STUDIA UBB CHEMIA, LXI, 1, 2016 (p. 305-316)
(RECOMMENDED CITATION)

Dedicated to Professor Mircea Diudea on the Occasion of His 65th Anniversary

QSAR STUDY OF PHENOTHIAZINES

ATENA PÎRVAN MOLDOVAN[a,*], SARA ERSALI[a], RALUCA POP[b]

ABSTRACT. A QSAR study on a set of 30 phenothiazines performed within a hypermolecule frame, to model their logP and LD_{50} values, is reported. The initial set of molecules was split into a training set and the test set; Cluj topological indices and some quantum mechanical descriptors have been used to derive the models, which were next tested for predictability by LOO, external validation and similarity clustering.

Key words: *phenothiazine, hypermolecule, LD_{50}, logP, topological indices.*

INTRODUCTION

Phenothiazine is an organic heterocyclic compound, of the class of thiazines, with the brute formula $S(C_6H_4)_2NH$, of which skeleton occurs in various antipsychotic, antihistaminic, antiemetic, etc. drugs. Phenothiazine was synthesized by Bernthsen in 1883 by melting the diphenylamine with sulfur; its medicamentous derivatives are currently synthesized by the cyclization of substituted diphenylamines or diphenyl sulfides. Synthesis of methylene blue was reported in 1876 and is still used as antiseptic, antihelminthic drug.

Phenothiazine antipsychotics, like chlorpromazine and prochlorperazine, are used to treat serious mental and emotional disorders, including schizophrenia and other psychotic disorders. Phenothiazine antipsychotics are classified into three groups, differing with respect to the substituent on nitrogen: the aliphatic compounds, piperidine compounds and piperazine derivatives. As antihistaminic, the promethazine is the most used phenothiazine.

[a] *Babeş-Bolyai University, Faculty of Chemistry and Chemical Engineering, 11 Arany Janos str., RO-400028, Cluj-Napoca, Romania*
[b] *University of Medicine and Pharmacy "Victor Babes" Timisoara, Faculty of Pharmacy, E. Murgu Square 2, 300041 Timisoara, Romania*
** Corresponding Author: atenamoldovan99@gmail.com*

Several water-soluble phenothiazines, such as methylene blue, methylene green, thionine, etc. can be electropolymerized, the resulted polymer inding industrial applications [1].

Quantitative structure–activity relationship (QSAR) studies, attempt to predict (in the light of the paradigm that relates a biological activity, or a physicochemical property, of a compound to its chemical structure) the activity of tested compounds and to suggest structural features which could enhance that biological activity, in the process of drug design [2-4].The concept of similarity is used in grouping chemical compounds according to their molecular structure, biological effects or physicochemical properties; it has found extensive use in drug discovery [4].

Topological indices are molecular descriptors, useful in QSAR studies; they are integer or real-valued numbers, derived from the connectivity and other topological matrices computed on the molecular graph associated to a molecule. Among thousands of topological indices, the Cluj indices, defined by Diudea [5, 6], are among the most simple and versatile ones in coding the chemical information of a molecular graph. Indices are calculated from the Cluj topological matrices, as half sum of matrix entries, by using the original TopoCluj software [7].

The octanol–water partition coefficient (log P) describes the chemical lipophilic/hydrophilic characteristics. Log P is the ratio of a chemical concentration in the octanol phase to its concentration in the aqueous phase of a two-phase system at equilibrium. Log P is involved in the passive transport of a drug molecule through cellmembrane [3].

This QSAR study was performed following Diudea's algorithm [8]; it is based on the alignment of molecules over a hypermolecule [9] and a correlation weighting procedure [10, 11] coupled with a predictive validation of the model descriptors within similarity clusters [12] performed for each molecule in the test set. The algorithm can be extended with other powerful statistical tools (e.g. PLS or PCA) but we limited here to the more common multi linear regression in achieving the best prediction of a chosen property, like logP or LD_{50}.

COMPUTATIONAL

The structures have been optimized at Hartree-Fock HF (6-31g(d,p)) level of theory, in gas phase, by Gaussian 09 [13]. Topological indices have been computed by TOPOCLUJ software [7]. The modeled properties: logP and LD_{50}, along with some of the molecular descriptors, like Charges, D3D, Detour, Distance, IE[CfMax], IP[CfMin], IP[CfMax], E HOMO (a.u.), HL gap (eV), Chemical potential (eV), Hardness (eV), Elecrophilicity (eV), are listed in Tables 1 and 2, respectively.

RESULTS AND DISCUSSION

Data set

A hypermolecule (Figure 1) was built up by superposing all the 30 molecules under study. The hypermolecule is considered to mimic the investigated statistical hyperspace [9] and works like a biological receptor, over which the ligands are aligned. According to this alignment, *binary vectors* were constructed, with 1 when for a given position of the hypermolecule exists an atom in the current molecule, and zero, otherwise. In the above binary vectors, the values 1 are next replaced by mass fragments. Table 1 lists the phenothiazines of the data set, with the properties to be modeled: logP and LD_{50} (intraperitoneal, mouse).

Table 1. List of studied phenothiazines with their name, CID and properties logP and LD_{50}.

No.	CID	logP	LD50 mg/kg	Name	Canonical Smiles
1	2726	5.41	14	chlorpromazine	CN(C)CCCN1C2=CC=CC=C2SC3=C1C=C(C=C3)Cl
2	2801	5.19	150	clomipramine	CN(C)CCCN1C2=CC=CC=C2CCC3=C1C=C(C=C3)Cl
3	2995	4.90	85	desipramine	CNCCCN1C2=CC=CC=C2CCC3=CC=CC=C31
4	3089	3.34	190	fonazine/dimetothiazine	CC(CN1C2=CC=CC=C2SC3=C1C=C(C=C3)S(=O)(=O)N(C)C)N(C)C
5	3781	3.66	62	isothipendyl	CC(CN1C2=CC=CC=C2SC3=C1N=CC=C3)N(C)C
6	4066	4.70	54	mequitazine	C1CN2CCC1C(C2)CN3C4=CC=CC=C4SC5=CC=CC=C53
7	4744	4.10	185	perazine	CN1CCN(CC1)CCCN2C3=CC=CC=C3SC4=CC=CC=C42
8	4747	3.52	115	periciazine	C1CN(CCC1O)CCCN2C3=CC=CC=C3SC4=C2C=C(C=C4)C#N
9	4748	4.20	64	perphenazine	C1CN(CCN1CCCN2C3=CC=CC=C3SC4=C2C=C(C=C4)Cl)CCO
10	4917	4.88	120	prochlorperazine	CN1CCN(CC1)CCCN2C3=CC=CC=C3SC4=C2C=C(C=C4)Cl
11	4926	4.55	140	promazine	CN(C)CCCN1C2=CC=CC=C2SC3=CC=CC=C31
12	4927	4.81	124	promethazine	CC(CN1C2=CC=CC=C2SC3=CC=CC=C31)N(C)C
13	5452	5.90	65	thioridazine	CN1CCCCC1CCN2C3=CC=CC=C3SC4=C2C=C(C=C4)SC
14	5566	5.03	120	trifluoperazine	CN1CCN(CC1)CCCN2C3=CC=CC=C3SC4=C2C=C(C=C4)C(F)(F)F
15	6075	5.60	140	mepazine/pecazine	CN1CCCC(C1)CN2C3=CC=CC=C3SC4=CC=CC=C42
16	6077	4.20	350	acetylpromazine	CC(=O)C1=CC2=C(C=C1)SC3=CC=CC=C3N2CCCN(C)C
17	6761	4.40	80	pipamazine	C1CN(CCC1C(=O)N)CCCN2C3=CC=CC=C3SC4=C2C=C(C=C4)Cl
18	10646	4.70	190	pyrathiazine	C1CCN(C1)CCN2C3=CC=CC=C3SC4=CC=CC=C42
19	14670	3.40	135	prothypendyl	CN(C)CCCN1C2=CC=CC=C2SC3=C1N=CC=C3
20	14677	5.23	183	methdilazine	CN1CCC(C1)CN2C3=CC=CC=C3SC4=CC=CC=C42
21	16414	4.80	119	7-hidroxyclorpromazine	CN(C)CCCN1C2=C(C=C(C=C2)O)SC3=C1C=C(C=C3)Cl
22	19396	3.40	185	oxomemazine	CC(CN1C2=CC=CC=C2S(=O)(=O)C3=CC=CC=C31)CN(C)C
23	19675	4.21	98	piperacetazine	CC(=O)C1=CC2=C(C=C1)SC3=CC=CC=C3N2CCCN4CCC(CC4)CCO
24	65535	4.90	225	diethazine	CCN(CC)CCN1C2=CC=CC=C2SC3=CC=CC=C31
25	65750	5.90	90	chlorproethazine	CCN(CC)CCCN1C2=CC=CC=C2SC3=C1C=C(C=C3)Cl
26	68223	4.20	115	fenethazine	CN(C)CCN1C2=CC=CC=C2SC3=CC=CC=C31
27	69500	3.80	210	difazin	CCN(CC)CC(=O)N1C2=CC=CC=C2SC3=CC=CC=C31
28	70413	3.90	163	opromazine	CN(C)CCCN1C2=CC=CC=C2S(=O)C3=C1C=C(C=C3)Cl
29	72287	4.68	58.5	levomepromazine	C[C@@H](CN1C2=CC=CC=C2SC3=C1C=C(C=C3)OC)CN(C)C
30	94280	4.96	206	dimetacrine	CC1(C2=CC=CC=C2N(C3=CC=CC=C31)CCCN(C)C)C

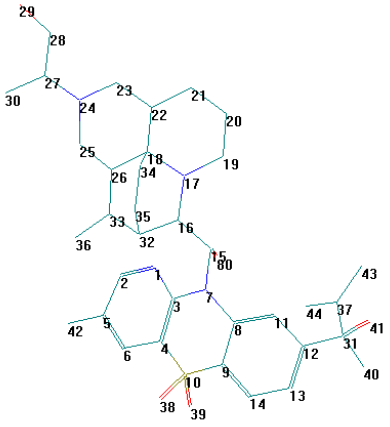

Figure 1. Hypermolecule comprising the features of the dataset

Table 2. The modeled properties: logP and LD_{50} and some of the molecular descriptors computed for 30 phenothiazines

Mol	logP	LD50 mg/kg	Charges	D3D	Detour	Distance	IE[CfMax]	IP[CfMin]	IP[CjMax]	E HOMO (a.u.)	HL gap (eV)	Chem.pot (eV)	Hardness (eV)	E.phil (eV)	SD(logP)	SD(LD50)
1	5.20	~	-0.105	959	2450	896	201	4630	1260	-0.297	11.040	-2.560	5.520	0.590	-54.235	~
2	5.20	150	-0.379	1040	2860	995	212	5030	1350	-0.302	11.340	-2.550	5.670	0.570	-53.820	-9000.05
3	4.90	85	0.060	800	2330	759	144	3770	958	-0.295	11.430	-2.450	5.720	0.520	-54.097	-9033.07
4	3.80	190	-0.280	1550	3810	1540	366	8090	2070	-0.308	10.680	-3.030	5.320	0.860	-55.200	-8928.76
5	3.50	62	-0.668	740	2150	744	156	3730	974	-0.297	11.040	-2.550	5.520	0.590	-55.528	-9040.66
6	4.60	54	-0.162	1130	3320	1090	219	8580	1670	-0.291	11.436	-2.203	5.718	0.424	-54.476	-9059.29
7	4.10	185	-0.341	1370	3500	1390	320	9070	2440	-0.291	11.170	-2.340	5.580	0.490	-54.779	-8932.61
8	3.50	115	-0.034	1610	4200	1720	398	11300	2980	-0.304	10.616	-2.947	5.308	0.818	-55.463	-9015.74
9	4.20	64	-0.065	1910	4560	2000	486	13500	3690	-0.294	11.284	-2.346	5.642	0.488	-54.828	-9034.78
10	4.90	120	-0.098	1530	3840	1540	353	10100	2670	-0.294	11.280	-2.347	5.640	0.488	-54.208	-8976.6
11	4.50	140	-0.196	874	2200	789	178	4030	1130	-0.293	11.180	-2.380	5.590	0.510	-54.807	-8956.05
12	4.80	124	-0.267	759	2150	744	156	3730	974	-0.273	10.690	-2.090	5.350	0.410	-54.571	-9007.79
13	5.90	65	-0.334	1560	3860	1440	311	9490	2390	-0.297	10.900	-2.620	5.450	0.610	-53.196	-9048.29
14	4.70	120	0.636	2110	4940	2080	492	13700	3610	-0.304	10.950	-2.800	5.480	0.720	-54.707	-8989.26
15	5.60	140	-0.268	1070	2870	982	162	6450	1420	-0.294	11.180	-2.410	5.590	0.520	-53.606	-8973.29
16	4.30	350	-0.065	1240	2990	1150	268	6080	1630	-0.301	10.360	-3.010	5.180	0.870	-54.393	-8762.6
17	4.40	80	0.074	1870	4540	1980	474	13300	3590	-0.267	10.471	-2.013	5.235	0.387	-54.529	-9045.99
18	4.70	190	-0.275	959	2560	889	176	5340	1280	-0.286	11.700	-2.107	5.673	0.391	-54.243	-8906.66
19	3.40	135	-0.642	807	2200	789	178	4030	1130	-0.280	10.610	-2.320	5.300	0.510	-55.764	-8988.92
20	4.60	183	-0.208	833	2540	859	146	5280	1130	-0.280	10.770	-2.240	5.380	0.470	-54.455	-8930.29
21	4.80	119	-0.192	1060	2720	1020	224	5300	1420	-0.297	11.020	-2.560	5.510	0.600	-54.235	-9000.05
22	3.40	185	0.123	1140	2880	1080	255	5450	1550	-0.333	11.220	-3.460	5.610	1.070	-55.560	-8931.28
23	4.00	~	-0.013	2250	5380	2410	595	16500	4480	-0.299	10.340	-2.960	5.170	0.850	-55.355	~
24	4.90	225	-0.286	951	2410	894	215	4570	1330	-0.291	11.080	-2.360	5.540	0.500	-54.243	-8906.66
25	5.90	90	-0.221	1170	2980	1190	314	6280	1940	-0.297	11.030	-2.560	5.510	0.590	-53.199	-9000.05
26	4.20	115	-0.228	711	1950	658	128	3300	832	-0.287	11.350	-2.122	5.675	0.397	-54.883	-8980.16
27	3.80	210	0.033	990	2610	979	242	4950	1450	-0.304	11.300	-2.620	5.650	0.610	-55.296	-8906.66
28	3.90	163	-0.048	1020	2690	993	223	5110	1390	-0.309	10.970	-2.930	5.480	0.780	-55.299	-8947.64
29	4.80	58.5	-0.323	1150	2950	1130	268	5900	1620	-0.282	10.920	-2.220	5.460	0.450	-54.273	-9038.46
30	5.40	206	-0.536	1020	2650	973	221	4930	1380	-0.262	10.510	-1.870	5.250	0.330	-53.731	-8906.92

Data reduction

In this step, the descriptors with variance <10% and intercorrelation > 0.80 (two descriptors highly correlated bring quite the same information on the molecule, one of them being sufficient) were discarded.

Correlation weighting was performed as follows: the correlation coefficients of the statistically significant positions in the hypermolecule were used to multiply the local descriptors, thus resulting new weighted vectors CD_{ij}. Next, these new descriptors are summed to give a global descriptor, $SD_i = \sum_j CD_{ij}$ which is a linear combination of the local correlating descriptors for the significant positions in the hypermolecule (for logP model, significant positions are 1, 10, 12, 18, 19, 21, 23, 24, 26, 27, 28, 30, 31, 32, 40, 80; for LD_{50} model, these are: 1, 12, 16, 18, 19, 21, 27, 28, 31, 32, 33, 34, 37, 38, 40, 41).

Basic equations that describe the relationships between values of property or biological activity of compounds and their structures were obtained:

$$\log P = 59.096 + SD_{\log P} \quad (1)$$

n=30; R^2=0.946; s=0.165; F=488.078

$$LD_{50} = 9113.289 + SD_{LD50} \quad (2)$$

n=28; R^2=0.956; s=13.964; F=566.487

(molecules 1 and 23 were outliers)

QSAR models (for case log P)

The models were performed on the training set (structures 11-30) and the best results (in decreasing order of R^2) are listed below in Tables 3 and 4.

Table 3. The best bivariate models for logP in the training set

Property		Descriptors	R^2
logP	SD(logP)	IP[CfMax]	0.9487
	SD(logP)	Detour	0.9482
	SD(logP)	D3D	0.9480
	SD(logP)	El.phil (eV)	0.9480
	SD(logP)	IE[CfMax]	0.9479
	SD(logP)	IE[CjMax]	0.9479
	SD(logP)	Chem.pot.(eV)	0.9474
	SD(logP)	E HOMO (a.u.)	0.9469
	SD(logP)	Charges	0.9467
	SD(logP)	Hardness(eV)	0.9465

Table 4. The best trivariate models for logP in the training set

Property	Descriptors			R^2
logP	IP[CfMin]	Chem.pot (eV)	SD(logP)	0.95114
	Chem.pot (eV)	SD(logP)	IP[CjMin]	0.95110
	IP[CjMax]	Chem.pot (eV)	SD(logP)	0.95109
	Chem.pot (eV)	SD(logP)	IP[CjMax]	0.95109
	Distance	Chem.pot (eV)	SD(logP)	0.95060
	Chem.pot (eV)	SD(logP)	IE[CfMax]	0.95026
	Hardness (eV)	SD(logP)	Detour	0.94954
	D3D	Hardness (eV)	SD(logP)	0.94920

Property	Descriptors			R^2
	SD(logP)	Detour	E HOMO (a.u.)	0.94919
	Charges	Chem.pot (eV)	SD(logP)	0.94919
	E HOMO (a.u.)	SD(logP)	D3D	0.94905
	E HOMO (a.u.)	SD(logP)	Distance	0.94904
	E HOMO (a.u.)	SD(logP)	Charges	0.94833
	Chem.pot (eV)	SD(logP)	E HOMO (a.u.)	0.94755

I. Monovariate regression

$$logP= 57.266+0.966\times SD_{logP} \quad (3)$$
$$n=20;\ R^2=0.946;\ s=0.172;\ F=317.17$$

II. Bivariate regression

$$logP= 57.299+0.968\times SD_{logP}+3.54\times 10^{-5}\times IP[CfMax] \quad (4)$$
$$n=20;\ R^2=0.949;\ s=0.173;\ F=157.155$$

III. Trivariate regression

$$logP= 56.341+0.948\times SD_{logP}+1.26\times 10^{-5}\times IP[CfMin]+0.097\times Chem.pot. \quad (5)$$
$$n=20;\ R^2=0.951;\ s=0.174;\ F=103.82$$

QSAR models (for case LD_{50})

The models were performed on the training set and the best results (in decreasing order of R^2) are listed below in Tables 5 and 6.

I. Monovariate regression

$$LD_{50}=9444.65+1.037\times SD_{LD50} \quad (6)$$
$$n=19;\ R^2=0.940;\ s=13.534;\ F=265.126$$

II. Bivariate regression

$$LD_{50}=9415.945+1.033\times SD_{LD50}-0.0266\times IE[CfMax] \quad (7)$$
$$n=19;\ R^2=0.943;\ s=13.607;\ F=131.573$$

Table 5. The best bivariate models for LD_{50} in the training set

Property	Descriptors		R^2
LD_{50}	SD(LD_{50})	IE[CfMax]	0.9427
	IE[CfMin]	SD(LD_{50})	0.9424
	IE[CjMin]	SD(LD_{50})	0.9424
	SD(LD_{50})	IP[CfMax]	0.9422
	SD(LD_{50})	IP[CjMax]	0.9421
	Distance	SD(LD_{50})	0.9414
	IP[CfMin]	SD(LD_{50})	0.9413
	IP[CjMin]	SD(LD_{50})	0.9412
	D3D	SD(LD_{50})	0.9412
	SD(LD_{50})	Detour	0.9407
	Charges	SD(LD_{50})	0.9401
	HL gap (eV)	SD(LD_{50})	0.9398

Property	Descriptors		R^2
	Hardness (eV)	$SD(LD_{50})$	0.9398
	$SD(LD_{50})$	E.phil (eV)	0.9398
	$SD(LD_{50})$	Chem.pot (eV)	0.9398
	$SD(LD_{50})$	E HOMO (a.u.)	0.9397

III. Trivariate regression

$LD_{50}=9606.267+1.055\times SD_{LD50}-0.2024\times IE[CfMax]+0.047\times Distance$ (8)
$n=19$; $R^2=0.947$; $s=13.498$; $F=89.557$

Table 6. The best trivariate models for LD_{50} in the training set

Property	Descriptors			R^2
LD_{50}	Distance	$SD(LD_{50})$	IE[CfMax]	0.9471
	D3D	$SD(LD_{50})$	Detour	0.9448
	IE[CfMax]	HL gap (eV)	$SD(LD_{50})$	0.9432
	IE[CfMax]	Hardness (eV)	$SD(LD_{50})$	0.9432
	$SD(LD_{50})$	E.phil (eV)	IE[CfMax]	0.9429
	$SD(LD_{50})$	Chem.pot (eV)	IE[CfMax]	0.9428
	$SD(LD_{50})$	E HOMO (a.u.)	IE[CfMax]	0.9427
	Chem.pot (eV)	Distance	$SD(LD_{50})$	0.9415
	IP[CjMin]	$SD(LD_{50})$	E HOMO (a.u.)	0.9412
	$SD(LD_{50})$	Detour	Chem.pot (eV)	0.9407

Model validation

(a) Leave-one-out

The performances in leave-one-out analysis [14] related to the models listed as best in Tables 3-6 are shown in Tables 7 and 8 .

Table 7. Leave-one-out analysis for the best logP models

	Descriptors	Q^2	R^2- Q^2
1	SD(logP)	0.9378	0.0085
2	SD(logP), IP[CfMax]	0.9285	0.0208
3	SD(logP), IP[CfMin], Chem.pot.	0.9096	0.0410

Table 8. Leave-one-out analysis for the best LD_{50} models

	Descriptors	Q^2	R^2- Q^2
1	$SD(LD_{50})$	0.9306	0.0092
2	$SD(LD_{50})$, IE[CfMax]	0.9251	0.0176
3	$SD(LD_{50})$, IE[CfMax], Distance	0.9188	0.0301

(b) External Validation

The values of logP and LD_{50} for the test sets (structures 1-10 for logP; structures 8, 11, 12, 16, 18, 20, 24, 26, 28 for LD_{50}) of phenothiazines were calculated by using Eqs 5 and 8, respectively.

The monovariate correlations are plotted in Figures 2 and 3.

$$\log P_{exp} = -0.674 + 1.132 \times \log P_{calc} \quad (9)$$
$$(n=10,\ R^2=0.940,\ s=0.171,\ F=126.298)$$

$$LD_{50exp} = 10.945 + 0.916 \times LD_{50\ calc} \quad (10)$$
$$(n=9,\ R^2=0.944,\ s=18.942,\ F=116.983)$$

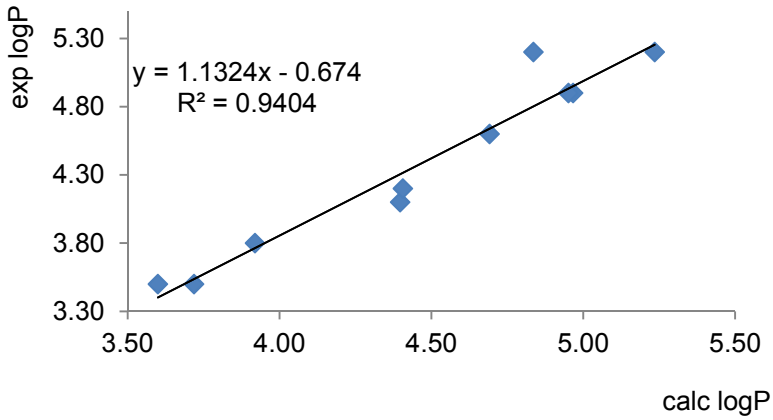

Figure 2. The plot $\log P_{exp}$ vs. $\log P_{calc.}$ for the test set (external validation).

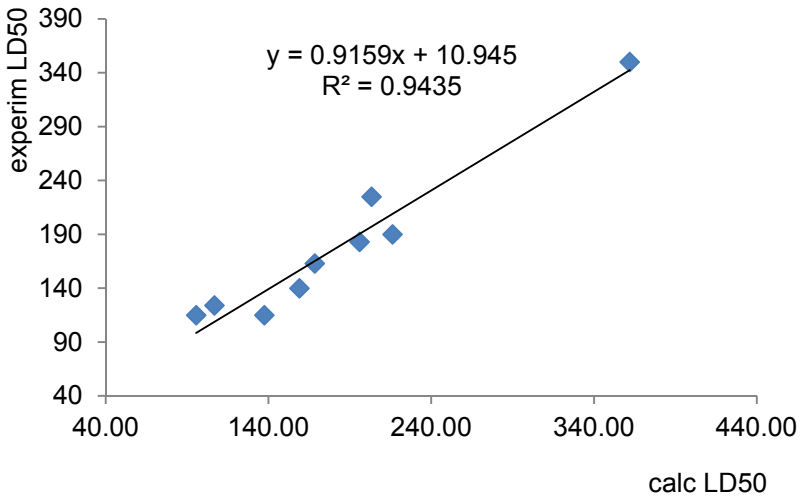

Figure 3. The plot LD_{50exp} vs. $LD_{50calc.}$ for the test set(external validation)

From Figures 2 and 3 one can see that our models show a good predictive ability.

(c) Similarity Cluster Validation

Validation can also be performed by using similarity clusters: each of the 10/9 molecules in the test set, is the leader of its own cluster, selected by 2D similarity among the 20/19 structures of the learning set (each cluster comprising about 12-15 molecules). The values of logP and LD_{50calc} were predicted by 10/9 new equations (the leader being left out) with the same descriptors as in Eqs. 5 and 8, respectively.

The monovariate correlation for logP

$$\log P_{exp} = -0.516 + 1.094 \times \log P_{calc\text{-}clusters} \quad (11)$$
$$n = 10,\ R^2 = 0.946,\ s = 0.162,\ F = 141.674$$

is plotted in Figure 4.

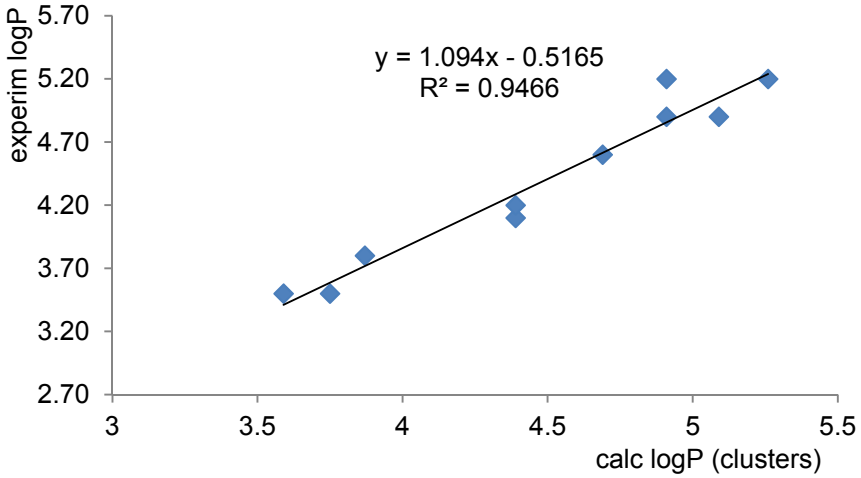

Figure 4. The plot $\log P_{exp}$ vs. $\log P_{calc}$ (by clusters of similarity) for the test set

The monovariate correlation for LD_{50}

$$LD_{50exp} = 10.99 + 0.914 \times LD50_{\ calc\text{-}clusters} \quad (12)$$
$$n = 9,\ R^2 = 0.951,\ s = 17.707,\ F = 134.87$$

is plotted in Figure 5.

Prediction of logP(R^2= 0.946), and LD_{50}(R^2= 0.951), is more accurate when using the similarity clusters, compared to the classical external validation of the model. We limited the model to three variables, keeping in mind the suggestions of Topliss and Costello [15].

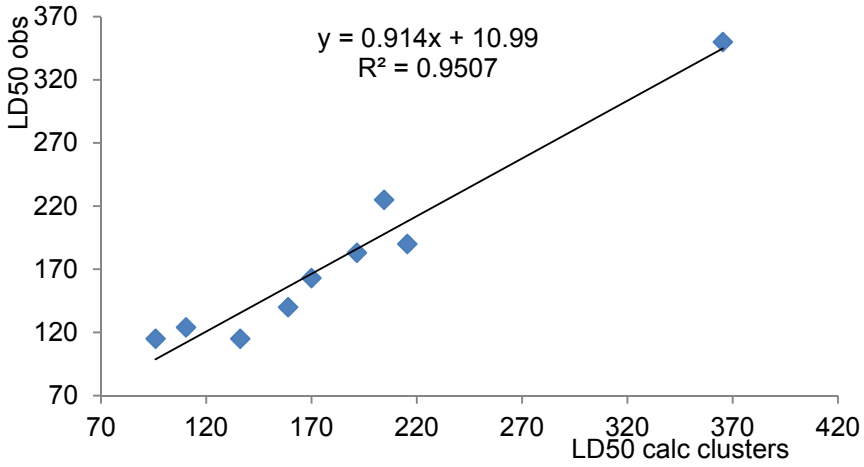

Figure 5. The plot LD_{50} vs. LD_{50calc} (by clusters of similarity) for the test set

CONCLUSIONS

A QSAR study for modeling logP and LD_{50} of a set of 30 phenothiazine derivates, downloaded from some well-known databases, is reported. The approach is based on correlation weighting and alignment over a hypermolecule, that mimics the investigated correlational space. The best models, derived on the learning set of phenothiazines, were validated by leave-one-out test, in the external test set and in a version of prediction, based on clusters of similarity. The models were built up around the „sum descriptor" SD_i that collects the ligand topological informations, a linear combination of local descriptors CD_{ij}, weighted by correlation coefficients of fitting the ligands (*i.e.*, molecules of the learning set) over the hypermolecule. The other topological, global descriptors were calculated by TOPOCLUJ software program. The clusters of similarity ensured the „congeneric state" of molecules on which the prediction of property/ activity is made for the molecules in the test set, thus surpassing the models found in the learning set and also in the test set, by external validation.

ACKNOWLEDGMENT

The authors are endebted to Prof. Dr. Mircea V. Diudea for valuable discussions and help in preparing the manuscript.

R E F E R E N C E S

1. M.V. Diudea, "*Fenotiazine şi medicamente structural înrudite*", PhD Thesis, Inst. Chem. Cluj, **1979**.
2. S.M. Free, Jr, J.W. Wilson, *J. Med. Chem.*, **1964**, *7*, 395.
3. H. Kubinyi, "*QSAR: Hansch analysis and related approaches*", R. Mannhold, Krogsgaard-Larsen P., Timmerman H. (Eds.),VCH Publishers, New York, **1993**.
4. M.V. Diudea, (Ed.), *QSPR/QSAR "Studies by Molecular Descriptors"*, NOVA, New York, **2001**.
5. M.V. Diudea, *MATCH Commun. Math. Comput. Chem*., **1997**, 35, 169.
6. M.V. Diudea, *J. Chem. Inf. Comput. Sci*., **1997**, *37*, 300.
7. O. Ursu, M.V. Diudea, TOPOCLUJ software program, Babes-Bolyai University, Cluj, 2005.
8. C.D. Moldovan, A. Costescu, G. Katona and M.V. Diudea, *MATCH Commun*. *Math*. *Comput*. *Chem.*, **2008**, *60*, 977.
9. A.T. Balaban, A. Chiriac, I. Motoc, and Z. Simon, "*Steric Fit in QSAR"* (*Lectures Notes in Chemistry*, Vol. 15), Springer, Berlin, **1980**.
10. A.A.Toropov, A.P.Toropova, *Internet El. J. Molec. Design*, **2002**, *1*, 108.
11. A.A.Toropov, A.P.Toropova, *J. Mol. Struct*. (*Theochem*), **2001**, *538*, 287.
12. P. Willett, J.M. Barnard, G.M. Downs, *J. Chem. Inf. Comput. Sci.*, **1998**, *38(6),* 983.
13. Gaussian 09, Gaussian Inc Wallingford CT, Revision A.1 (2009) Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery JA, Peralta JE, Ogliaro F, Bearpark M, Heyd JJ, Brothers E, Kudin KN, Staroverov VN, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam NJ, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas Ö, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ.
14. L. Jäntschi, LOO Analysis (LOO: leave one out), Academic Direct Library of Software, **2005**, Available at: *http://l.academicdirect.org/Chemistry/SARs/SARs/loo/*
15. J.G. Topliss*,* R.J. Costello, *J. Med*. *Chem*., **1972**, *15*, 1066.