STUDIA UNIVERSITATIS BABEŞ-BOLYAI, CHEMIA, XLIX, 2, 2004

Dedicated to Professor Valer Fărcăşan at his 85th anniversary

HYDANTOIN DERIVATIVES HPLC-RT LIPOPHILICITIES: A QSPR STUDY

OLEG URSU[a], MIRCEA V. DIUDEA[b]

[a,b] *Faculty of Chemistry and Chemical Engineering Babes-Bolyai University, 400028 Cluj, Romania*

ABSTRACT. A quantitative structure-property relationship QSPR investigation was performed on the lipophilicities of a number of hydantoin derivates as measured RT-HPLC retention times provided by Scholl et al[1]. The lipophilicities (S) were correlated with a series of graph theoretical, geometrical, and electronic descriptors provided by TOPOCLUJ 3.0 software package from PM3-optimized geometry. In second step these descriptors were incorporated into the descriptor matrix to build several QSPRs in view of obtaining prediction models for lipophilicity.

INTRODUCTION

Hydantoins find important applications as medicines (e.g., as anticonvulsant drugs in treatment of epilepsy) and as agrochemicals. Lipophilicity represents the affinity of a molecule or a moiety for a lipophilic environment. It is commonly measured by its distribution behavior in a biphasic system, either liquid-liquid (e.g., partition coefficient in 1-octanol/water) or solid-liquid (retention on reversed-phase high-performance liquid chromatography (RT-HPLC) or thin-layer chromatography (TLC) system). Lipophilicity plays a vital role in physicochemical, environmental, and biological, processes as it determines the transport phenomena in vivo such as through cell membrane barrier. With the advent of inexpensive and rapid computation, there has a remarkable growth interest in the field of quantitative structure-property relationships QSPR and use of multivariate linear regression and latent variables methods to model relevant properties as a function of molecular structural parameters (*i*. *e*., molecular descriptors).

DATA ACQUISITION

Data set. The series of hydantoin HPLC-RT (Table 1) reported by Scholl et al[1] was used in this QSPR study; lipophilicities could be expressed as partition coefficient in 1-octanol/water *logP* biphasic system, but lipophilicities derived from HPLC-RT are preferred due to the following advantages: analytical procedure, easy use of mixtures, and absence of incertitude in determination of concentration.

Derivation of descriptors. The structures of all the hydantoins (Figure 1) were drawn with HYPERCHEM molecular modeling software package.[2] After preoptimization using Molecular Mechanics (MM+, HYPERCHEM), the geometries of these compounds were further optimized using the semiempirical PM3 parameterization of the HYPERCHEM. The resulting output files containing the refined geometry represent the input for the TOPOCLUJ to calculate the molecular

descriptors. This provided a number of electronic, topological, and geometrical descriptors, used in building the descriptors matrix with dimensions 635 x 73. It was subsequently employed for correlating the lipophilicity values of the hydantoins.

TABLE 1.

HPLC-RT Lipophilicity (S) Data of Substituted Hydantoins

No.	S	No.	S	No.	S
1		26	-0.197	51	-0.147
2	-0.157	27	-0.183	52	-0.170
3	-0.157	28	-0.222	53	-0.191
4	-0.174	29	-0.224	54	-0.165
5	-0.180	30	-0.233	55	-0.174
6	-0.144	31	-0.202	56	-0.184
7	-0.142	32	-0.190	57	-0.174
8	-0.145	33	-0.173	58	-0.182
9	-0.141	34	-0.210	59	-0.193
10	-0.186	35	-0.164	60	-0.191
11	-0.157	36	-0.200	61	-0.190
12	-0.168	37	-0.210	62	-0.225
13	-0.211	38	-0.191	63	-0.233
14	-0.209	39	-0.197	64	-0.211
15	-0.227	40	-0.222	65	-0.208
16	-0.193	41	-0.209	66	-0.197
17	-0.178	42	-0.260	67	-0.236
18	-0.190	43		68	-0.216
19	-0.231	44	-0.221	69	-0.219
20	-0.193	45	-0.208	70	-0.210
21	-0.210	46	-0.229	71	-0.240
22	-0.192	47	-0.176	72	-0.196
23	-0.213	48	-0.157	73	-0.248
24	-0.187	49	-0.157	74	-0.207
25	-0.209	50	-0.199	75	-0.206

Molecular structure of the compounds is presented in Figure 1.

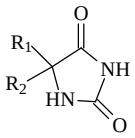
A large variety of descriptors have been used: electronic (VEA, VRA, VED, VRD indices[2]) geometrical (3D-Wiener index[3]), and topological Wiener[4]

$$W = (1/2) \sum_i \sum_j [\mathbf{D}]_{ij} \quad (1)$$

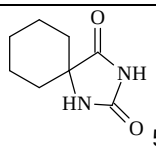
where $[D]_{ij}$ denote the entries in the distance matrix D which are just the just the topological distances d_{ij} between *i* and *j*.

The Cluj indices are calculated[5.] as half-sum of the entries in a Cluj symmetric matrix, **M**, (**M = CJD, CJΔ, CFD, CFΔ**), defined as follows:

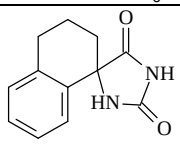
$$[\mathbf{UM}]_{ij} = \max_{k=1,2,\ldots} \left| V_{i,j,p_k} \right| \quad (2)$$



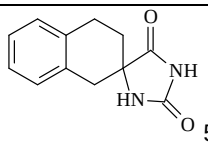
No.	R_1	R_2	No.	R_1	R_2	No.	R_1	R_2
1	H	Me	18	Ph	EtOOC	35	Et	3-ClC_6H_4
2	H	Ph	19	Ph	Ph	36	Me	4-ClC_6H_4
3	H	$PhCH_2$	20	Ph	$PhCH_2$	37	Et	4-ClC_6H_4
4	H	$PhCH_2CH_2$	21	Et	4-MeC_6H_4	38	^{n}Pr	4-ClC_6H_4
5	H	$^{c}C_6H_{11}CH_2CH_2$	22	^{n}Pr	4-MeC_6H_4	39	^{i}Pr	4-ClC_6H_4
6	Me	^{c}Pr	23	^{n}Bu	4-MeC_6H_4	40	^{n}Bu	4-ClC_6H_4
7	Me	^{i}Pr	24	^{i}Pr	4-MeC_6H_4	41	$(CH_3)CHCH_2$	4-ClC_6H_4
8	Me	$(CH_3)CHCH_2$	25	$(CH_3)CHCH_2$	4-MeC_6H_4	42	$^{n}C_7H_{15}$	4-ClC_6H_4
9	Me	^{t}Bu	26	Me	4-$^{t}BuC_6H_4$	43	$HOOCCH_2$	4-ClC_6H_4
10	Me	$(CH_3)CH(CH_2)_2$	27	Et	4-$MeOC_6H_4$	44	$MeOOCCH_2$	4-ClC_6H_4
11	Ph	Me	28	^{i}Pr	4-$MeOC_6H_4$	45	Me	4-BrC_6H_4
12	Ph	Et	29	$(CH_3)CHCH_2$	4-$MeOC_6H_4$	46	^{n}Bu	4-BrC_6H_4
13	Ph	^{n}Pr	30	^{n}Bu	4-$MeOC_6H_4$	47	Et	4-O_2NC6H4
14	Ph	^{i}Pr	31	^{n}Pr	4-$MeOC_6H_4$	48	Me	2-thienyl
15	Ph	^{n}Bu	32	Et	4-FC_6H_4	49	Me	2-furanyl
16	Ph	$^{c}C_6H_{11}$	33	Et	2-FC_6H_4	50	Me	2-benzofuranyl
17	Ph	MeOOC	34	Et	3-ClC_6H_4			



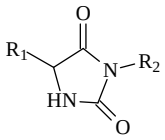
51

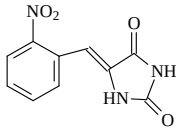


52



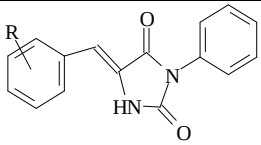
53





57

No.	R_1	R_2
54	$PhCH_2$	Me
55	Me	$PhCH_2$
56	$PhCH_2$	$PhCH_2$



No.	R	No.	R	No.	R
58	H	64	2-Cl	70	4-Br
59	2-Me	65	3-Cl	71	4-F_3C
60	3-Me	66	2,6-Cl	72	2-MeO
61	4-Me	67	2,4-Cl	73	4-$PhCH_2O$
62	2,4,6-Me	68	2-Br	74	4-MeS
63	4-^{i}Pr	69	3-Br	75	4-NC

Figure 1. Structures of Hydantoins

$$V_{i,j,p_k} = \{v \mid v \in V(G); d_{iv} < d_{jv};\ \ (i,v)_h \cap p_k = \{i\};\ p_k \in D(G) \text{ or } \Delta(G)\};\ h,k = 1,2,... \tag{3}$$

$$IE(M) = (1/2)\sum_i \sum_j [\mathbf{M}]_{ij}[\mathbf{A}]_{ij} \tag{4}$$

$$IP(M) = (1/2)\sum_i \sum_j [\mathbf{M}]_{ij} \tag{5}$$

or from an asymmetric Cluj matrix, by

$$IE2(UM) = (1/2)\sum_i \sum_j [\mathbf{UM}]_{ij}[\mathbf{UM}]_{ji}[\mathbf{A}]_{ij} \tag{6}$$

$$IP2(UM) = (1/2)\sum_i \sum_j [\mathbf{UM}]_{ij}[\mathbf{UM}]_{ji} \tag{7}$$

In the above, *D*(*G*) and Δ(*G*) represents the distance and detour sets in *G*.

The number defined on edge, *IE*, is an *index* while the number defined on path, *IP* is a *hyper-index*. Note that the operators IE and IP, as well as $IE2$ and $IP2$ may be applied to both symmetric and asymmetric matrices. When the last two operators are calculated on a symmetric matrix, the terms of sum represent squared entries in that matrix.

Randic[7] and DSI[6] index use the relation (8) to describe features of molecules.

$$\chi = \sum_{(i,j)\in E(G)} (\delta_i \delta_j)^{-1/2} \tag{8}$$

where δ_i and δ_j represent the corresponding vertex degrees. In analogy to Randic index was defined the DSI index by using group electronegativity valences EVG_i:

$$DS_{\boldsymbol{i}} = \sum_{j:(i,j)\in E(G)} (EVG_i EGV_j)^{-1/2} \tag{9}$$

$$DS = \sum_i DS_{\boldsymbol{i}} \tag{10}$$

For the definition of other indices see.[5,8]

REGRESSION ANALYSIS

The quantitative relationship between calculated descriptors and property is a difficult task due to large number of descriptors, and there is no guarantee that the best subset is found. There is the possibility to eliminate descriptors irrelevant to property of interest by stepwise regression eventually followed by more sophisticated statistical methods analysis like PLS, CoMFA.

PLS is a projection method in which the data matrix represented as a set of n points in an m-dimensional space is projected on to a k-dimensional hyperplane, in such way that the coordinates of the projection are good predictors of some *y* property. PLS allows the simultaneous use of strongly intercorrelated x-descriptors by focusing the systematic covariances in the X block in a few latent variables.

In this work we use the following methodology: (a) divide the initial set by property values in 4 subsets and one external prediction set; (b) autoscaling descriptors matrix; (c) PLS with Leave one out (LOO) (MATLAB[8] and PLS toolbox[9]) modeling of 4 subsets of data; the number of factors is determined by improvement of percent variance captured by model (PRESS), and additional factor is taken at each 2% improvement of PRESS; (d) model validation with external prediction set.

The regression analysis summary and best PLS models are presented in Table 2 and Figure 2.

TABLE 2

Correlations of Lipophilicities of Hydantoins by PLS method

Subset	R^2	RMSECV	Percent cumulative variance	No. of LVs	Lipophilicity interval	Compounds in the model
A	0,984	0,0010843	98,42	4	-(0,14-0,17)	7,8,51,49,3,2,11,35, 54,52
B	0,985	0,001969	98,54	9	-(0,17-0,20)	33,57,55,4,47,17,5,27, 10,32,18,60,53,22,59 16,72,66,39,26,50,36
C	0,982	0,00094949	98,22	10	-(0,20-0,23)	31,75,65,45,25,14,70,3 7,64,13,23,44,40,28,46
D	0,998	0,00048575	99,79	4	-(0,23-0,26)	19,63,30,67,73,42

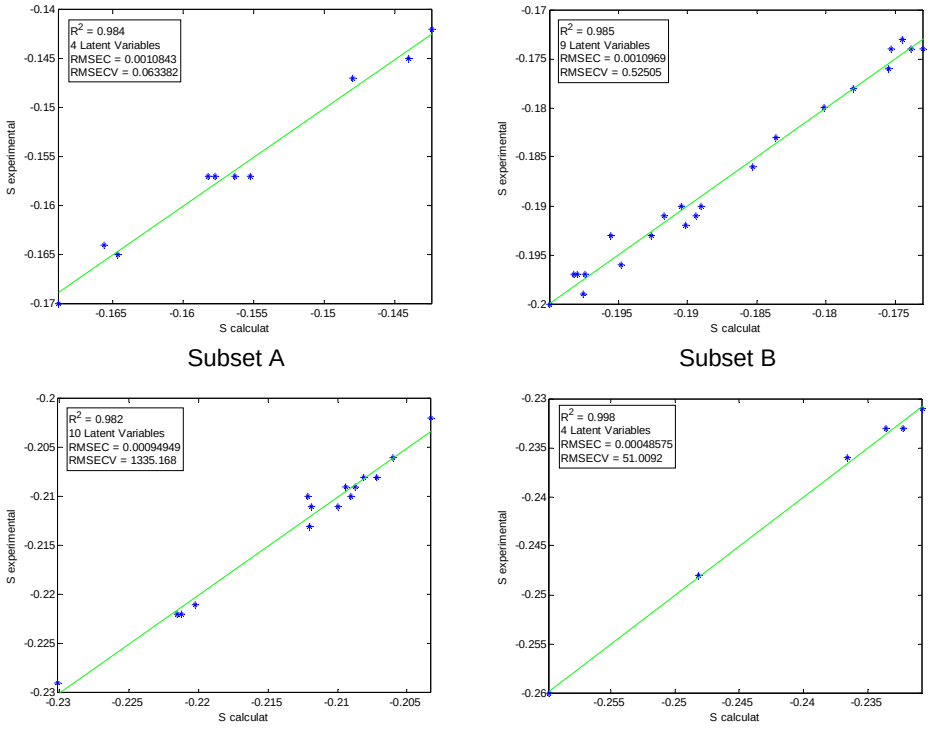


Figure 2. Best PLS models for chosen Hydantoins subsets

Validation of the PLS models. For each compound from external prediction subset, the PLS model for lipophilicity was chosen according to property value. The prediction summary and graph is presented in Table 3 and Figure 3.

TABLE 3.

Experimental and predicted lipophilicities for testing model robustness.

No.	*S* Experimental	*S* Calculated	RES	R^2	F
71	-0.240	-0.230	0.0021	0.9059	163.6176
15	-0.227	-0.212	0.0091		
62*	**-0.225**	**-1.358**	-		
29	-0.224	-0.203	0.0163		
69	-0.219	-0.213	0.0012		
68	-0.216	-0.222	-0.0103		
21	-0.210	-0.207	0.0000		
34	-0.210	-0.210	-0.0029		
41	-0.209	-0.217	-0.0105		
74	-0.207	-0.203	0.0016		
20	-0.193	-0.199	-0.0064		
38	-0.191	-0.186	0.0049		
61	-0.190	-0.192	-0.0023		
24	-0.187	-0.194	-0.0066		
56	-0.184	-0.188	-0.0034		
58	-0.182	-0.197	-0.0140		
12	-0.168	-0.161	0.0097		
48	-0.157	-0.157	0.0050		
6	-0.144	-0.146	0.0048		
9	-0.141	-0.146	0.0015		

*Excluded from validation subset

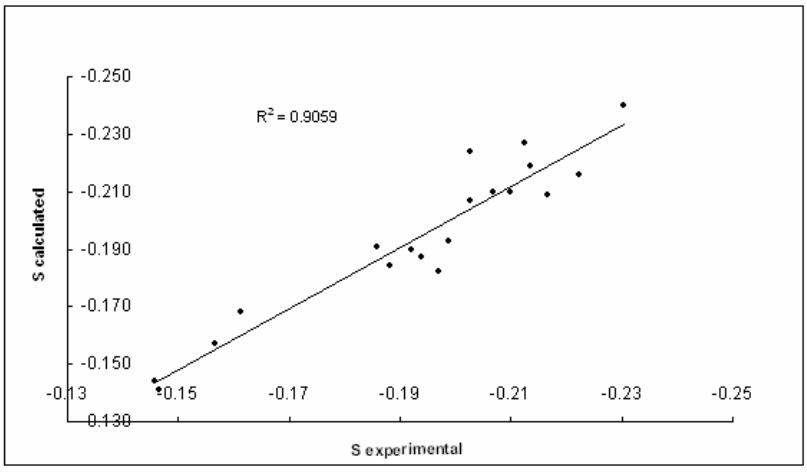


Figure 3. Plot of calculated vs experimental lipophilicities

CONCLUSIONS

This study provided improved correlations (R^2=0.9059) of the lipophilicity data of hydantoins by using TOPOCLUJ descriptors. Recall that the best literature result was R^2=0.829[10]. This study is useful in clustering and diagnosis of hydantoin compounds. For predicting new structures with desired lipophilicity, similarity studies are in progress in our laboratory.

REFERENCES

[1] Scholl, S.; Koch, A.; Henning, D.; Kempter, G.; Kleinpeter, The influence of structure and lipophilicity of hydantoin derivatives on anticonvulsant activity, E. Struct. Chem. 1999, **10**, 355-366

[2] Alexandru T. Balaban, Dan Ciubotariu, and Mihai Medeleanu, Topological indices and real number vertex invariants based on graph eigenvalues or eigenvectors, J. Chem. Inf. and Comput. Sci. 1992, **31**, 4, 517-523

[3] Bogdanov, B.; Nikolić, S.; Trinajstić, N. On the Three-Dimensional Wiener Number, *J. Math. Chem*. 1989, **3**, 299-309.

[4] Wiener, H., Structural determination of paraffin boiling points. J. Am. Chem. Soc. 1947, **69**, 17-20

[5] Lorentz Jantschi, Gabriel Katona and Mircea V. Diudea, Modeling Molecular Properties by Cluj Indices, MACH **41**, 2000, 151-188

[6] Mircea V. Diudea, Ivan Gutman, Jantschi Lorentz, Molecular topology, Nova Science Publishers, New-York, 2002.

[7] Randić, M. Characterization of molecular branching, J. Am. Chem. Soc. 1975, **97**, 6609-6615.

[8] Matlab 6.5.1, Mathworks Inc., www.mathworks.com

[9] PLS Toolbox 3.0 Eigenvector Research Inc., www.eigenvector.com

[10] Alan R. Katritzky, Subbu Perumal, Ruslan Petrukhin, and Erich Kleinpeter, CODESSA-Based Theoretical QSPR Model for Hydantoin HPLC-RT Lipophilicities, J. Chem. Inf. and Comput. Sci. 2001, **41**, 3, 569-574