ACADEMIA ROMÂNĂ
Revue Roumaine de Chimie
http://web.icf.ro/rrch/

Rev. Roum. Chim.,
2015, *60*(7-8), 665-676

Dedicated to Professor Valer Farcasan on the occasion of his 95th anniversary

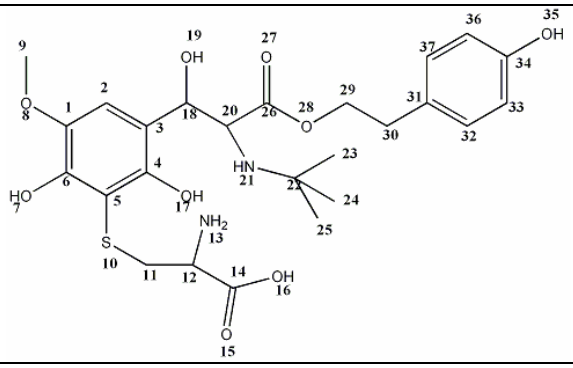
PREDICTION OF LIPOPHILICITY OF CATECHOLAMINE RELATED COMPOUNDS BASED ON THE HYPERMOLECULE CONCEPT

Ioana Anamaria SIMA, Mircea V. DIUDEA and Costel SÂRBU*

Babeş-Bolyai University, Faculty of Chemistry and Chemical Engineering, 11 Arany Janos str., RO-400028, Cluj-Napoca, Roumania

Received October 20, 2014

A set of 38 catecholamine related compounds was submitted to a novel QSPR approach based on weighting and alignment of the molecules over a hypermolecule. Any QSPR approach assumes that a macroscopic property of a chemical compound depends on the molecular structure or the topological indices, which are derived from the molecular topology or geometry. In this study the indices were calculated using TOPOCluj software and these were adjacency, connectivity, detour, distance, IE[CjMax], IE[CjMin], IP[CjMax], IP[CjMin] and Randic. Further, the calculated indices were correlated with the lipophilicity coefficient, logP, using multivariate regression and genetic algorithms.



INTRODUCTION

Catecholamines are a group of biological active amines, which contain in their structure a catechol (benzene with two hydroxyl side groups) and a side-chain amine. The catecholamines comprise the endogenous substances dopamine, noradrenaline (norepinephrine) and adrenaline (epinephrine) as well as numerous artificially synthesized compounds such as isoprenaline.[1] Their investigation constitutes a prominent chapter in the history of physiology, biochemistry and pharmacology.

All catecholamines are synthesized from the amino acid L-tyrosine according to the following sequence: tyrosine → dopa (dihydroxyphenylalanine) → dopamine → norepinephrine (noradrenaline) → epinephrine (adrenaline).[2] Catecholamines are synthesized in the brain, in the adrenal medulla, and by some sympathetic nerve fibres, playing a key role in nutrient metabolism and the generation of body heat (thermogenesis).[3] They stimulate not only oxygen consumption but also consumption of fuels, such as glucose and free fatty acids, thereby generating heat. They stimulate glycogenolysis and the breakdown of triglycerides, the stored form of fat, to free fatty acids (lipolysis).[2] They also have a role in the regulation of secretion of multiple hormones. For example, dopamine inhibits prolactin secretion, norepinephrine stimulates gonadotropin-releasing hormone secretion, and epinephrine inhibits insulin secretion by the beta cells of the islets of Langerhans of the pancreas.[3]

* Corresponding author: csarbu@chem.ubbcluj.ro

Quantitative structure-activity relationship (QSAR) is a powerful method for the design of bioactive compounds and the prediction of corresponding activity with physical and chemical properties.[4] It describes how a known biological activity can differ as a function of molecular descriptors derived from the chemical structure of a set of molecules. Many physiological activities of a molecule can be associated with their composition and structure. Molecular descriptors, which are numerical depictions of the molecular structures, are used for performing QSAR analysis.[5] Nowadays Quantitative structure-property relationship (QSPR) is also well established and correlates varied proprieties (such as lipophilicity, expressed as logP) of a compound with its molecular structure, trough a variety of descriptors (topological indices, which are derived from the molecular topology or geometry). In this study the topological indices used were the Cluj indices and they were defined by Diudea.[6, 7]

The lipophilicity is a property often used in strategies proposed to enhance the passive internalization of drugs into cells.[8] In fact, lipophilicity is an important endpoint used extensively in medicinal chemistry and environmental toxicology in predicting biological and hazardous effects of chemicals.[9] The lipophilicity is experimentally determined as partition coefficient (log P) between two immiscible phases (usually octanol-water), but it may be also computationally expressed.[10] Furthermore, it is a major experimental and theoretical tool in drug design. The lipophilicity of a solute controls its distribution among body fluids, liquid-rich phases, and tissue proteins.[11, 12]

In view of the above mentioned, in this study a set of 38 catecholamines and related compounds was submitted to a novel QSPR approach based on weighting and alignment of the molecules over a hypermolecule, and prediction of lipophilicity using the Cluj-topological indices.

DATA SET AND CORRELATING ALGORITM

A series of 38 catecholamines and related compounds (Table 1) was divided randomly in two groups, 28 molecules for the training set (molecules no.: 1, 2, 4-6, 8-11, 13, 14, 16-18, 20, 22-25, 27, 28, 30-32) and 10 molecules for the test set (molecules no.: 3, 7, 12, 15, 19, 21, 26, 29, 33), in order to develop the QSPR model and to test its applicability.

Table 1

Molecular structures of the investigated compounds

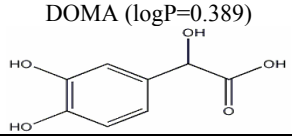
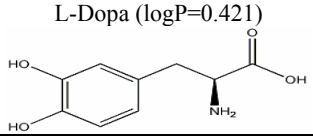
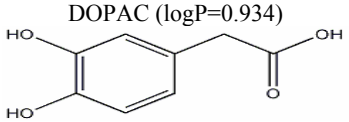
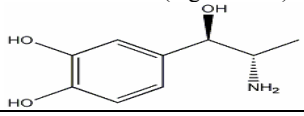
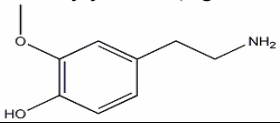
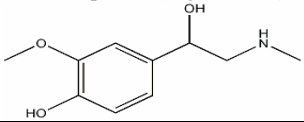
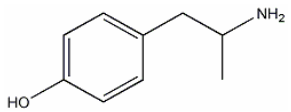
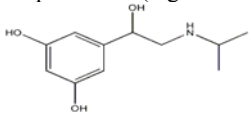
No.	Compound	No.	Compound
1	DOMA (logP=0.389) 	**20**	L-Dopa (logP=0.421) 
2	DOPAC (logP=0.934) 	**21**	Levonordefrin (logP=0.268) 
3	3-Methoxytyramine (logP=0.973) 	**22**	Metanephrine (logP=0.573) 
4	4-Hydroxyamphetamine (logP=1.367) 	**23**	Metaprotenerol (logP=1.049) 

Table 1 (continued)

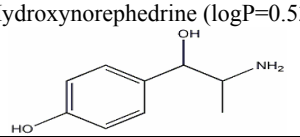
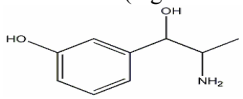
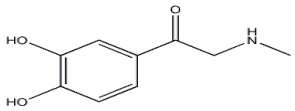
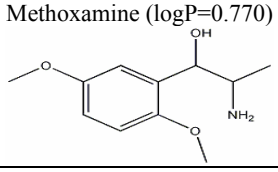
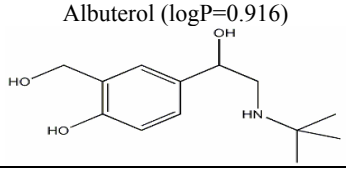
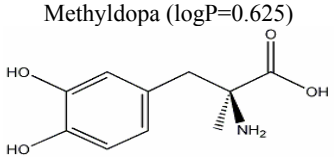
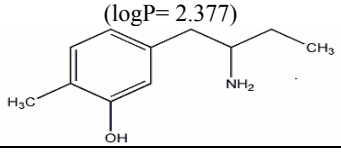
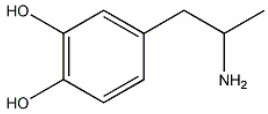
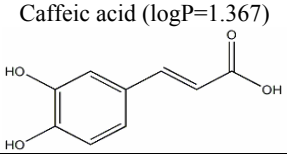
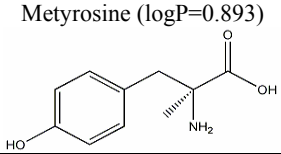
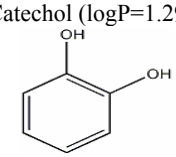
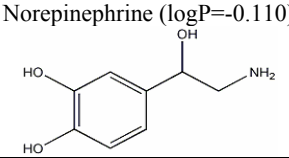
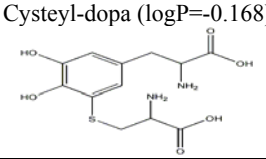
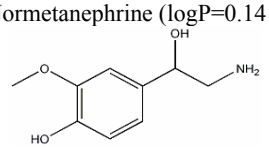
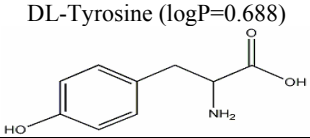
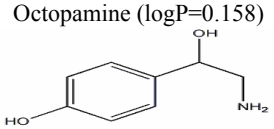
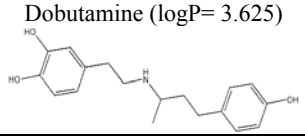
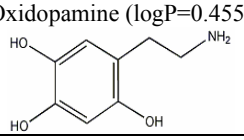
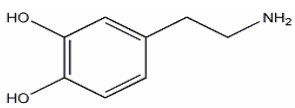
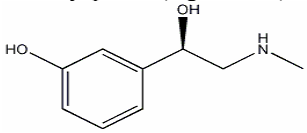
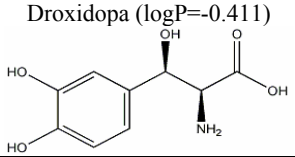
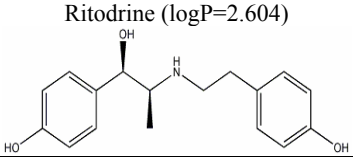
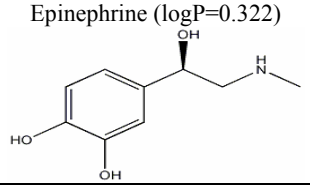
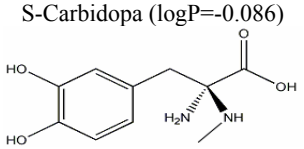
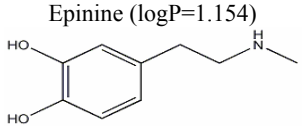
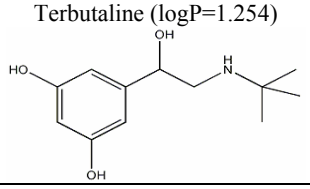
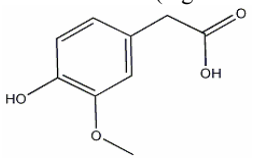
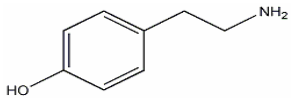
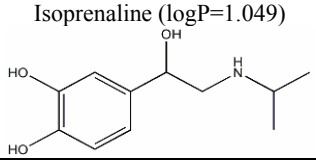
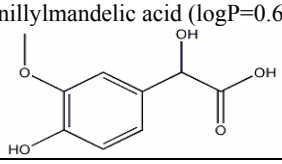
5	4-Hydroxynorephedrine (logP=0.535) 	**24**	Metaraminol (logP=0.535) 
6	Adrenalone (logP=0.719) 	**25**	Methoxamine (logP=0.770) 
7	Albuterol (logP=0.916) 	**26**	Methyldopa (logP=0.625) 
8	Alpha-ethyl-3-hydroxy-4-methylphenethylamine (logP= 2.377) 	**27**	Methyldopamine (logP=1.100) 
9	Caffeic acid (logP=1.367) 	**28**	Metyrosine (logP=0.893) 
10	Catechol (logP=1.295) 	**29**	Norepinephrine (logP=-0.110) 
11	Cysteyl-dopa (logP=-0.168) 	**30**	Normetanephrine (logP=0.141) 
12	DL-Tyrosine (logP=0.688) 	**31**	Octopamine (logP=0.158) 
13	Dobutamine (logP= 3.625) 	**32**	Oxidopamine (logP=0.455) 
14	Dopamine (logP=0.722) 	**33**	Phenyephrine (logP=0.590) 
15	Droxidopa (logP=-0.411) 	**34**	Ritodrine (logP=2.604) 

Table 1 (continued)

16	Epinephrine (logP=0.322) 	**35**	S-Carbidopa (logP=-0.086) 
17	Epinine (logP=1.154) 	**36**	Terbutaline (logP=1.254) 
18	Homovanillic acid (logP=1.185) 	**37**	Tyramine (logP=0.990) 
19	Isoprenaline (logP=1.049) 	**38**	Vanillylmandelic acid (logP=0.640) 

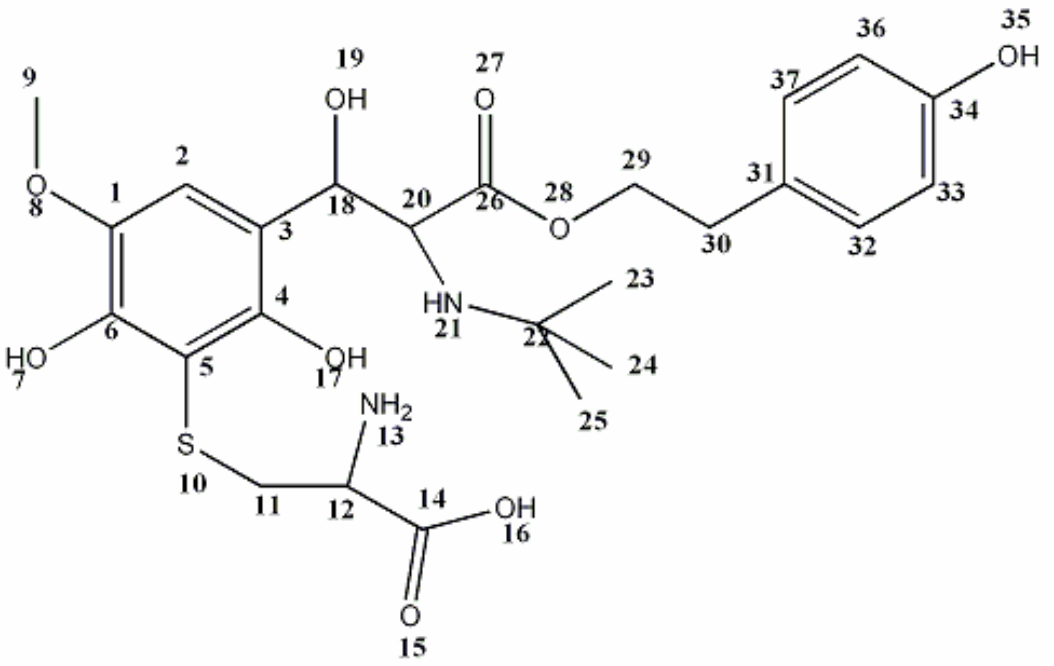


Fig. 1 – The hypermolecule.

First the molecules were drawn in HyperChem and each of them was optimized at molecular mechanics (MM+) level of theory. The correlating algorithm followed a few steps: (1) generate the hypermolecule; (2) calculate the molecular descriptors; (3) find the best regression equations by correlating the topological indices with the chosen property (logP) and (4) test the predictive capability of the model.

In order to achieve the model, the structure is encoded in a numerical form. The arrangement of substituent groups, on the catecholamine derivatives, can be accounted for by the hypermolecule concept[13] viewed as the union of

the molecules forming the correlating space. A binary vector was assigned to each molecule by aligning them over the hypermolecule (Fig. 1): 1- for a common feature in a given position of the hypermolecule and 0- for an empty position (Table 2). Next, the binary vector was weighted by the mass of "hydride" fragments composing each molecule and the weighted vector (Table 3) was used in the data-reduction step and correlation weighting procedure,[9] which are described in detail in recent papers from literature.[14-17]

Then the topological indices were calculated using TOPOCLUJ software[18] and the following descriptors are procured into consideration for developing the model: sumative descriptor (SD), adjacency, connectivity, detour, distance, DS, IEmax, IEmin, IPmax, IPmin and Randic (Table 4).

Table 2

The binary vectors of the investigated molecules

Compound	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1
2	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	1
3	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	0	1
4	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	1
5	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	1	1
6	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1
7	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	1	1
8	1	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	1	0	1
9	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	1
10	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
11	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	0	1	0	1
12	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	1
13	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	1
14	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	1
15	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1
16	1	1	1	1	1	1	1	0	0	1	0	0	0	0	0	0	0	1	1	1
17	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	1
18	1	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	1	0	1
19	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1
20	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	1
21	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1
22	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1
23	1	1	1	1	1	1	0	1	0	1	0	0	0	0	0	0	0	1	1	1
24	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	1	1	1
25	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	0	1	1	1	1
26	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	1
27	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	1
28	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	1
29	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1
30	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1
31	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	1	1
32	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	0	1
33	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	1	1	1
34	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	1	1
35	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	1
36	1	1	1	1	1	1	0	1	0	1	0	0	0	0	0	0	0	1	1	1
37	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	1
38	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1

Table 3

Correlation weighted descriptors

Compound	H7	H8	H9	H10	H11	H16	H17	H19	H21	H22	H24	H25	H26	H27	H29	H35	SD
1	-0.501	-0.253	0	0	0	0	0	-0.874	0	0	0	0	0	0	0	0	-1.628
2	-0.501	-0.253	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.754
3	-0.501	-0.253	0.284	0	0	0	0	0	-0.498	0	0	0	0	0	0	0	-0.967
4	-0.501	0	0	0	0	0	0	0	-0.498	0	0	0	0.359	0	0	0	-0.640
5	-0.501	0	0	0	0	0	0	-0.874	-0.498	0	0	0	0.359	0	0	0	-1.514
6	-0.501	-0.253	0	0	0	0	0	-0.874	-0.498	0.192	0	0	0	0	0	0	-1.934
7	-0.501	0	0	0	0	0	0	-0.874	-0.498	0.192	0.249	0.258	0	0	0	0	-1.175
8	0	0	0	0.195	0	0	0	0	-0.498	0	0	0	0.359	0	0	0	0.056
9	-0.501	-0.253	0	0	0	0	0	0	0	0	0	0	0.359	-0.692	0	0	-1.086
10	-0.501	-0.253	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.754
11	-0.501	-0.253	0	0	-0.547	-0.074	0	0	-0.498	0	0	0	0.359	-0.692	0	0	-2.206
12	-0.501	0	0	0	0	0	0	0	-0.498	0	0	0	0.359	-0.692	0	0	-1.332
13	-0.501	-0.253	0	0	0	0	0	0	0	0	0	0	0	0	0.696	1.445	1.387
14	-0.501	-0.253	0	0	0	0	0	0	-0.498	0	0	0	0	0	0	0	-1.252
15	-0.501	-0.253	0	0	0	0	0	-0.874	-0.498	0	0	0	0.359	-0.692	0	0	-2.459
16	-0.501	0	0	0.195	0	0	0	-0.874	-0.498	0.192	0	0	0	0	0	0	-1.486
17	-0.501	-0.253	0	0	0	0	0	0	-0.498	0.192	0	0	0	0	0	0	-1.060
18	-0.501	0	0	0.195	-0.547	0	0	0	0	0	0	0	0	0	0	0	-0.853
19	-0.501	-0.253	0	0	0	0	0	-0.874	-0.498	0.192	0.249	0	0	0	0.696	0	-0.989
20	-0.501	-0.253	0	0	0	0	0	0	-0.498	0	0	0	0.359	-0.692	0	0	-1.584
21	-0.501	-0.253	0	0	0	0	0	-0.874	-0.498	0	0	0	0.359	0	0	0	-1.767
22	-0.501	-0.253	0.284	0	0	0	0	-0.874	-0.498	0.192	0	0	0	0	0	0	-1.650
23	0	-0.253	0	0.195	0	0	0	-0.874	-0.498	0.192	0.249	0	0	0	0	0	-0.989
24	0	-0.253	0	0	0	0	0	-0.874	-0.498	0	0	0	0.359	0	0	0	-1.266
25	0	-0.253	0.284	0	0	0	-0.309	-0.874	-0.498	0	0	0	0.359	0	0	0	-1.290
26	-0.501	-0.253	0	0	0	0	0	0	-0.498	0	0	0	0.359	-0.692	0	0	-1.584
27	-0.501	-0.253	0	0	0	0	0	0	-0.498	0	0	0	0.359	0	0	0	-0.893
28	-0.501	0	0	0	0	0	0	0	-0.498	0	0	0	0.359	-0.692	0	0	-1.332
29	-0.501	-0.253	0	0	0	0	0	-0.874	-0.498	0	0	0	0	0	0	0	-2.126
30	-0.501	-0.253	0.284	0	0	0	0	-0.874	-0.498	0	0	0	0	0	0	0	-1.842
31	-0.501	0	0	0	0	0	0	-0.874	-0.498	0	0	0	0	0	0	0	-1.874
32	-0.501	-0.253	0	0	0	0	-0.309	0	-0.498	0	0	0	0	0	0	0	-1.561
33	0	-0.253	0	0	0	0	0	-0.874	-0.498	0.192	0	0	0	0	0	0	-1.433
34	-0.501	0	0	0	0	0	0	-0.874	0	0	0	0	0	0	0.696	1.445	0.766
35	-0.501	-0.253	0	0	0	0	0	0	-0.498	0.192	0	0	0.359	-0.692	0	0	-1.392
36	0	-0.253	0	0.195	0	0	0	-0.874	-0.498	0.192	0.249	0.258	0	0	0	0	-0.731
37	-0.501	0	0	0	0	0	0	0	-0.498	0	0	0	0	0	0	0	-0.999
38	-0.501	-0.253	0.284	0	0	0	0	-0.874	0	0	0	0	0	0	0	0	-1.344

Table 4

Topological descriptors computed using TOPOCLUJ software

Compound	LogP	SD	Adjacency	Connectivity	Detour	Distance	DS	IE[CjMax]	IE[CjMin]	IP[CjMax]	IP[CjMin]	Randic
1	0.389	-1.628	13	17	380	250	9.9	76	180	260	510	6
2	0.934	-0.754	12	16	330	210	9.2	61	150	210	430	5.6
3	0.973	-0.967	12	15	330	200	9.1	60	150	210	420	5.8
4	1.367	-0.640	11	14	260	170	8.4	49	130	160	340	5.2
5	0.535	-1.514	12	15	310	200	9.1	62	150	200	410	5.6
6	0.719	-1.934	13	17	390	260	9.9	81	190	280	540	6.1
7	0.916	-1.175	17	20	770	560	13	200	370	810	1400	7.8
8	2.377	0.056	13	16	400	260	9.9	84	190	300	570	6.1
9	1.367	-1.086	13	18	400	270	10	88	190	320	600	6.1
10	1.295	-0.754	8	11	120	60	6.2	13	53	41	100	3.8
11	-0.168	-2.206	21	26	1400	1000	16	380	650	1700	2700	9.7
12	0.688	-1.332	13	17	390	270	9.9	90	190	310	580	6.1
13	3.625	1.387	23	29	1800	1300	18	480	900	2700	5000	11
14	0.722	-1.252	11	14	270	160	8.4	45	120	150	320	5.2
15	-0.411	-2.459	15	19	530	370	11	130	260	440	800	6.9
16	0.322	-1.486	13	16	390	260	9.8	81	190	280	540	6.1
17	1.154	-1.060	12	15	330	210	9.1	66	160	240	460	5.7
18	1.185	-0.853	13	17	400	260	9.9	79	190	290	550	6.1
19	1.049	-0.989	15	18	560	400	11	140	270	530	920	7
20	0.421	-1.584	14	18	470	320	11	110	230	390	710	6.5
21	0.268	-1.767	13	16	380	250	9.8	76	180	260	510	6
22	0.573	-1.650	14	17	470	310	11	100	230	370	690	6.7
23	1.049	-0.989	15	18	570	390	11	140	270	530	880	7
24	0.535	-1.266	12	15	320	200	9.1	62	140	210	390	5.6
25	0.770	-1.290	15	18	560	360	11	120	250	440	750	7.1

Table 4 (continued)

26	0.625	-1.584	15	19	540	370	11	130	260	450	820	6.8
27	1.100	-0.893	12	15	330	210	9.1	61	150	210	430	5.6
28	0.893	-1.332	14	18	450	320	11	110	220	360	670	6.4
29	-0.110	-2.126	12	15	320	200	9.1	58	150	190	400	5.6
30	0.141	-1.842	13	16	390	250	9.8	75	180	270	520	6.2
31	0.158	-1.874	11	14	260	160	8.4	46	120	150	320	5.2
32	0.455	-1.561	12	15	330	200	9.1	57	150	190	390	5.6
33	0.590	-1.433	12	15	320	210	9.1	67	150	230	420	5.7
34	2.604	0.766	22	28	1500	1100	17	390	760	2000	3800	10
35	-0.086	-1.392	16	20	620	440	12	160	300	550	980	7.4
36	1.254	-0.731	16	19	660	470	12	170	310	650	1100	7.3
37	0.990	-0.999	10	13	210	130	7.7	35	100	120	250	4.8
38	0.640	-1.344	14	18	460	300	11	96	220	340	650	6.6

Table 5

Best models in describing log P for the training set of molecules

No.	Descriptors	R^2	Adjust. R^2	St. Error	F
1	***SD***	***0.9118***	***0.9084***	***0.2477***	***268.74***
2	IP min	0.3611	0.3365	0.6666	14.696
3	IP max	0.3145	0.2881	0.6905	11.926
4	IE min	0.2696	0.2415	0.7128	9.5972
5	***SD*, *Adjacency***	***0.9121***	***0.9050***	***0.2522***	***129.64***
6	SD, Randic	0.9119	0.9048	0.2525	129.36
7	SD, DS	0.9119	0.9048	0.2525	129.35
8	SD, IE max	0.9119	0.9048	0.2525	129.35
9	SD, Distance	0.9119	0.9048	0.2525	129.33
10	SD, Connectivity	0.9118	0.9048	0.2525	129.30
11	SD, Detour	0.9118	0.9048	0.2525	129.29
12	***SD*, *Adjacency*, *Randic***	***0.9198***	***0.9098***	***0.2458***	***91.799***
13	SD, Adjacency, Distance	0.9188	0.9087	0.2473	90.544
14	SD, IE max, IP max	0.9178	0.9076	0.2488	89.356
15	SD, Distance, IP max	0.9176	0.9073	0.2492	89.059
16	SD, Detour, IP max	0.9163	0.9059	0.2511	87.595
17	SD, IE min, IP max	0.9148	0.9041	0.2533	85.916
18	SD, IP max, Randic	0.9129	0.9020	0.2562	83.872
19	SD, IP min, Randic	0.9129	0.9020	0.2562	83.826
20	SD, DS, IP max	0.9129	0.9020	0.2562	83.823
21	SD, Connectivity	0.9125	0.9016	0.2568	83.421
22	***SD*, *Adjacency*, *Randic*, *DS***	***0.9250***	***0.9119***	***0.2429***	***70.894***
23	SD, Adjacency, Randic, Connectivity	0.9242	0.9110	0.2441	70.128
24	SD, IP max, Distance, DS	0.9228	0.9094	0.2463	68.776
25	SD, IP max, Distance, Randic	0.9205	0.9066	0.2501	66.545

RESULTS AND DISCUSSION

1. QSPR models – by multivariate regression

The models were developed using the molecules from the training set and the best results are listed below (Table 5).

At first, only one-dimensional models were selected. Then the most accurate model was picked up. After that, all two-dimensional models produced by adding a new attribute to the first model were evaluated. Again the best model was chosen and supplemented by a new attribute. All the considered regression models must pass a test based on the value of Fisher statistics of all their regression determination coefficients (R^2).

If a term has the value of R^2 for the regression equation less than a specified threshold, this term is removed from the model. Thus, the process of adding new terms stops either when all attributes are included in the model or when no new term can be added without violating this criterion. The best equation is produced by the system based on determination coefficient (R^2) and the best fit regression is that in which the values for R^2 are closer to 1.

1.1. Model validation – external validation

For the applicability of the model, a step of external validation was performed on the best model obtained. Thus, the test molecules were submitted to the model in order to calculate log P using equation no. 22 from Table 5. The results are presented in Table 6, and Figure 2 represents the correlation log P-pred vs. log P-exp.

Table 6

Predicted log P for the test set of molecules using external validation

Compound	LogP- Pred	LogP-Exp
3	1.159	0.973
7	0.774	0.916
12	0.711	0.688
15	-0.598	-0.411
19	0.892	1.049
21	0.193	0.268
26	0.185	0.625
29	-0.094	-0.110
33	0.643	0.590
38	0.826	0.640

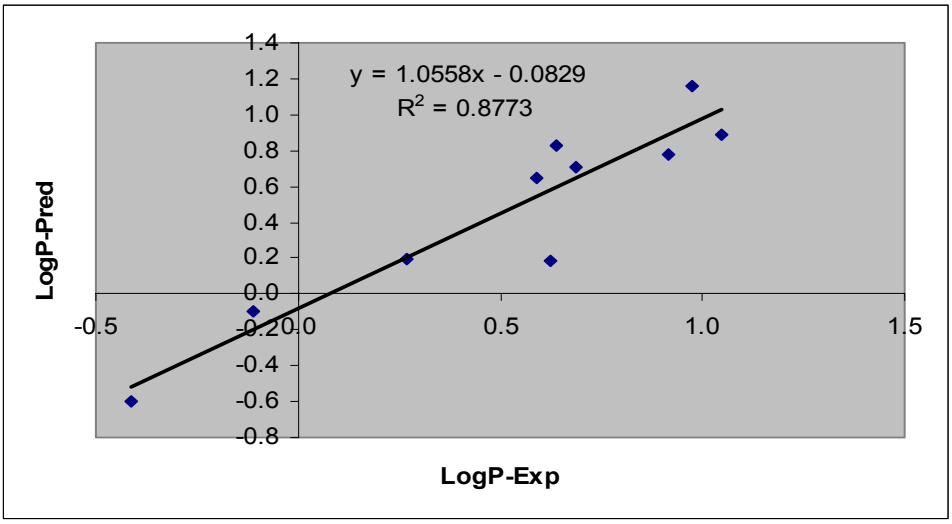


Fig. 2 – Plot of the regression log P-pred vs. log P-exp.

1.2. Model validation – similarity cluster

The similarity cluster validation was made using the similarity analysis from the Topocluj molecular topology software. The analysis consisted of finding similarities between target molecules, which in this case are the test molecules, and the molecules from the training set. Then for each molecule in the test set a regression equation was obtained using the topological indices of the molecules in the training set which had a degree of similarity grater than 70% with the respective molecule. The equations thus obtained were used to predict the values of log P for each molecule from the test set and data are listed in Table 7. It can be observed in Fig. 3, that the determination coefficient of log P-pred vs. log P-exp is far better in this case than in the external validation.

Table 7

Predicted log P for the test set using similarity cluster validation

Compound	LogP- Pred	LogP-Exp
3	1.006	0.973
7	0.744	0.916
12	0.569	0.688
15	-0.659	-0.411
19	0.971	1.049
21	0.267	0.268
26	0.359	0.625
29	-0.068	-0.110
33	0.599	0.590
38	0.820	0.640

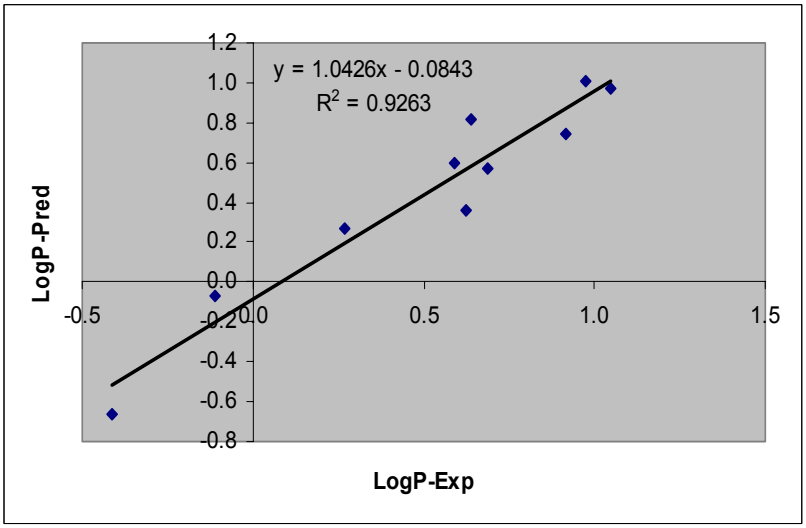


Fig. 3 – Plot of the regression logP-pred vs. logP-exp.

Table 8

Best models in describing log P for the training set of molecules

No.	Descriptors	R^2	Adjust. R^2	St. Error	F
1	SD, Adjacency, Randic	0.9198	0.9098	0.246	91.77
2	***SD*, *Adjacency*, *DS*, *Randic***	***0.9250***	***0.9119***	***0.243***	***70.88***
3	SD, Adjacency, IE max, Randic	0.9198	0.9059	0.251	65.97
4	SD, Adjacency, Distance, Randic	0.9199	0.9059	0.251	66.00
5	SD, Adjacency, IEmin, Randic	0.9200	0.9061	0.251	66.11
6	SD, Adjacency, IPmax, Randic	0.9199	0.9060	0.251	66.08
7	SD, Adjacency, Detour, Randic	0.9198	0.9059	0.251	65.99
8	SD, Adjacency, IPmin, Randic	0.9199	0.9060	0.251	66.05
9	SD	0.9118	0.9084	0.248	268.62
10	SD, Adjacency, Conectivity, Randic	0.9242	0.9110	0.244	70.11

2. QSAR models – by genetic algorithms

Genetic algorithms (GA) are an evolutionary method widely used for complex optimisation problems in several fields such as robotics, chemistry and QSAR/QSPR.[19, 20] Using the same topological indices (Table 4) QSPR models were generated for describing log P for the training set of molecules, using MobiDigs software, which allows searching for regression models by developing optimal model populations using genetic algorithms. The best obtained models are listed in Table 8.

2.1. Model validation – by external validation

The best obtained model was submitted to external validation in order to test its applicability. Thus, for the test molecules log P was predicted using equation no. 2 from Table 9. The results are presented in Table 9, and Figure 4 represents the correlation log P-pred vs. log P-exp.

Table 9

Predicted log P
for the test set of molecules using external validation

Compound	LogP- Pred	LogP-Exp
3	1.194	0.973
7	0.766	0.916
12	0.716	0.688
15	-0.491	-0.411
19	1.039	1.049
21	0.212	0.268
26	0.305	0.625
29	-0.100	-0.110
33	0.660	0.590
38	0.731	0.640

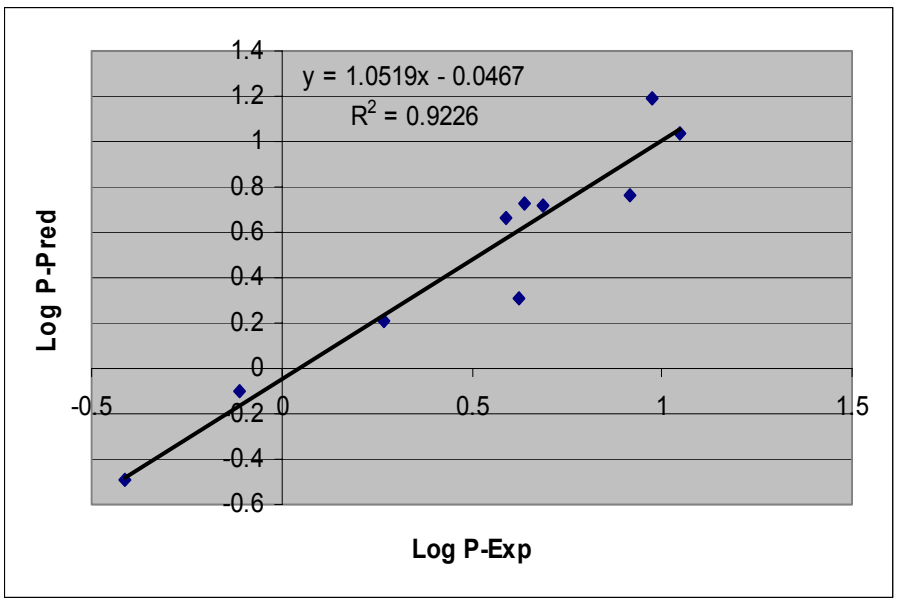


Fig. 4 – Plot of the regression log P-pred vs. log P-exp.

CONCLUSIONS

A set of thirty eight catecholamines and related compounds was submitted to a novel QSPR approach based on the alignment of all structures over a hypermolecule, thus obtaining a powerful topological descriptor, the summative descriptor (SD), for the prediction of lipophilicity (log P). The set of molecules was divided in two groups, the first group (training set) was used to develop the QSPR models by multivariate regression and also by genetic algorithms, and the second group (test set) was used to validate the obtained models.

The results indicate that the QSPR model obtained using multivariate regression has good predictive capacity in case of external validation but in case of validation by similarity clusters the results were significantly improved, from a determination coefficient of 0.8773 in the first case, to 0.9263 in the second case, respectively. Also the QSPR model obtained using genetic algorithms provided similar results, with a coefficient of correlation of 0.9226, thus supporting the idea that the new QSPR approach is of great use in predicting the lipophilicity of catecholamine related compounds.

Acknowledgements: This work was possible due to the financial support of the Sectorial Operational Program for Human Resources Development 2007-2013, co-financed by the European Social Fund, under the project number

POSDRU/159/1.5/S/132400 with the title "Young successful researchers – professional development in an international and interdisciplinary environment" and by Roumanian Ministry of Education, Research, Youth and

Sport through research project PN-II-ID-PCE-2011-3-0366.

REFERENCES

1. R.S. Chen, W.H. Huang, H. Tong, Z.L. Wang and J.K. Cheng, *Anal. Chem*., **2003**, *75*, 6341-6345.
2. H.E. Himwich and W.A. Himwich, "Biogenic amines", Elsevier, Amsterdam, 1964.
3. S. Shioda, I. Homma and N. Kato, "Transmitters and modulators in health and disease", Springer, Tokyo, 2009.
4. D.H. Cho, S.K. Lee, B.T. Kim and K. Tai, *Bull. Korean Chem. Soc*., **2001**, *22*, 388-394.
5. P.P. Latha and J.S. Sharmila, *Int. J. Chem. Res.*, **2010**, *2*, 32-38.
6. M.V. Diudea, *MATCH Commun. Math. Comput. Chem.*, **1997**, *35*, 169-183.
7. M.V. Diudea, *J. Chem. Inf. Compu. Sci.*, **1997**, *37*, 300-305.
8. R. Pignatello, S. Guccione, S. Forte, C. Di Giacomo, V. Sorrenti, L. Vicari, G.U. Barretta, F. Balzanoc and G. Puglisi, *Bioorg. Med. Chem*., **2004**, *12*, 2951-2964.
9. A.A. Toropov and A. P. Toropova, *J. Mol. Struc.-Theochem.*, **2001**, *538*, 197-199.
10. I.V. Tetko, J. Gasteiger, R. Todeschini, A. Mauri, D. Livingstone, P. Ertl, V.A. Palyulin, E.V. Radchenko, N.S. Zefirov, A.S. Makarenko, V.Y. Tanchuk and V.V. Prokopenko, *J. Comput. Aid. Mol. Des.*, **2005**, *19*, 453-463.
11. A. Testa, P.A. Carrupt, P. Gaillard, F. Billois and P. Weber, *Pharm. Res.*, **1996**, *13*, 335-343.
12. M.J. Waring, *Expert Opin. Drug Discov.*, **2010**, *5*, 235-248.
13. A.T. Balaban, A. Chiriac, I. Motoc and Z. Simon, "Steric Fit in QSAR (Lectures Notes in Chemistry, Vol. 15)", Springer, Berlin, 1980.
14. A.M. Harsa, *Studia UBB Chemia*, **2014**, *1*, 111-123.
15. T.E. Harsa, *Studia UBB Chemia*, **2014**, *1*, 99-110.
16. R. Maties, B. Szefler, I. Ionut and B. Tiperciuc, *Studia UBB Chemia*, **2012**, *4*, 121-133.
17. T.E. Harsa, A.M. Harsa and B. Szefler, *Cent. Eur. J. Chem.*, **2014**, *12*, 365-376.
18. O. Ursu and M.V. Diudea, TOPOCLUJ software program, Babeş-Bolyai University, Cluj, 2005.
19. D.E. Goldberg, "Genetic Algorithms in Search, Optimization and Machine Learning", Addison-Wesley, Massachusetts, 1989.
20. R. Wehrens and L.M.C. Buydens, *Trends Anal. Chem.*, **1998**, *17*, 193-203.