STUDIA UBB CHEMIA, LX, 4, 2015 (p. 185-192)
(RECOMMENDED CITATION)

QSAR STUDY ON SEROTONIN DERIVATIVES

SARA ERSALI[a] AND MIRCEA V. DIUDEA[a*]

ABSTRACT. A set of 40 serotonin derivatives, downloaded from the PubChem database, was submitted to a QSAR study, following Diudea's algorithm, in the frame of a hypermolecule, which mimics the „alignment" of drug molecules to the biological receptors. The best models describing log P of this set of serotonins were validated by the leave-one-out procedure, in the external test set and in a new version of prediction by using clusters of similar molecules. The structures have been optimized at PM3 level of theory. Topological indices have been computed by TOPOCLUJ software program.

Keywords: *Serotonin, QSAR, Hypermolecule, log P.*

INTRODUCTION

Serotonin or 5-hydroxytryptamine (5-HT) is a monoamine neuro-transmitter. Biochemically derived from tryptophan, serotonin is primarily found in the gastrointestinal tract, blood platelets, and the central nervous system (CNS) of animals, including humans. It is popularly thought to be a contributor to feelings of well-being and happiness [1].

Approximately 90% of the human body's total serotonin is located in the enterocheromaffin intestine cells, where it is used to regulate intestinal movements [2,3]. It is also synthesized in serotonergic neurons of the CNS, where it has various functions, including the regulation of mood, appetite, and sleep. Serotonin also has some cognitive functions, including memory and learning. Modulation of serotonin at synapses is thought to be realized by several classes of antidepressants. Serotonin secreted from the enterocheromaffin cells eventually arrives in the blood, acting on blood *platelets*, which store it. When the platelets bind to a clot, they release serotonin, where it serves as a vasoconstrictor and helps to regulate hemostasis and blood clotting. Serotonin also is a growth factor for some types of cells, possibly involved in wound healing. There are various serotonin receptors.

[a] *Babeş-Bolyai University, Faculty of Chemistry and Chemical Engineering, 11 Arany Janos Str., RO-400028, Cluj-Napoca, Romania.*
** Corresponding author: diudea@chem.ubbcluj.ro*

Serotonin is metabolized mainly to 5-HIAA, in the liver; metabolism starts with oxidation of serotonin by monoamine oxidase to the corresponding aldehyde. This is followed by oxidation by aldehyde dehydrogenase to 5-HIAA, an indole acetic acid derivative and finally excreted by the kidneys. Because of serotonin's growth-promoting effect on cardiac myocytes [4], it could be involved in hearth cancer.

In addition to animals, serotonin is found in fungi and plants [5]. Serotonin's presence in insect venoms and plant spines cause pain, which is a side-effect of serotonin injection. Serotonin is produced by pathogenic amoebae, and its effect on the intestine causes diarrhea. Its presence in seeds and fruits may stimulate the digestive tract into expelling the seeds.

STRUCTURAL MOLECULAR DATA

A set of 40 serotonin derivatives, taken from PubChem Database [6], were divided into a training set (the first 25 molecules) and a test set (the last 15 molecules), taken randomly; the modelled property was log P (Table 1).

On the set of 40 serotonins, a hypermolecule [7] was built up, as a reunion of their substructures (Figure 1).

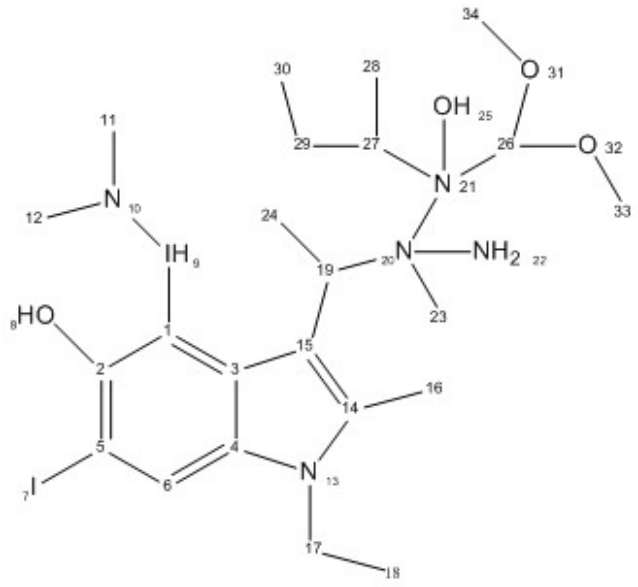

Figure 1. The hypermolecule built on 40 serotonins of the dataset

Table 1. Molecular structures (in SMILES code) of Serotonin derivatives with their log P and CID (taken from PubChem)

Mol.	Canonical SMILES	log P	CID
1	C1=CC2=C(C=C1O)C(=CN2)CCN	0.2	5202
4	C1=CC2=C(C=C1O)C(=CN2)CCCN	0.6	19438874
5	CC(C)CC1=CNC2=C1C=C(C=C2)O	2.5	69951223
6	CCC(C)C1=CNC2=C1C=C(C=C2)O	2.6	69950621
7	CC(CNC)C1=CNC2=C1C=C(C=C2)O	1.1	57825670
8	C1=CC(=C(C2=C1NC=C2CCCN)I)O	2	57620180
10	CC(CC1=CNC2=C1C=C(C=C2)O)NC	1.2	23531057
14	CC(C)(CC1=CNC2=C1C=C(C=C2)O)N	0.8	15137728
16	CC(CN)C1=CNC2=C1C=C(C=C2)O	0.5	115292
18	CN(C)CCC1=CNC2=C1C=C(C=C2)O	1.2	10257
19	CC(CC1=CNC2=C1C=C(C=C2)O)N	0.6	2107
20	CC(C)N(CCC1=CNC2=C1C=C(C=C2)O)C(C)C	2.8	71360804
21	CC(C)(C)NCC1=CNC2=C1C=C(C=C2)O	1.2	71040304
22	CCCN(CCC)CCC1=CNC2=C1C=C(C=C2)O	3	169764
23	C1=C2C(=CC(=C1O)I)NC=C2CCCN	2	57620181
24	C1=CC2=C(C=C1O)C(=CN2)CCNO	0.2	53831394
25	CC1=C(C=CC2=C1C(=CN2)CCN)O	1.4	23373052
27	CCN1C=C(C2=C1C=CC(=C2)O)CCN	1.3	70378769
28	CNCCC1=CN(C2=C1C=C(C=C2)O)C	1.5	11447050
29	CN1C=C(C2=C1C=CC(=C2)O)CCN	1	440945
30	CC(=O)NCCC1=CNC2=C1C=C(C=C2)O	0.5	903
31	CC(CC1=CN(C2=C1C=C(C=C2)O)C)N	1.4	72523641
32	C1=CC2=C(C=C1O)C(=CN2)CCNC(=O)O	0.5	67228616
33	CC1=C(C2=C(N1)C=CC(=C2)O)CCNC	1.9	57825649
40	CC1=C(C2=C(N1)C=CC(=C2)O)CC(C)N	1.8	72523635
2	C1=CC2=C(C=C1O)C(C=N2)CCN	0.1	46228506
3	C1=CC2=C(C=C1O)C(=CN2)C=CN	0.6	50937459
9	C1=CC2=C(C=C1O)C(=CN2)CCN(F)F	1.3	54301972
11	CC(C)C1=CNC2=C1C=C(C=C2)O	2	22669491
12	CCCC1=CNC2=C1C=C(C=C2)O	2.2	15366338
13	CCC1=CNC2=C1C=C(C=C2)O	1.7	15366337
15	CC1=CNC2=C1C=C(C=C2)O	1.3	192734
17	CCN(CC)CCC1=CNC2=C1C=C(C=C2)O	1.9	26395
26	C1=C2C(=CC(=C1O)F)NC=C2CCN	0.3	194142
34	CC(C1=CNC2=C1C=C(C=C2)O)O	0.4	18615721
35	CNC(=O)NCCC1=CNC2=C1C=C(C=C2)O	0.3	18360650
36	CN(C)CC1=C(C=CC2=C1C(=CN2)CCN)O	0.8	10922434
37	CC(=NCCC1=CNC2=C1C=C(C=C2)O)C(=O)O	0.7	10043102
38	CCC(=O)NCCC1=CNC2=C1C=C(C=C2)O	0.9	594440
39	C1=C2C(=C(C(=C1F)O)F)C(=CN2)CCN	1.2	194143

COMPUTATIONAL DETAILS

The structures have been optimized at Molecular Mechanics level of theory (Table 2).

Table 2. Log P, correlating descriptor SD, and topological indices (total adjacency AD, 3D distance sum D3D and detour DE) for the set of. 40 serotonins in Table 1

Mol.	log P	SD	AD	D3D	DE
1	0.2	-19.518	14	198	568
2	0.1	-19.518	14	198	678
3	0.6	-19.592	14	198	667
4	0.6	-18.122	15	254	658
5	2.5	-17.664	15	238	772
6	2.6	-17.859	15	232	795
7	1.1	-18.807	16	291	780
8	2	-17.442	16	291	768
9	1.3	-18.539	16	305	658
10	1.2	-18.611	16	297	790
11	2	-17.785	14	184	667
12	2.2	-18.122	14	198	1318
13	1.7	-18.048	13	152	904
14	0.8	-19.039	16	279	1350
15	1.3	-19.153	12	117	794
16	0.5	-19.256	15	232	678
17	1.9	-17.939	18	441	673
18	1.2	-18.734	16	305	792
19	0.6	-19.060	15	238	795
20	2.8	-17.039	20	564	673
21	1.2	-18.544	17	355	928
22	3	-16.839	20	606	783
23	2	-18.122	16	305	928
24	0.2	-19.430	15	254	797
25	1.4	-18.839	15	230	785
26	0.3	-19.518	16	286	568
27	1.3	-18.812	16	299	568
28	1.5	-18.313	15	237	790
29	1	-18.674	17	384	558
30	0.5	-19.206	16	281	568
31	1.4	-18.377	17	384	471
32	0.5	-19.206	16	295	386
33	1.9	-17.994	15	238	1048
34	0.4	-18.889	14	184	672
35	0.3	-19.539	18	467	558
36	0.8	-18.838	18	374	1070
37	0.7	-19.054	19	543	1062
38	0.9	-19.206	18	467	1214
39	1.2	-18.838	16	277	1070
40	1.8	-17.985	16	277	786

Topological indices have beed computed by TOPOCLUJ software [8]; some of them (Conectivity =C, Total adjacency = Ad, Detour = De, Distance = Di, D3D, SD), and log P are listed in Table 2.

RESULTS AND DISCUSSION

In this paper, mass fragments are used for weighting the binary vector of ligand superposition over the statistically significant positions of the hypermolecule, according to Diudea's algorithm [9].

Data reduction

The local correlation-weighted descriptors are summed to give SD global descriptor, over the following significant positions in the hypermolecule: H1, H14, H16, H17, H20, H21, H22, H23, H24, H26, H27, H28, H29, H30, H32, H34. SD correlation with log P= 19.839+0.999XSD, R2=0.839, n=40, s=0.310, F=199.179, and the best results are listed below and in Table 3.

QSAR models

The models were performed on the training set (the first 25 structures in Table 1) and the best results are listed below and in Table 3. The number of descriptors was limited to three, to fulfil the considerations of Topliss and Costello [10].

Table 3. Best models in describing log P in the training set of serotonin in Table 1

	Descriptors	R^2	Adjust. R^2	St. Error	F
1	**SD**	0.847	0.841	0.321	128.053
2	**AD**	0.277	0.246	0.699	8.830
3	**D3D**	0.236	0.203	0.719	7.128
4	**SD, D3D**	0.8483	0.8346	0.3277	61.552
5	**SD, Di**	0.8482	0.8344	0.3278	61.466
6	**SD, CjDe**	0.8483	0.8345	0.3277	61.515
7	**SD, CfDe**	0.8482	0.8344	0.3279	61.331
8	**SD, D3D, AD**	0.8588	0.8387	0.3235	42.603
9	**SD, CjDi, CjDe**	0.8575	0.8371	0.3251	42.139
10	**SD, CjDe, CfDi**	0.8558	0.8352	0.3270	41.552
11	**SD, Di, D3D**	0.8520	0.8308	0.3313	40.298
12	**SD, De, CjDi**	0.8509	0.8296	0.3325	39.972

(i) Monovariate regression

$$\log P = 20.392 + 1.031 \times SD$$

(ii) Bivariate regression

$$\log P = 20.818 + 1.050 \times SD - 0.0003 \times D3D$$

(iii) Trivariate regression

$$\log P = 15.76 + 1.02 \times SD + 0.4 \times D3D - 0.005 \times AD$$

Model Validation

(a) Leave-one-out

The performances in leave-one-out analysis related to the models listed as the best in Table 3 are presented in Table 4 [11, 12].

Table 4. Leave-one-out analysis for best log P models (Table 3).

	Descriptors	Q^2	R^2-Q^2	St. Error$_{loo}$	F_{loo}
1	**SD**	**0.8272**	0.205	0.3421	110.134
4	**SD, D3D**	**0.8105**	0.037	0.3583	98.371
8	**SD, D3D, AD**	**0.7848**	0.074	0.3817	83.911

(b) External Validation

The values log P for the test set of serotonins (Table 1, last 15 molecules) were calculated by using the best equation (with three variables) in Table 3, entry 8. Data are listed in Table 5 and the monovariate correlation: $\log P = 0.35 \times \log P_{calc.} + 0.813$; n=15; R^2=0.664; s=0.41; F=25.785 is plotted in Figure 2.

Table 5. Calculated values of log P for the molecules in the test set (Table 1)

Mol.	log P	log $P_{calc.}$
2	0.1	0.10
3	0.6	0.02
9	1.3	1.25
11	2	1.93
12	2.2	1.52
13	1.7	1.47
15	1.3	0.16
17	1.9	1.87
26	0.3	0.24
34	0.4	0.81
35	0.3	0.10
36	0.8	1.32
37	0.7	0.55
38	0.9	0.44
39	1.2	1.10

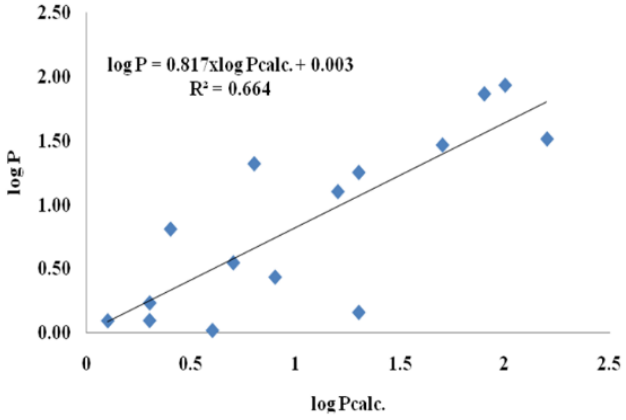

Figure 2. The plot log P vs. log $P_{calc.}$ for the test set (external validation)

Clearly, the model (Table 3, entry 8) is not a predictive one; we only used this model to select the best independent descriptors for log P in the set of studied serotonins.

(c) Similarity Cluster Validation

Validation can be performed by calculating log P for the molecules in the test set with equations learned on clusters of similarity: each of the 15 molecules is the leader in its own cluster, selected by (2D) similarity among the 25 structures of the initial learning set [13]. The values log $P_{calc.}$ for each of the

15 molecules in the test set were computed by 15 new equations (the leader being left out) with the same descriptors as in eq. 8, Table 3. Data are listed in Table 6 and the monovariate correlation: $\log P = 0.184 \times \log P_{calc.} + 1.014$; n=15; R^2=0.913; s=0.21; F=136.053 is plotted in Figure 3.

Table 6. Calculated values of log P by similarity clusters, for the molecules in the test set (Table 1)

Mol.	log P	log $P_{calc.}$
2	0.1	0.05
3	0.6	0.12
9	1.3	1.14
11	2	1.92
12	2.2	1.66
13	1.7	1.20
15	1.3	1.10
17	1.9	1.88
26	0.3	0.25
34	0.4	0.52
35	0.3	0.16
36	0.8	0.70
37	0.7	0.40
38	0.9	0.52
39	1.2	1.15

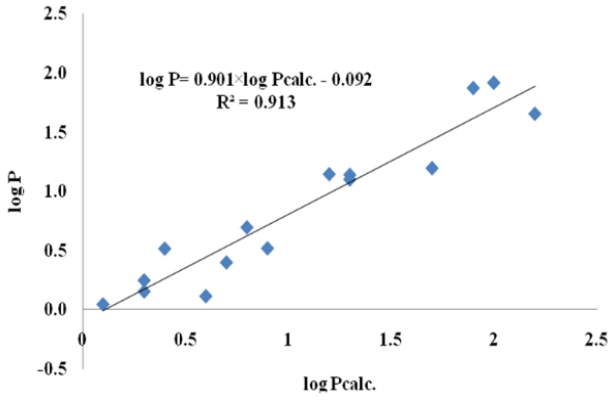

Figure 3. The plot log P vs. log P calc. for the test set (similarity clusters)

From Figure 3 one can see the better fitting of data predicted by clustering the test set according to (2D) similarity of each of the 15 molecular structures, taken as leaders.

CONCLUSIONS

A set of 40 serotonin derivatives, downloaded from the PubChem database, was submitted to a QSAR study. The best models have been validated in the external test set and in a new version of validation/prediction by using clusters of similarity, that provides „quasi-congeneric" structures, important for a prediction that surpasses the model. This punctual prediction is much more important in comparison with a more general model but with a lower predictability.

REFERENCES

[1]. S.N. Young, *Rev. Psychiatr. Neurosci*. **2007**, *32*, 394.
[2]. M.W. King, Med. Biocchem. Page. Indiana Univ. School Med. Retrieved 1 Dec. **2009**.
[3]. M. Berger, J.A. Gray, B.L. Roth, *Annu. Rev. Med*. **2009**, *60*, 355.
[4]. P. Bianchi, D.R. Pimentel, M.P. Murphy, W.S. Colucci, A. Parini, *Federation of American Societies for Experimental Biology Journal*, **2005**, *19*, 641.
[5]. K. Kang, S. Park, Y.S. Kim, S. Lee, K. Back, *Appl. Microbiol. Biotechnol*. **2009**, *83*, 27.
[6]. PubChem database, accessed 20.08. **2014**.
[7]. A.T. Balaban, A. Chiriac, I. Motoc, Z. Simon, Steric Fit in QSAR (Lectures Notes in Chemistry, Vol. 15), Springer Berlin, **1980**.
[8]. O. Ursu, M.V. Diudea, TOPOCLUJ software program, Babes-Bolyai University, Cluj, **2005**.
[9]. C.D. Moldovan, A. Costescu, G. Katona, M.V. Diudea., *J. Math. Chem.,* **2008**, *45*, 442.
[10]. J.G. Topliss, R.J. Costello, *J. Med. Chem.*, **1972**, *15*, 1066.
[11]. D.M. Hawkins, S.C. Basak, D. Mills, *J. Chem. Inf. Comput. Sci*., **2003**, *43*, 579.
[12]. L. Jäntschi, AcademicDirect Library of software, **2005**. Available at: *http://l.academicdirect.org/Chemistry/SARs/SARs/loo/*
[13]. A.M. Harsa, T.E. Harsa, S.D. Bolboacă, M.V. Diudea, *Curr. Comput.-Aided Drug Des*. **2014**, *10*, 115.