STUDIA UBB CHEMIA, LX, 1, 2015 (p. 165-176) (RECOMMENDED CITATION)

QSAR STUDY ON (1-METHYLPIPERIDIN-4-YL) PROPANOATE DERIVATIVES BY SIMILARITY CLUSTER PREDICTION

TEODORA E. HARSA^a, ALEXANDRA M. HARSA^a, AND MIRCEA V. DIUDEA^a*

ABSTRACT. Derivatives of (1-methylpiperidin-4-yl) propanoate, with similar biological characteristics, have been reported for patients with Alzheimer disease [1,2]. QSAR study was performed on a set of 40 (1-methylpiperidin-4-yl) propanoate derivatives, downloaded from the PubChem database and aligned over a hypermolecule that mimics the investigated correlational space. The best models describing log P of these (1-methylpiperidin-4-yl) propanoate derivatives were validated by leave-one-out procedure [3] and best predicted within similarity clusters.

Keywords: (1-methylpiperidin-4-yl) propanoate, log P, QSAR, Hypermolecule, similarity cluster

1. INTRODUCTION

1-[11C]methylpiperidin-4-yl propanoate (i.e. [11C]PMP) and N [11C] methyl piperidin-4-yl acetate ([11C]MP4A) are used as radioligands in positron emission tomography (PET) [4,5]. Synthesis of [11C]PMP can be achieved by direct N-methylation of 4-piperidinyl propanoate with [11C]methyl trifluoromethanesulfonate at room temperature in dimethyl-formamide [6].

Topological indices are numerical representations of the chemical structures, computed on the basis of molecular graph [7]. Topological indices can be easily calculated with TOPOCLUJ software and they have shown good correlation with log P [8,9].

QSAR/QSPR methodologies (Quantitative Structure Activity Relationships/ Quantitative Structure Property Relationships) attempt to correlate molecular structures with their properties or biological activities. QSAR is widely used in pharmacology, environmental search, and agricultural chemistry, etc. [10].

^a Babeş-Bolyai University, Faculty of Chemistry and Chemical Engineering, 11 Arany Janos str., RO-400028, Cluj-Napoca, Romania. * Corresponding author: diudea@gmail.com

Multiple linear regressions (MLR) relate e.g. log P (measuring the lipophilic character or drug-likeness of a given molecule) and topological indices, as shown in refs. [11-13].

2. DATA SET

A set of 40 molecular structures, derivatives of (1 methylpiperidin-4-yl) propanoate, have been downloaded from the Pubchem database [14] (Table 1), together with their log P. The set was split into the training set and test set (25 and 15 molecules, respectively, randomly chosen). The structures have been optimized at Hartree-Fock HF (6-31g(d,p)) level of theory. The calculations were performed in gas phase by Gaussian 09 [15].

Mol.	Canonical SMILES	Log P	CID
1	CCC(=0)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

Table 1. (1-methylpiperidin-4-yl) propanoate molecular structures
(in SMILES code) and their log P (taken from PubChem).

QSAR STUDY ON (1-METHYLPIPERIDIN-4-YL) PROPANOATE DERIVATIVES BY ...

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



Figure1: Hypermolecule

A hypermolecule (Figure 1) was built up as the reunion of the structural features in all 40 molecules under study [16,17].

3.COMPUTATIONAL DETAILS

Topological indices have been computed by TOPOCLUJ software [8]; some of them (Total adjacency = Adj, Conectivity=C, Detour = De, Distance = Di, D3D), SD_{k} ,k=1,2, HOMO (in au) and log P are listed in Table 2.

4. RESULTS AND DISCUSSION

4.1. Mass fragments description (case 1)

4.1.1.Data reduction (for log P)

In the step of data reduction, all the descriptors with the variance Var<20% and those with intercorrelation larger than 0.80 have been discarded.

Correlation weighting was performed on all the positions in the hypermolecule: the correlating coefficients of the statistically significant positions of the hypermolecule were used to multiply the local descriptors. Next, the correlating weighed local descriptors are summed to give a global descriptor, $SD_i = \sum_j CD_{ij}$ [18-20]. This new descriptor is a linear combination of the local weighed correlating descriptors for the significant positions in the Hypermolecule: H4, H8, H9, H10, H13, H14, H15, H18, H21, H23, H25, H27, H29, H32, H33,

Eq: $\log P = 2.849 + 1.0001 \times SD_1$, R²=0.907, St. Error=0.327, F=0.327

Mol.	log P	SD1	SD2	НОМО	Adj	Di	С	CjDi
1	1.2	-1.421	-1.839	-9.207	15	402	16	512
2	3.1	-0.146	-0.567	-9.146	17	562	18	704
3	2.9	0.031	-1.034	-9.175	16	480	17	601
4	3.8	1.026	0.423	-9.025	19	750	20	856.5
5	1.5	-1.410	-2.431	-9.263	13	276	14	352
6	1.2	-1.233	-2.294	-9.322	12	216	13	275
7	1.2	-1.632	-2.248	-9.261	16	537	17	659
8	1.1	-1.429	-2.541	-9.250	12	217	13	281
9	0.9	-1.647	-2.251	-9.272	15	440	16	548
10	2.6	-0.441	-1.187	-9.161	16	531	17	615.5
11	3.2	0.148	-0.884	-9.032	17	574	18	666.5
12	2.2	-0.258	-0.965	-9.244	16	510	17	594.5
13	1.9	-1.088	-2.500	-9.269	15	358	16	433.5
14	4	1.004	0.600	-8.847	15	368	15	444.5
15	3.4	0.971	-0.354	-9.153	17	518	18	610.5
16	0.5	-1.891	-2.983	-9.533	17	588	19	734
17	4.6	1.866	0.247	-9.680	17	556	18	697
18	2	-0.588	-1.523	-9.699	17	582	19	725
19	2.9	-0.588	-0.638	-9.648	15	366	16	458
20	1.1	-1.284	-2.417	-9.801	15	382	17	489.5

Table 2. log P, correlating	descriptors SD, HO	MO energy (au) ar	nd topological	indices
for the set of	(1-methylpiperidin-4-	-yl) propanoates in	Table 1	

21	1.5	-1.284	-2.136	-9.769	16	452	18	578.5
22	-0.1	-2.919	-3.136	-9.694	16	475	18	595
23	1	-1.676	-2.716	-9.501	19	778	21	960
24	0.5	-1.676	-2.653	-9.524	18	667	20	831
25	1.4	-1.447	-2.106	-9.653	20	981	22	1190
26	0.7	-2.227	-3.415	-9.644	16	475	18	595
27	1.9	-0.595	-1.734	-9.665	21	1130	23	1361
28	1.6	-1.094	-2.044	-9.631	20	964	22	1173
29	1.1	-2.164	-2.636	-9.656	19	834	21	1021
30	1.6	-1.330	-2.554	-9.682	20	888	22	1064
31	2.4	-1.233	-1.162	-9.484	11	168	12	212.5
32	2	-1.147	-2.363	-9.282	14	355	15	443
33	2.3	-0.595	-1.387	-9.628	22	1281	24	1534
34	3	-0.160	-0.669	-9.501	23	1308	25	1562
35	1.5	-1.676	-2.243	-9.571	18	667	20	831
36	1.3	-1.871	-2.518	-9.562	18	682	20	846
37	1	-2.082	-2.814	-9.598	17	564	19	698
38	0.5	-2.082	-2.746	-9.599	16	475	18	595
39	2	-0.962	-1.615	-9.562	22	1096	23	1327
40	2	-1.219	-2.170	-9.137	15	434	17	546

QSAR STUDY ON (1-METHYLPIPERIDIN-4-YL) PROPANOATE DERIVATIVES BY ...

4.1.2. QSAR models (for log P)

The models were performed on the training set (the 25 structures in Table 1) and the best results (in decreasing order of R^2) are listed below and in Table 3.

(i) Monovariate regression

 $\log P = 2.741 + 1.014 \times SD_1$

(ii) Bivariate regression

 $\log P = 2.764 + 1.013 \times SD_1 - 0.00004 \times D3D$

(iii) Three-variate regression

 $\log P = 1.901 + 0.948 \times SD_1 + 0.008 \times De - 0.008 \times CjDi$

(iv) Four-variate regression

 $\log P = -7.216 + 0.0.828 \times SD_1 - 1.011 \times HOMO + 0.969 \times Adj. - 0.879 \times C$

TEODORA E. HARSA, ALEXANDRA M. HARSA, MIRCEA V. DIUDEA

	Descriptors	R ²	Adjust. R ²	St. Error	F
1	SD ₁	0.930	0.927	0.311	304.509
2	НОМО	0.140	0.103	1.088	3.742
3	CjDi	0.009	-0.034	1.168	0.219
4	Di	0.004	-0.039	1.171	0.102
5	De	0.003	0.040	1.172	0.074
6	SD ₁ , D3D	0.930	0.923	0.318	145.776
7	SD ₁ , Di	0.930	0.923	0.318	145.658
8	SD ₁ , De	0.930	0.923	0.318	145.638
9	SD ₁ , CjDi	0.930	0.923	0.318	145.723
10	SD ₁ , Adj.	0.930	0.923	0.318	145.732
11	SD₁, De, CjDi	0.944	0.936	0.292	117.235
12	SD ₁ , Adj., C	0.939	0.931	0.302	108.542
13	SD ₁ , De, D3D	0.936	0.927	0.311	102.306
14	SD₁, CjDi, CjDe	0.936	0.926	0.312	101.724
15	SD ₁ , Di, D3D	0.932	0.922	0.320	95.885
16	SD ₁ , D3D, HOMO	0.930	0.920	0.325	92.813
17	SD ₁ , HOMO, C	0.930	0.920	0.325	92.783
18	SD ₁ , HOMO, Adj., C	0.952	0.943	0.275	99.593
19	SD ₁ , De, Di, D3D	0.938	0.926	0.313	75.728
20	SD1,HOMO, Di, D3D	0.932	0.931	0.328	68.492

Table 3. Best models in describing log P in the training set of (1-methylpiperidin-4-yl) propanoate in Table1

4.1.3. Model Validation (for log P)

(a) Leave-one-out

The best models from Table 3 describing log P of these (1methylpiperidin-4-yl) propanoate derivatives were validated by leave-oneout LOO procedure, as listed in Table 4.

	Descriptors	\mathbf{Q}^2	R ² -Q ²	St. Error _{ioo}	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	SD₁, De, CjDi	0.916	0.028	0.34	250.326
19	SD ₁ , HOMO, Adj., C	0.922	0.03	0.328	272.132

Table 4. LOO for some models in Table 3

QSAR STUDY ON (1-METHYLPIPERIDIN-4-YL) PROPANOATE DERIVATIVES BY ...

(b) External Validation

The values log P for the test set of (1-methylpiperidin-4-yl) propanoates were calculated by using equation cf. entry 11, Table 3. Data are listed in Table 5 and the monovariate correlation: R^2 =0.869, St. Error=0.331, F=86.123 is plotted in Figure 2.



(c) Similarity Cluster Validation

Validation can also be performed by calculating log P for the molecules in the test set by using clusters of similarity: each of the 15 molecules is the leader of its own cluster, selected by 2D similarity among the 25 structures of the initial learning set. The values log P calc. were computed by 15 new equations (the leader being left out) with the same descriptors as in eq. 11, Table 3. Data are listed in Table 6 and the monovariate correlation: $\log P = 1.009 \times \log P_{calc.} - 0.339$ R²=0.953, St. Error=0.197, F=265.759 is plotted in Figure 3.

TEODORA E. HARSA, ALEXANDRA M. HARSA, MIRCEA V. DIUDEA

Table 6. Calculated					
value	values of log P for the				
mole	cules in	the test			
s	et (Tab	le 1)			
Mol.	log P	log P _{calc.}			
2	3.1	2.51			
3	2.9	2.65			
4	3.8	3.89			
5	1.5	1.09			
30	1.6	1.51			
31	2.4	1.83			
32	2	1.55			
33	2.3	2.07			
34	3	2.60			
35	1.5	1.00			
36	1.3	0.82			
37	1	0.68			
38	0.5	0.45			
39	2	1.66			
40	2	1.80			



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

4.2. Partial charges description (case 2)

4.2.1.Data reduction (for log P)

This new descriptor is a linear combination of the local correlating descriptors for the significant positions in the hypermolecule (i.e. H1, H4, H7, H8, H11, H12, H13, H14, H15, H16, H17, H22, H24, H32, H33)

Eq., $\log P = 3.667 + 1.001 \times SD_2$ R²=0.892, St. Error=0.352, F=313.391

4.2.2. QSAR models (for log P)

The models were performed on the training set (the 25 structures in Table 1) and the best results (in decreasing order of R^2) are listed below and in Table 7.

(i) Monovariate regression

 $\log P = 3.704 + 1.071 \times SD_2$

(ii) Bivariate regression

 $\log P = 3.182 + 1.078 \times SD_2 - 0.057 \times HOMO$

(iii) Three-variate regression

 $\log P = 2.847 + 1.088 \times SD_2 - 0.110 \times HOMO - 0.0002 \times De$

(iv) Four-variate regression

 $\log P = -0.844 + 1.003 \times SD_2 - 0.478 \times HOMO + 0.386 \times Adj. - 0.360 \times C$

172

QSAR STUDY ON (1-METHYLPIPERIDIN-4-YL) PROPANOATE DERIVATIVES BY ...

	Descriptors	R^2	Adjust. R ²	St. Error	F
1	SD ₂	0.917	0.914	0.36	256.093
2	НОМО	0.192	0.157	1.127	5.470
3	De	0.007	0.036	1.250	0.157
4	Di	0.005	0.039	1.251	0.110
5	SD ₂ , De	0.918	0.911	0.366	123.875
6	SD ₂ , HOMO	0.918	0.910	0.368	122.668
7	SD ₂ , Di	0.918	0.911	0.366	123.881
8	SD ₂ , CjDi	0.918	0.911	0.366	123.843
9	SD ₂ , D3D	0.918	0.911	0.366	123.969
10	SD ₂ , HOMO, De	0.920	0.904	0.381	57.284
11	SD ₂ , CjDi, HOMO	0.919	0.907	0.374	79.230
12	SD ₂ , D3D, De	0.919	0.907	0.375	78.909
13	SD ₂ , Di, D3D	0.919	0.907	0.375	78.921
14	SD ₂ , De, CjDi	0.918	0.907	0.375	78.835
15	SD ₂ , Adj, C	0.918	0.907	0.375	78.634
16	SD ₂ , HOMO, C, Adj,	0.921	0.905	0.378	58.100
17	SD ₂ , C, De, Adj	0.920	0.904	0.381	57.263
18	SD ₂ , HOMO, De, Di	0.919	0.903	0.383	56.755

Table 7. Best models in describing log P in the training set of.(1-methylpiperidin-4-yl) propanate in Table1.

4.2.3. Model Validation (for log P)

(a) Leave-one-out

The best models in Table 7 describing log P of these (1methylpiperidin-4-yl) propanoate derivatives were validated by leave-one-out procedure, as listed in Table 8.

Table 8. LOO for some models in Table 7

	Descriptors	Q^2	R^2-Q^2	St. Error _{loo}	Floo
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

TEODORA E. HARSA, ALEXANDRA M. HARSA, MIRCEA V. DIUDEA

(b) External Validation

The values log P for the test set of (1-methylpiperidin-4-yl) propanoate were calculated by using equation cf. entry 10, Table 7. Data are listed in Table 9 and the monovariate correlation: $\log P = 1.010 \times \log P_{calc.} - 0.291$ R²=0.833, St. Error=0.286, F=64.662 is plotted in Figure 4.



(c) Similarity Cluster Validation

Validation was performed by calculating log P for the molecules in the test set, similar to that in the Section 4.1.3. The values log $P_{calc.}$ were computed with the same descriptors as in eq. 10, Table 7. Data are listed in Table 10 and the monovariate correlation: $\log P = 1.108 \times \log P_{calc.} - 0.425$ R²=0.945, St. Error=0.164, F=222.647 is plotted in Figure 5.

QSAR STUDY ON (1-METHYLPIPERIDIN-4-YL) PROPANOATE DERIVATIVES BY ...

Table 10. Calculated values of log P by similarity clusters, for the molecules in the test set					
Mol.					
26	0.7	0.10			
27	1.9	1.78			
28	1.6	1.56			
29	1.1	0.89			
30	1.6	0.95			
31	2.4	2.35			
32	2	1.64			
33	2.3	2.17			
34	3	2.93			
35	1.5	1.28			
36	1.3	0.98			
37	1	0.66			
38	0.5	0.41			
39	2	1.91			
40	2	1.61			



Figure 5. The plot log P vs. log P calc. for the test set (similarity cluster validation)

CONCLUSIONS

A set of 40 derivatives of (1-methylpiperidin-4-yl) propanoates, downloaded from the PubChem database, has been submitted to a QSAR study, involving the hypermolecule concept.

The set was split into a learning set and a test set, the last one being used for the validation of the models, in the so-called external set validation. Also, the validation was made by a new version of prediction by using similarity clusters.

ACKNOWLEDGMENTS

This paper is a result of a doctoral research made possible by the financial support of the Sectorial Operational Program for Human Resources Development 2007-2013, co-financed by the European Social Fund, under the project POSDRU/159/1.5/S/137750 - "Doctoral and postoctoral programs - support for increasing research competitiveness in the field of exact Sciences".

TEODORA E. HARSA, ALEXANDRA M. HARSA, MIRCEA V. DIUDEA

REFERENCES

- [1]. F. Liu, Y. Liang, C. Caoa, N. Zhou, Analytica Chimica Acta, 2007, 594, 279.
- [2]. C.S. Lee, S.J. Chung, J.S. Kim, Journal of Clinical Neurology, 2006, 2, 83.
- [3]. L. Jäntschi, LOO Analysis (LOO: leave one out), AcademicDirect Library of software, 2005, Available at:
 - http://l.academicdirect.org/Chemistry/SARs/MDF_SARs/loo/
- [4]. H. Shinotoh, K. Fukushi, S. Nagatsuka, T. Irie, Curr Pharm Des. 2004, 10, 13, 1505.
- [5]. D.M. Jewett, T.B. Nguyen, D.E. Kuhl, M.R. Kilbourn, Nuclear Medicine and Biology, 1998, 25, 751.
- [6]. S.E. Snyder, L. Tluczek, D.M. Jewett, T.B. Nguyen, D.E. Kuhl, M.R. Kilbourn, Nucl Med Biol. 1998, 25, 8, 751.
- [7]. Q. Shu, J. Wang, B. Peng, D. Wang, G. Wang, Fuel, 2008, 87, 3586.
- [8]. O Ursu, M.V. Diudea, TopoCluj software program. Babes-Bolyai University Cluj, 2005.
- [9]. I.A. Aguayo-Villarreal, V. Hernández-Montoya, N.A. Rangel-Vázquez, M.A. Montes-Morán, Journal of Molecular Liquids, 2014, 196, 326.
- [10]. F. Liu, Y. Liang, C. Caob, N. Zhoua, Talanta, **2007**, 72, 1307.
- [11]. C.A. Nunes, M. P. Freitas, Journal of Microbiological Methods, 2013, 94, 217.
- [12]. T. Cheng, Y. Zhao, X. Li, F. Lin, Y. Xu, X. Zhang, Y.Li, R. Wang, L. Lai, Journal of Chemical Information and Modelin, 2007, 47 (6), 2140.
- [13]. C.D. Moldovan, A. Costescu, G. Katona, M.V. Diudea, MATCH Commun. Math. Comput. Chem., 2008, 60, 977.
- [14]. Pub. Chem. database, accessed 1.09. 2014.
- [15]. M.J. Frisch, G.W. Trucks, et. all, Gaussian Inc Wallingford CT, 2009.
- [16] N.J.Richmond, C.A. Abrams, P.R. Wolohan, E. Abrahamian, P. Willett, R.D. Clark, J Comput Aided Mol Des, 20, 2006, 567.
- [17]. A.T. Balaban, A. Chiriac, I. Motoc, Z. Simon, Springer, Berlin, 15, 1980.
- [18]. T.E. Harsa, A.M. Harsa, B. Szefler, Cent. Eur. J. Chem., 2014, 12, 365.
- [19]. A.M. Harsa, T.E. Harsa, S.D. Bolboacă, M.V. Diudea, Current Computer-Aided Drug Design, BEntham Science, 2014, 2, 10, 115.
- [20]. T.E. Harşa,, Studia UBB Chemia, LIX, 1, 2014, 99.