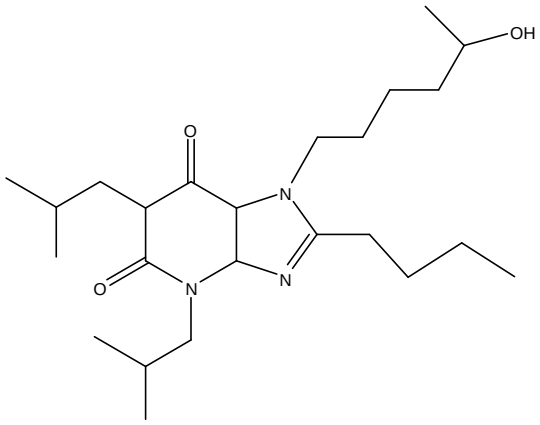

Figure 45(a): Pharmacophore model

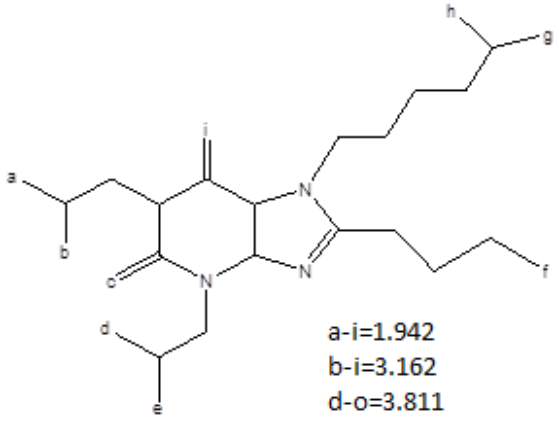

Figure 45(b): Distances within pharmacophore features (Å).

B. Model validation (for LD50)

(a) Leave-one-out

Analiza leave-one-out pentru cel mai bune modele din tabelul 46 sunt prezentate în tabelul 47.

Table 47. Leave-one-out for LD50

	Descriptors	**Q^2**	**R^2-Q^2**	**St. Error$_{loo}$**	**F$_{loo}$**
1	SD	0.873	0.018	165.428	157.876
5	SD, IE_{max}	0.913	0.023	134.202	251.841
11	SD, IE_{max}, IE_{min}	0.908	0.026	140.407	228.086

(b) External Validation

The values LD50 for the test set of caffeine were calculated by using equation 11 in Table 46. Data is plotted in Figure 47, n=15; R^2=0.929; s=153.272; F=169.735.

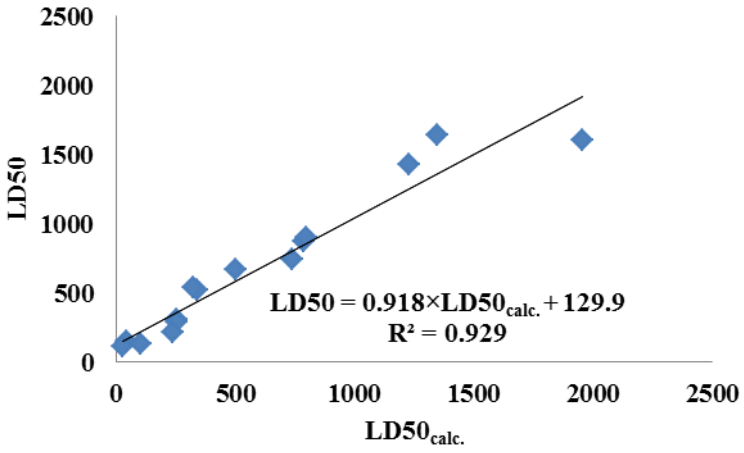

Figure 47. LD50 vs. LD50 $_{calc}$ for the test set (external validation)

(c) Similarity Cluster Validation

Validation can also be performed by using clusters of similarity: each of the 15 molecules in the test set (chosen as the best scored in the docking set) is the leader of its own cluster, selected by 2D similarity among the 20 structures of the learning set (each cluster comprising about 14-17 molecules). The values LD50 for the test set of caffeine were calculated by using the learning equations (with the same descriptors as in eq. 11, Table 46) from each of the 15 clusters. Data is plotted in Figure 48.n=15; R^2=0.951; s=127.328; F=251.832.

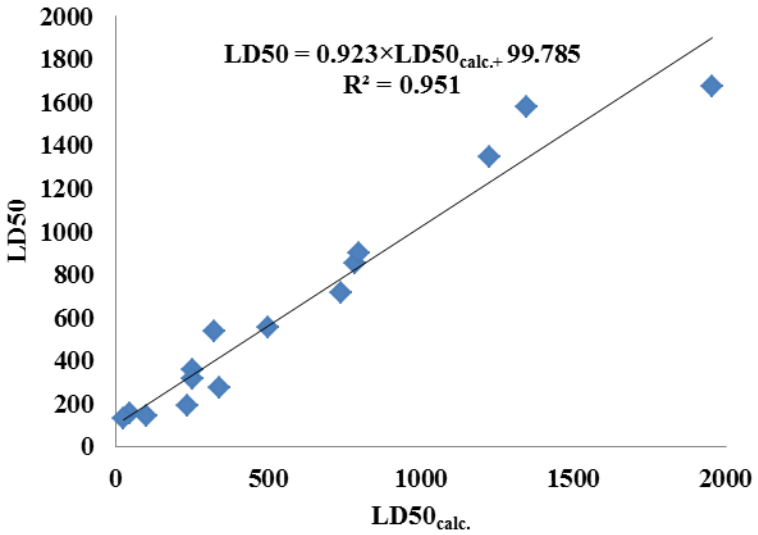

Figure 48. LD50 vs. LD50 $_{calc..}$ by similarity clusters

In this paper a qsar study on 40 caffeine derivatives, docked on the protein (4FH3), was reported. Molecular docking was performed to investigate the binding modalities of ligands toward possible targets comprised in poly (A) polymerase Cid1 (4FH3). A further QSAR study suggested that LD50 is not a result of interaction of caffeines with Cid1 protein, the docking energies being not correlated with the reported toxicity. However, the docking information was helpful in the choice of leaders for the similarity test set, increasing the accuracy of the predicted LD50 values.

2.7. QSAR and Docking Study on Indolizines by Similarity Clustering

In the present study, a molecular docking analysis has been performed on indolizine derivatives on the proteins 1GA0, 4O0Z, 4O10, then we made a QSAR study to predicting IC50 of indolizine derivative

A set of 25 indolizine were taken from PubChem Database (Table 47) and were divided into a training set (15 molecules) and a test set (10 molecules), taken randomly. The property chosen for modeling was IC50 (μM) 15-LO from soybeans) (see Table 47)[28].

Table 47. *IC50(μM) 15-LO from soybeans, cod Smiles and CID (PubChem)*

	Canonical SMILES	CID	IC_{50}
1	C1=C(C=C[N]2C1=CC(=C2C3=CC=CC=C3)C4=CC=CC=C4)C#N	482634	30±2

2	C1=C(C=C[N]2C1=C(C(=C2C3=CC=CC=C3)C4=CC=CC=C4)OC)C#N	10853428	33±2
3	C1=CC=C(C=C1)COC2=C3C=C(C=CN3C(=C2C4=CC=CC=C4)C5=CC=CC=C5)C#N	57399922	31±2
4	[N]12C(=C(C(=C1C=C(C=C2)C#N)CC3=CC=CC=C3)C4=CC=CC=C4)C5=CC=CC=C5	491919	27±2
5	[N]12C(=C(C(=C1C=C(C=C2)C#N)C3=CC=CC=C3)C4=CC=CC=C4)C5=CC=CC=C5	482635	28±1
6	COC1=CC=CC=C1C2=C3C=C(C=CN3C(=C2C4=CC=CC=C4)C5=CC=CC=C5)C#N	10644682	20±1
7	COC1=CC=CC(=C1)C2=C3C=C(C=CN3C(=C2C4=CC=CC=C4)C5=CC=CC=C5)C#N	10668699	20±1
8	C1=CC=C(C=C1)C2=C(N3C=CC(=CC3=C2C=O)C#N)C4=CC=CC=C4	491918	29±1
9	CC(=O)C1=C2C=C(C=CN2C(=C1C3=CC=CC=C3)C4=CC=CC=C4)C#N	491856	23±1
10	C1=CC(=CC=C1C2=C(N3C=CC(=CC3=C2)C#N)C4=CC=C(C=C4)F)F	57394641	30±6
11	O(C1=C4[N](C(=C1C2=CC=CC=C2)C3=CC=CC=C3)C=C(C=C4)C#N)C(=O)C	10066595	31±2
12	CC(C1=C2C=C(C=CN2C(=C1C3=CC=CC=C3)C4=CC=CC=C4)C#N)O	491916	26±2
13	CC1=C2C=C(C=CN2C(=C1C3=CC=CC=C3)C4=CC=CC=C4)C#N	491857	27±2
14	C1=CC=C(C=C1)C2=C(N3C=CC(=CC3=C2C(=O)C4=CC=CC=C4)C#N)C5=CC=CC=C5	491920	23±1
15	C1=CC=C(C=C1)C2=C(N3C=CC(=CC3=C2CO)C#N)C4=CC=CC=C4	491917	26±1
16	C1=CC=C(C=C1)C2=C(N3C=CC(=CC3=C2OS(=O)(=O)C4=CC=CC=C4)C#N)C5=CC=CC=C5	57394621	24±3
17	CN(C)S(=O)(=O)OC1=C2C=C(C=CN2C(=C1C3=CC=CC=C3)C4=CC=CC=C4)C#N	57396377	25±2
18	C1=CC(=CC=C1C2=C(N3C=CC(=CC3=C2)C#N)C4=CC=C(C=C4)Cl)Cl	57392897	33±7
19	C1=C(C=C[N]2C1=C(C(=C2C3=CC=CC=C3)C4=CC=CC=C4)OCC5=CC=C(C=C5)C)C#N	57401703	37±4
20	C1=C(C=C[N]2C1=CC(=C2C3=CC=CC=C3)C4=CC=CC=C4)C#N	482634	30±2
21	CC1=CC=C(C=C1)C2=C(N3C=CC(=CC3=C2)C#N)C4=CC=C(C=C4)C	57392898	27±3
22	COC(C1=CC=CC=C1)C2=C3C=C(C=CN3C(=C2C4=CC=CC=C4)C5=CC=CC=C5)C#N	491877	21±3
23	CS(=O)(=O)OC1=C2C=C(C=CN2C(=C1C3=CC=CC=C3)C4=CC=CC=C4)C#N	53855501	22±2
24	CCS(=O)(=O)OC1=C2C=C(C=CN2C(=C1C3=CC=CC=C3)C4=CC=CC=C4)C#N	57403346	28±6
25	C1=CC=C(C=C1)COCOC2=C3C=C(C=CN3C(=C2C4=CC=CC=C4)C5=CC=CC=C5)C#N	57403432	46±4

(c) Similarity Cluster Validation

Validation can be performed by calculating IC50 for the molecules in the test set with equations learned on clusters of similarity: each of the 10 molecules is the leader in its own cluster best scored in the docking step, selected by (2D) similarity among the 15 structures of the initial learning set. The values IC50calc. for each of the 10 molecules in the test set were computed by 10 new equations (the leader being left out) with the same descriptors as in eq. 11, Table 7. The data is shown in Figure 52, n=10; R2=0.955; s=1.329; F=171.735.

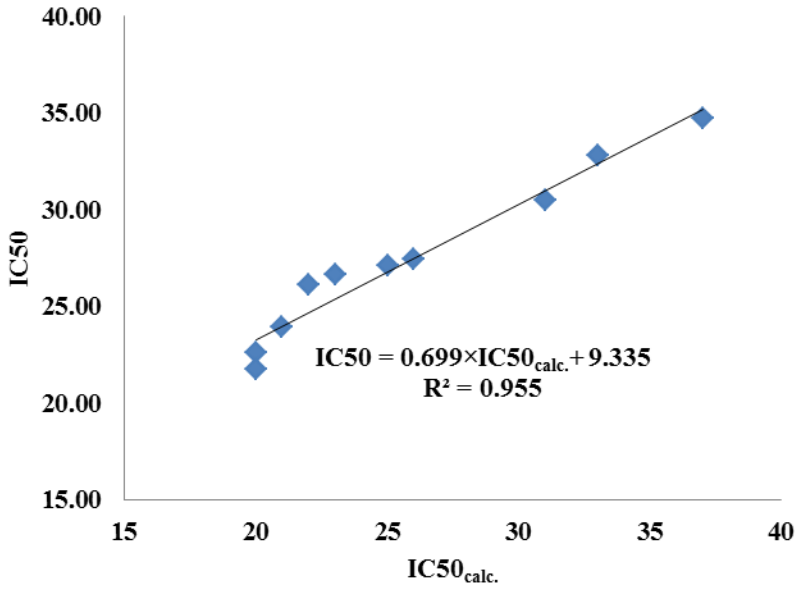

Figure 52. IC50 vs. $IC50_{calc.}$ by similarity clusters

Beta lactamase and nicotinamide phosphoribosyltransferase, has been investigated for its potential binding affinity with selective indolizine derivatives. The docking result of the study of 25 molecules demonstrated that the binding energies when we use Beta lactamase were in the range of -9.8 kcal/mol to -6.1 kcal/mol, when use Nicotinamide phosphoribosyltransferase were in the range of -9.7 kcal/mol to -5.2 kcal/mol.

2.8. QSAR and Molecular Docking Studies of Serotonin Derivatives

In this paper we have carried out a docking study to identify the geometric description of a pharmacophore [29, 30] in the interaction of this class of ligands with protein 3ADX and 2YX8 . A set of 35 derivatives of serotonin were taken from PubChem Database and have been screened molecular docking (Table 55). The property chosen for modeling was LD50 (on rat, intraperitoneal route administered).

Table 55. Serotonin molecular structures (in SMILES code) and LD50 (taken from PubChem)

Nr. Crt.	**Canonical SMILES**	log P	LD50
1	C1=CC=C2C(=C1)C(=CN2)CCN	1.6	100
2	CNCCC1=CNC2=CC=CC=C21	2.1	158
3	CC1=CC2=C(C=C1)NC=C2CCN	1.9	50
4	C1=CC=C2C(=C1)C(=CN2)CC(=O)O	1.4	150
5	C1=CC=C2C(=C1)C(=CN2)C=O	1.7	600
6	CC(=O)OC1=CNC2=CC=CC=C21	2	600
7	C1=CC=C2C(=C1)C(=CN2)CCO	1.8	351
8	CN(C)CCC1=CNC2=C1C=C(C=C2)O	1.2	290

9	C1=CC=C2C(=C1)C=CN2	2.1	117
10	CN(C)CCC1=CNC2=C1C(=CC=C2)O	2.1	196
11	CN1C(=O)CC2=CC=CC=C21	1	1177
12	CCC(=O)NCCC1=CNC2=CC=CC=C21	2.3	900
13	C1=CC=C2C(=C1)C(=CN2)CCC(=O)O	1.8	100
14	CCNCCC1=CNC2=CC=CC=C21	2.4	562
15	C1=CC=C2C(=C1)C=C(N2)O	2.4	400
16	CC(=O)C1=CC2=C(C=C1)NC=C2	2	450
17	C1=CC2=C(C=CN2)C=C1C(=O)CO	1.3	600
18	CC(=O)CC1(C2=CC=CC=C2NC1=O)O	0.2	800
19	CN(C)CCC(C1=CNC2=CC=CC=C21)O	1.8	767
20	COC1=CC2=C(C=C1)NC=C2CC(=O)O	1.4	98
21	C1=CC2=C(C=CN2)C=C1O	2	1000
22	COC1=CC2=C(C=C1)NC=C2	2.1	370
23	CC(CC1=CNC2=CC=CC=C21)N	2	20
24	CCC(CC1=CNC2=CC=CC=C21)N	2.5	400
25	CN(C)CC1=CNC2=CC=CC=C21	1.8	122
26	C1=CC=C2C(=C1)C(=CN2)C(=O)CO	1.4	700
27	CCC(=O)NCCC1=CNC2=C1C=C(C=C2)OC	1.3	850
28	CC1=C(C2=C(N1)C=CC(=C2)OC)CCN(C)C	2.7	100
29	COC1=CC2=C(C=C1)NC=C2CCN	0.5	176
30	CC(=O)C1=CNC2=C1C(=O)CCC2	0.5	233
31	CC1=C(NC2=C1C(=O)CCC2)C(=O)C	1.2	533
32	CC1=CC2=C(C=C1)NC=C2C(=O)CO	1.8	600
33	COC1=CC2=C(C=C1)NC=C2C(=O)CO	1.1	600
34	COC1=CC2=C(C=C1)C(=CN2)C(=O)CO	1.1	600
35	CNS(=O)(=O)CC1=CC2=C(C=C1)NC=C2CCN(C)C	0.9	200

Model validation

(a) Leave-one-out

Leave-one-out analysis related to the models listed as best in Table 60 are presented in Table 61.

Table 61. Leave-one-out for LD50

	Descriptors	Q^2	R^2-Q^2	St. $Error_{loo}$	F_{loo}
1	SD	0.823	0.029	125.176	107.368
5	SD, C	0.819	0.04	126.660	104.331
11	SD, CjDi, CjDe	0.812	0.052	129.079	99.603

(b) External Validation

The values LD50 for the test set of serotonin were calculated by using equation 11 in Table 60. Data is plotted in Figure 58, n=10, R^2=0.742; s=182.028; F=22.983.

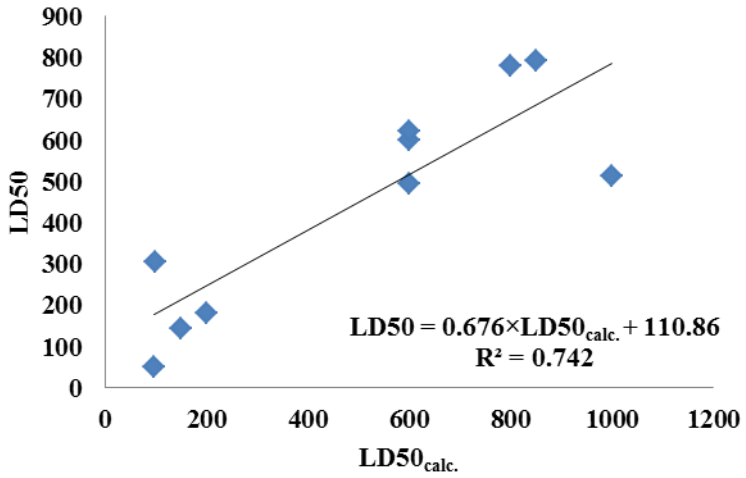

Figure 58. LD50 vs. LD50 $_{calc}$. for the test set (external validation)

(c) Similarity Cluster Validation

The values LD50 for the test set of serotonin were calculated by using equation 11 in Table 60. Data is plotted in Figure 7, n=10; R^2=0.962; s=69.361; F=205.390.

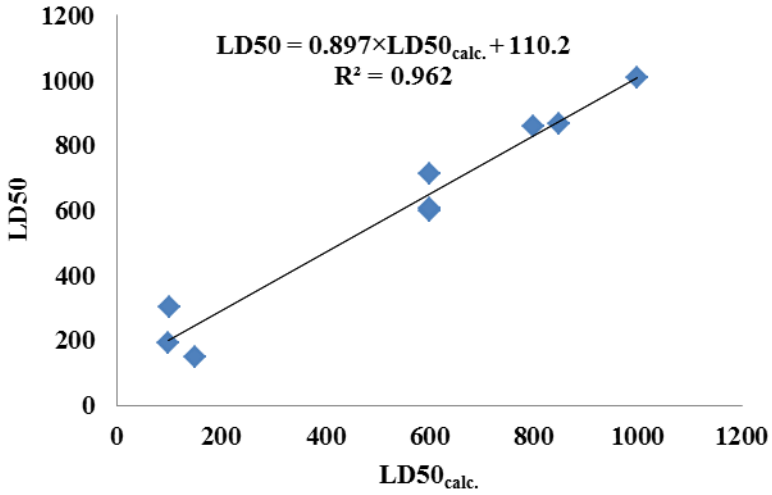

Figure 59 . LD50 vs. LD50 $_{calc..}$ by similarity clusters

Prediction of LD50 is more accurate when using the similarity clusters (R^2= 0.962), compared to the classical external validation of the model (R^2= 0.742). The binding energies provided by the docking step correlates very poorly (R^2= 0.009 for 3ADX and R^2= 0.242 for 2YX8 with the LD50 values); however, selection of the leaders in the similarity test as the best docked ligands clearly improved the

LD50 prediction. The binding energies provided by the docking step correlates very good (n=10; R^2=0.849; s=0.165; F=4.508 for 3ADX with the LD50, HOMO, LUMO ω and π values)[31].

2.9. QSAR and docking studies of Chlorpropamide derivates

The present work focuses on the molecular docking analysis of chlorpropamide and its analogues against protein [32]. This work was carried out by molecular docking studies to determine whether 40 molecules of chlorpropamide interact whit protein CYTOCHROME P450 2C9 and to find the orientation that enhances this interaction as well as minimizing the total energy of the interaction complex [33,34].

A set of 40 were taken from PubChem Database (Table 62); the set was divided into a training set (25 molecules) and a test set (15 molecules), taken randomly. The property chosen for modeling was log P (calculated, Table 1) and LD50 (on mouse, oral route administrated, Table 62

Table 62. Chlorpropamide molecular structures and their log P (taken from PubChem)

Mol.	Canonical SMILES	log P	LD50
1	CCCNC(=O)NS(=O)(=O)C1=CC=C(C=C1)Cl	2.3	580
2	CCCCNC(=O)NS(=O)(=O)C1=CC=C(C=C1)C	2.3	490
3	CCCCNC(=O)NS(=O)(=O)C1=CC=C(C=C1)CO	1.1	490
4	CCCCNC(=O)NS(=O)(=O)C1=CC=C(C=C1)C(=O)O	1.7	490
5	COCCOC1=CN=C(N=C1)NS(=O)(=O)C2=CC=CC=C2	1.1	2800
6	C1=CC(=CC=C1N)S(=O)(=O)NC(=S)N	-0.7	3240
7	C1COCCN1S(=O)(=O)C2=CC=C(C=C2)N	0.1	2150
8	CC(=O)NCCC1=CC=C(C=C1)S(=O)(=O)Cl	1.6	200
9	CC1=CC=C(C=C1)S(=O)(=O)NC(=NCC2=CC=CC=C2)SC	3.8	300
10	CC(=O)NC1=CC=C(C=C1)S(=O)(=O)NC(=NCC2=CC=CC=C2)SC	2.6	300
11	CCOC(=O)NC1=CC=C(C=C1)S(=O)(=O)N2CCOCC2	0.6	2500
12	CC1=CC(=NC(=N1)NS(=O)(=O)C2=CC=C(C=C2)N)C	0.3	50000
13	COC1=CN=C(N=C1)NS(=O)(=O)C2=CC=C(C=C2)N	0.4	16000
14	C1=C(C(=CC(=C1Cl)Cl)Cl)NS(=O)(=O)C2=C(C(=CC(=C2Cl)Cl)Cl)O	6.4	179
15	CC1=C(ON=C1C)NS(=O)(=O)C2=CC=C(C=C2)N	1	6800
16	C1=CC(=CC(=C1)S(=O)(=O)N)N	-0.4	4500
17	C1=CC(=C(C=C1S(=O)(=O)N)Cl)N	0.8	2850
18	CNS(=O)(=O)C1=CC(=C(C=C1)N)Cl	1.2	3000
19	CC(=O)NC1=C(C=C(C=C1)S(=O)(=O)N)Cl	0.1	5000
20	C1=CC(=CC=C1S(=O)(=O)N)Br	1.4	1700
21	CCS(=O)(=O)C1=CC=C(C=C1)F	1.6	542
22	COS(=O)(=O)C1=CC=CC=C1	1.2	250

23	C1=CC(=CC=C1O)S(=O)(=O)O	0.2	6400
24	C1=CC(=CC=C1OS(=O)(=O)C2=CC=C(C=C2)Cl)Cl	4.3	1475
25	CC1=C(C=C(C=C1)S(=O)(=O)O)N(CCCl)CCCl	2.3	200
26	CS(=O)(=O)OC1=C(C=C(C=C1)Cl)Cl	2.8	1070
27	C1=CC(=C(C=C1S(=O)(=O)N)Cl)N	0.8	2950
28	CCS(=O)(=O)C1=CC=C(C=C1)N	0.2	1000
29	C(CS(=O)(=O)C1=C(C(=C(C(=C1Cl)Cl)C#N)Cl)Cl)CCl	4.3	300
30	CCCCNS(=O)(=O)C1=CC=CC=C1	2.1	2500
31	C1=CC=C(C=C1)NS(=O)(=O)C2=CC=C(C=C2)F	2.3	750
32	CC(=O)NC1=C(C=C(C=C1)S(=O)(=O)NC)Cl	0.5	3000
33	CNS(=O)(=O)C1=CC=C(C=C1)O	0.9	2500
34	C1=CC(=CC=C1N)S(=O)(=O)N	-0.6	3000
41	C1CCC(CC1)NC(=O)NS(=O)(=O)C2=CC=C(C=C2)Cl	3.4	1525
36	CC1=NC(=NC=C1)NS(=O)(=O)C2=CC=C(C=C2)N	0.1	25000
37	COC1=NC(=NC(=C1)NS(=O)(=O)C2=CC=C(C=C2)N)OC	1.6	3200
38	CC1=CC(=NC(=N1)NS(=O)(=O)C2=CC=C(C=C2)N)C(F)(F)F	0.8	4150
39	C1=CC(=CC(=C1)S(=O)(=O)NC2=NC=C(C=N2)Cl)N	1.1	700
40	COC1=NC=NC(=C1)NS(=O)(=O)C2=CC=C(C=C2)N	0.8	4680

(b) External Validation

The values $LD50_{calc.}$ for each of the 15 molecules in the test set were chosen based on the lowest energy docking and computed with the same descriptors and the eq. 10, Table 66. Data are plotted in Figure 64; n=15; R^2=0.761; s=3510.249; F=41.371 5.

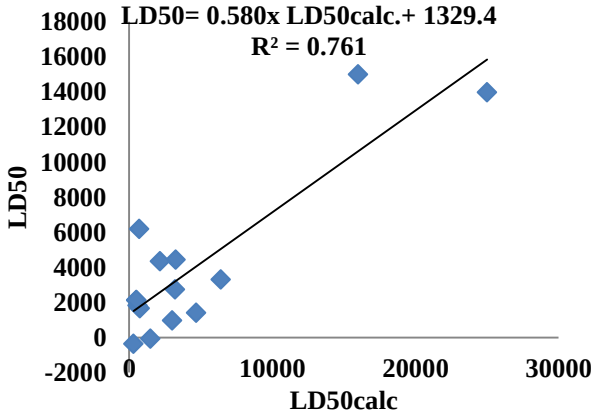

Figure 64. LD50 vs. $LD50_{calc}$ for the test set (external validation)

(c) Similarity Cluster Validation

The values LD50 calc. for each of the 15 molecules in the test set were chosen based on the lowest energy docking**.** Data are plotted in Figure 65, n=15; R^2=0.996 s=472.684; F=2985.482.

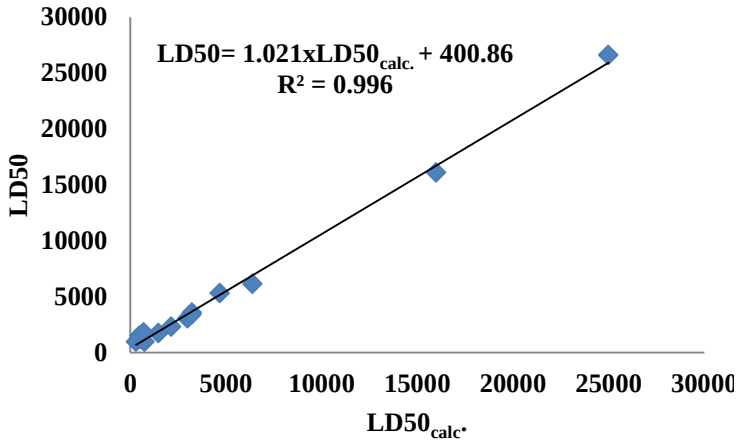

Figure 65. LD50 vs. LD50 $_{calc.}$.

by similarity clusters

CYTOCHROME P450 2C9 has been investigated for its potential binding affinity with selective chlorpropamide derivatives.

The docking result of the study of 40 chlorpropamide molecules demonstrated that the binding energies were in the range of -8.9 kcal/mol to -4.4 kcal/mol, with the minimum binding energy of -8.9 kcal/mol.

3. General conclusion

• Download of sets of molecules (flavonoids, caffeine, indolyl, 1-methylpiperidin-4-yl propanoate, dopamine, indolizine, serotonin and cloropropamide) from the database PubChem.

• The set was divided into a school set and a test set, the latter being used to validate the models, the so-called external validation set.

• Splitting data set was taken random or docking energy minimum.

• protein-ligand docking was performed AutoDock4.2 program.

• QSAR model validation method leave-one-out

• Validation was done through a new version of prediction using clusters of similarity.

• Prediction log P values, LD50 and IC50 by clusters of similarity is greater than that obtained through

external validation.

References (selected):

1. G. Di Carlo, N. Mascolo, A.A. Izzo, F. Capasso, Life Sci., **1999**, 65, 337–353.

2. R.Malek and N. Mousseau, J Chem Phys, **2010,** 135, 6, 7723–7728.

3. A.T. Balaban, A. Chiriac, I. Motoc, and Z. Simon, Steric Fit in QSAR (Lectures Notes in Chemistry, Vol. 15), Springer, Berlin, **1980**, Chap. 6.21.

4. S. F. Sousa, A. J. Ribeiro, J. T. S. Coimbra, R. P. P. Neves, S. A. Martins, H. N. S. Moorthy, P. A. Fernandes, M. J. Ramos, Curr. Med. Chem. **2013**, 20, 5, 2296–314.

5. C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney., Advanced Drug Delivery Reviews, **1997**, 23 (1-3), 3-25.

6. D. Voet, J. Voet, C. Pratt, Life at the Molecular Level, **2008**, 840.

7. C. D. Nicholson, Psychopharmacology, **1990**, 101, 147–159.

8. D. Kruzlicova, M. Danihelova, M. Veverka, Nova Biotechnologica et Chimica , **2012**, 1-11.

9. T. E. Harsa, A. M. Harsa, Beata Szefler, Cent. Eur. J. Chem., **2014**, 12, 3, 365-376

10. M. V. Diudea, MATCH Commun. Math. Comput. Chem. **1997**, 35, 169.

11. P. R. Brodfuehrer, B. Chen, T. R. Sattelberg. P. R. Smith, J. P. Reddy, D. R. Stark, S. L. Quinlan, J. G. Reid, J. Org. Chem., **1997**, 62, 9192-9202.

12. M. T. Lin, H. J. Tsay, W. H. Su, F. Y. Chueh, Am. J. Physiol. Integr. Comp. Physiol, **1998**, 274, 1260–1267.

13. L. Imeri, M. Mancia, S. Bianchi, MR Opp, Neuroscience, **1999**, 95, 445–452.

14. A. D. DeWeese, and T. W. Schultz, Environ. Toxicol. **2001**, 16,54–60.

15. T. E. Harsa, A. M. Harsa, L. Jantschi and M. V. Diudea, Journal of Chemical and Pharmaceutical Research, **2015**, 7, 3, 2378-239.

16. H. Shinotoh, K. Fukushi, S. Nagatsuka, T. Irie, Curr Pharm Des. **2004**, 10,13, 1505–17.

17. D. M. Jewett, T. B. Nguyen, D. E. Kuhl, M. R. Kilbourn, Nuclear Medicine and Biology, **1998**, 25, 751–754.

18. S.E. Snyder, L. Tluczek, D.M .Jewett, T.B. Nguyen, D.E. Kuhl, M.R. Kilbourn, Nucl Med Biol**.1998**, 25, 8, 751–4.

19. T. E. Harsa, A. M. Harsa, and M. V. Diudea, Studia Univ. "Babes-Bolyai", 1, **2015**, 165-176.

20. N. E.Anden, A. Carlsson, A. Dahlstroem, K. Fuxe, N. A. Hillarp, K. Larsson, Life Sci **1964**, 3, 523–530.

21. M. Jaber, S. W. Robinson, C. Missale, M. G. Caron, Neuropharmacol., **1996**, 35, 1503–1519.

22. Y. Li, Y. Xie, Y. Qin, Sensors and Actuators B, **2014**, 191, 227– 232.

23. M. Moreno-Smith, S. J. Lee, C. Lu, A. S. Nagaraja, G. He, R. Rupaimoole, H. D. Han, N. B. Jennings, J. W. Roh, M. Nishimura, Y. Kang, J. K. Allen, G. N. Armaiz, K. Matsuo, M.M.K. Shahzad, J. Bottsford-Miller, R. R. Langley, S. W. Cole, S. K. Lutgendorf, Z. H. Siddik, A. K. Sood, Neoplasia, **2013**, 15,502–510.

24. O. G. Dadrass, M. A. Sobhani , A. Shafiee, Mahmoudian M., Daru, **2014**,12,1.

25. H. Zhang, Y. Wang, F. Xu, Journal of Molecular Structure, **2014**, 1076, 153–159.

26. The RCBS Protein data bank. [http://www.rcsb.org/pdb].

27. N. M. Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, G. R. Hutchison, Open Babel: An open chemical toolbox". Journal of Cheminformatics 3, **2011**, 33.

28. R. Abagyan, M. Totrov, Curr Opin Chem Biol., **2001**, 5,375-82.

29. S. J. Tepper, A. M. Rapoport, F. D. Sheftell, Agonist. Arch Neurol., 7, **2002**, 1084-1088.

30. J. E. Schaffer, Curr Opin Lipidol **2003**, 14, 281–287.

31. T. Terada, K. Sawada, H. Saito, Y. Hashimoto, K. Inui, Eur J Pharmacol,. **2000**, 392, 1-2 ,11-7.

32. V. Balavignesh, E. Srinivasan, N. G. Ramesh Babu, N. Saravanan, Int. J. of Pharm. & Life Sci. (IJPLS), **2013**, 4, 4, 2548-2558.

33. S. Y. Lu, Y. J. Jiang, J. Lv, T. X. Wu, QS and W. L. Zhu, J Mol Graph Model, **2010**, 28, 766-774.

34. P. Srinivasan, A. Sudha, A. Shahul Hameed, K. Prasanth M. Karthikeyan, Journal of Pharmacy Research, **2011**, 4, 1, 136-140.