HCA and PCA in study of structure property relationships of newly synthesized hydantoins and its molecular descriptors

Tatjana Djaković Sekulić
Department of Chemistry, Biochemistry and Environmental Protection, Faculty of Sciences, University of Novi Sad, Novi Sad, Republic of Serbia – tatjana.djakovic-sekulic@dh.uns.ac.rs

Zagorka Lozanov-Crvenković*
Department of Mathematics and Informatics, Faculty of Sciences, University of Novi Sad, Novi Sad, Republic of Serbia – zlc@dmi.uns.ac.rs

Abstract

The biological activity of the molecule greatly depends on its structural properties. Numerical expressions of these properties are reflected in various molecular descriptors. In the paper molecular descriptors that are of interest for the activity of 29 newly synthesized hydantoins were analyzed using multivariate exploratory analysis. Hierarchical cluster analysis and principal component analysis had been used to visualize grouping patterns among molecular descriptors as well as investigated compounds. Recognition of the similarity between investigated compounds and grouping pattern gives a starting point for understanding deeper level of physiological processes which may also provide base for future development and further derivatization or modification to obtain more potent and selective drugs.

Keywords: HCA; PCA; hydantoins; molecular descriptors.

1. Introduction

Hydantoins are a well known class of antiepileptic drugs used in the treatment of epilepsy for more than 70 years [1-3]. Although hydantoin based drugs are symptomatically effective in only 60-70% of patients, efforts to introduce a new generation of antiepileptic drugs did not give better results [4]. Besides anticonvulsive a broad spectrum of a different activities are recognized [5-8].

One of the goals of chemists is to predict factors that are important for the activity instead to search whether already synthesized compound is active or not. For that purpose some empirical rules such as Lipinski's "rule of five" (ROF) are very useful. In general, rule takes into account molecular properties important for a drug's pharmacokinetics, such as absorption, distribution, metabolism, and excretion (ADME) [9]. These properties are hydrogen bond donors/acceptors capability, molecular weight and partition coefficient (lipophilicity). All those properties of the molecules are usually expresses numerically by various molecular descriptors or determined experimentally.

In this study 29 newly synthesized hydantoins were investigated in order to understand better the structure property relationships of new derivatives with the emphasis to the properties that are connected with ROF. Molecular descriptors that express these properties were both calculated and experimental obtained. Similarities and differences between the tested compounds and its properties will be explored and visualized using principal component analysis (PCA) and hierarchical cluster analysis (HCA).

2. Investigated compounds and calculations

The general structure of investigated compounds is presented in Figure 1. These compounds were synthesized according to procedure that has been published earlier [10].

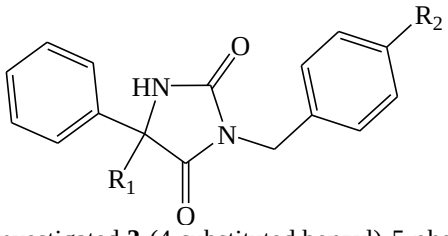

Figure 1. General formula of investigated **3**-(4-substituted benzyl)-5-phenylhydantoins. R_1 is ethyl (for compounds **1-7**), phenyl (**8-15**), methyl (**16-22**) and *tert*-butyl (**23-29**) group; R_2 is H (**1**, **8**, **16** and **23**), CH_3(**5**, **11**, **17** and **26**), OCH_3 (**2**, **9**, **22** and **27**), CN (**3**, **10**, **20** and **24**), NO_2 (**7**, **15**, **21** and **28**), Cl (**6**, **12**, **18** and **29**), Br (**4**, **14**, **19** and **25**), and $C(CH_3)_2$ (**13**).

Molecular descriptors were calculated using DRAGON software (Talete srl, Dragon (Software for Molecular Descriptor Calculation) Version 6.0 - 2014 - http://www.talete.mi.it/)). Following descriptors were calculayted: ALOGP (octanol-water partition coefficient), number of acceptor atoms for hydrogen bonds (nHAcc), number of donor atoms for hydrogen bonds (nHDon), MW (molecular weight), Druglikeness, Drug-Score, AMR (molar refractivity), TPSA(NO) (topological polar surface area using N,O polar contributions), RNB (number of rotatable bonds), Sp (sum of atomic polarizabilities), Pol (polarity number), Se (sum of atomic Sanderson electronegativities), IAC (total information index on atomic composition), Hy (hydrophilic factor), and Solubility.
Hierarchical clustering and principal component analysis were performed by statistical package Statistica 12 (StatSoft Inc., Tulsa, OK, USA), university license for Novi Sad University.

3. Experimental determination of descriptors

High-performance thin-layer chromatography was used for experimental determination of lipophilicity of investigated compounds. For that purpose the solutions of investigated compounds were prepared by dissolving of 0.02 g of each compound in 2 cm^3 of ethanol. Chromatographic conditions were: stationary phase was HPTLC of silica gel 60 RP-18 F254s (Merck, Darmstadt, Germany) and mobile phase methanol–water (with increasing methanol content from 50% to 90%, v/v; increment of 10%). After development the compounds were detected under UV illumination at 254 nm. The R_M values were calculated according to $R_M=log\ ((1/R_F)-1)$, where R_F is the retardation factor. The R_M value of each compound was examined as a function of mobile phase composition (ϕ):

$$R_M = R_M^0 - S\cdot\phi \quad (1)$$

Intercept R_M^0 of the Equation (1) is the most commonly used parameter for lipophilicity expression, *i.e.* molecular descriptor logP. Moreover the slope S and the ratio $C_0 = R_M^0/S$ are known in literature as alternative to R_M^0 [11]. In this study all three lipophilicity parameters will be considered, listed in Table 1 (columns 8-10).

4. Results and discussion

Molecular properties included in the Lipinski's "rule of five", namely partition coefficient (ALOGP), number of donor atoms for hydrogen bonds (nHDon), number of acceptor atoms for hydrogen bonds (nHAcc) and molecular weight (MW) are listed in Table 1, columns 1-4, respectively. According to the ROF numerical values of four mentioned descriptors should be close to five or a multiple of five. For each of these criteria, nearly 80–90% of the actual drugs examined fell below the cutoff range [9].
Column 5 in Table 1 lists the score for Lipinski's four descriptors. Except for compounds nos. **7**, **13**, **15** and **28** the remain hydantoins have score equal to 4. That means that the majority of investigated compounds are predicted to have good absorption or permeability properties. Compounds **7**, **13**, **15** and **28** (compounds with R_2=NO_2 and tBu) have score 3 that shows that the prediction for the same properties is indeterminate.
Lipinski rule do not predict activity but it defines the profile of common chemical, physical and/or physiological properties of successful drugs. Positive match of properties with such a profile points to drug-likeness. Thus the compounds with Lipinski's score of 4 are not necessary good drug candidates with high drug-likeness. For example, compound no. **10** is one with the lowest drug-likeness score

of -2.78 (column 7). Generally, negative value of drug-likeness score is observed for the hydantoins with cyano (**10**, **20, 24**) and nitro (**7**, **15**, **21** and **28**) group.

Table 1. Values of the molecular descriptors and Lipinski's number score for investigated compounds.

No	R_1/R_2*	ALOGP	nHDon	nHAcc	MW	Lipinski number	Drug-likeness	Drug score	R_M^0	S	C_0
		1	2	3	4	5	6	7	8	9	10
1	Et/H	3.18	1	4	294.38	4	9.52	0.83	3.23	3.94	0.82
2	Et/MeO	3.17	1	5	324.41	4	9.79	0.81	3.14	3.86	0.81
3	Et/CN	3.06	1	5	319.39	4	1.11	0.66	2.76	3.60	0.77
4	Et/Br	3.93	1	4	373.27	4	7.69	0.68	4.93	5.81	0.85
5	Et/Me	3.67	1	4	308.41	4	8.22	0.78	4.13	4.97	0.83
6	Et/Cl	3.85	1	4	328.82	4	10.41	0.72	4.30	5.15	0.84
7	Et/NO_2	3.08	1	6	339.38	3	-0.62	0.54	3.83	4.87	0.79
8	Ph/H	3.90	1	4	342.42	4	5.62	0.73	3.85	4.55	0.85
9	Ph/MeO	3.88	1	5	372.45	4	5.92	0.72	4.11	4.86	0.85
10	Ph/CN	3.77	1	5	367.43	4	-2.78	0.35	3.24	4.00	0.81
11	Ph/Me	4.38	1	4	356.45	4	4.28	0.67	4.32	4.95	0.87
12	Ph/Cl	4.56	1	4	376.86	4	6.57	0.61	4.73	5.37	0.88
13	Ph/tBu	5.30	1	4	398.54	3	2.36	0.46	5.97	6.51	0.92
14	Ph/Br	4.64	1	4	421.31	4	3.97	0.55	5.22	5.88	0.89
15	Ph/NO_2	3.79	1	6	387.42	3	-4.52	0.35	4.76	5.71	0.83
16	Me/H	2.66	1	4	280.35	4	7.94	0.86	3.97	4.97	0.80
17	Me/Me	3.15	1	4	294.38	4	6.22	0.82	3.54	4.27	0.83
18	Me/Cl	3.32	1	4	314.79	4	8.46	0.77	3.66	4.36	0.84
19	Me/Br	3.41	1	4	359.24	4	5.68	0.74	3.67	4.31	0.85
20	Me/CN	2.54	1	5	305.36	4	-0.9	0.52	2.23	3.02	0.74
21	Me/NO_2	2.55	1	6	325.35	3	-2.6	0.44	2.49	3.21	0.78
22	Me/MeO	2.64	1	5	310.38	4	7.89	0.85	3.30	4.28	0.77
23	iPr/H	3.50	1	4	308.41	4	8.01	0.8	3.12	3.51	0.89
24	iPr/CN	3.38	1	5	333.42	4	-0.38	0.51	3.51	4.18	0.84
25	iPr/Br	4.25	1	4	387.30	4	6.2	0.64	5.11	5.86	0.87
26	iPr/Me	3.99	1	4	322.44	4	6.73	0.75	4.75	5.55	0.86
27	iPr/MeO	3.49	1	5	338.44	4	8.41	0.79	3.77	4.53	0.83
28	iPr/NO_2	3.40	1	6	353.41	3	-2.08	0.43	6.03	6.93	0.87
29	iPr/Cl	4.17	1	4	342.85	4	8.94	0.68	4.05	4.58	0.88
Cutoff values[a]		≤5	≤10	≤5	≤500						

[a]Values recommended for good absorption or permeability prediction.
* Acronyms for functional groups: Me for methyl, Et for ethyl, tBu for tert-butyl, Ph for phenyl, MeO for methoxy, CN for cyano, NO_2 for nitro, Cl for chloro, and Br for bromo.

Finally, drug score given in column 7 is a combination of drug-likeness, lipophilicity, solubility, molecular weight and toxicity. It takes value from 0 to 1 (a score of 1 indicates that a compound is a good drug candidate, whereas a score of 0 indicates that a compound will likely not be a drug). The lowest drug score have compound number **10** designating that this hydantoin is not an good orally bioavailable. Drug score higher of 0.8 was observed for the compounds **1**, **2**, **16**, **17**, **22** and **23** - all with moderate lipophilicity.

In order to explore and visualize the similarities and differences between the tested compounds in the space of molecular descriptors hierarchical clustering and principal component analysis were applied. For the HCA was apply by the Ward linkage method using Euclidean distance as the similarity measure. The resulting dendrogram is presented at Figure 2.

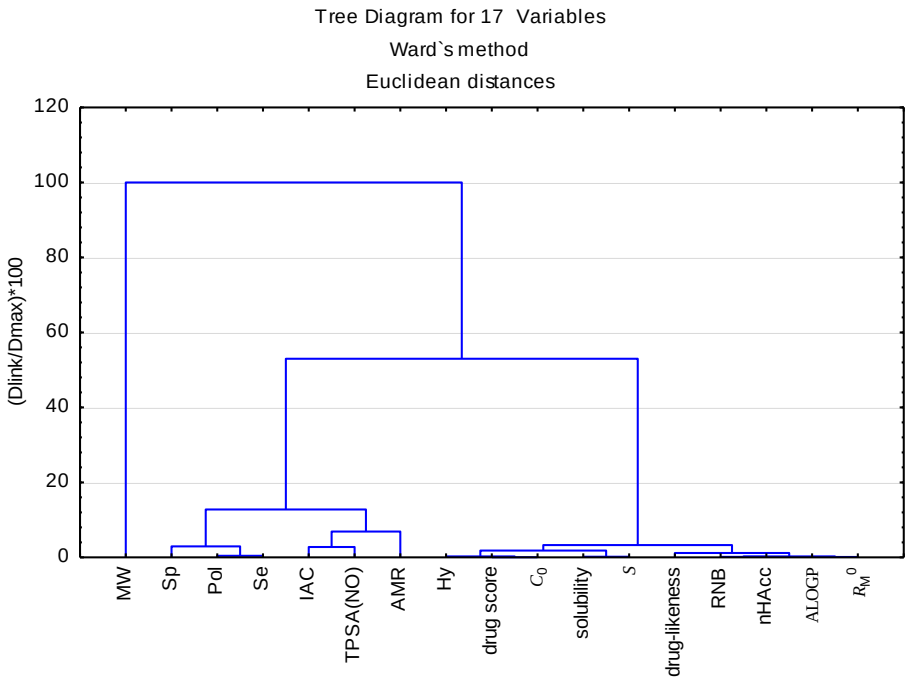

Figure 2. Dendrogram of studied molecular descriptors in the samples space.

The dendrogram constructed for the molecular descriptors in the samples space (see Figure 2b) reveals three main classes:

- Class A, containing of six variables divided in two subgroups: 1) Sp, Pol and Se; and 2) IAC, TPSA(NO) and AMR.
- Class B, contains two subgroups: 1) Hy, drug score, C_0, Solubility and S; 2) drug-likeness, RNB, nHAcc, ALOGP, and R_M^0.
- Class C, containing one molecular descriptor MW (molecular weight).

All three experimentally obtained parameters (R_M^0, S and C_0) that express lipophilicity were in the same class together with calculated ALOGP. Moreover, lipophiliicity either chromatographic or calculated is closely connected with the drug score and drug-likeness. This does not surprise because lipophilicity is a fundamental physicochemical property that plays a crucial role in the ADME features of the molecules. Same as drug score and drug-likeness, lipophilicity is not possible to evaluate using a single simple structural feature.

The presented dendrogram in Figure 2 reveal the data structure but do not enable interpretation of the relationships between the studied compounds and molecular descriptors. One of the most popular methods used in reduction of data dimensionality is principal components analysis. For the PCA 13 molecular descriptors were taken as variables for analysis while lipophilicity data (R_M^0, S, C_0 and ALOGP) were taken as supplementary variables.

The obtained PCA model has four significant principal components (PCs) that describe 97.5% of data variance. Score plots and loading plots are presented in Figure 3.

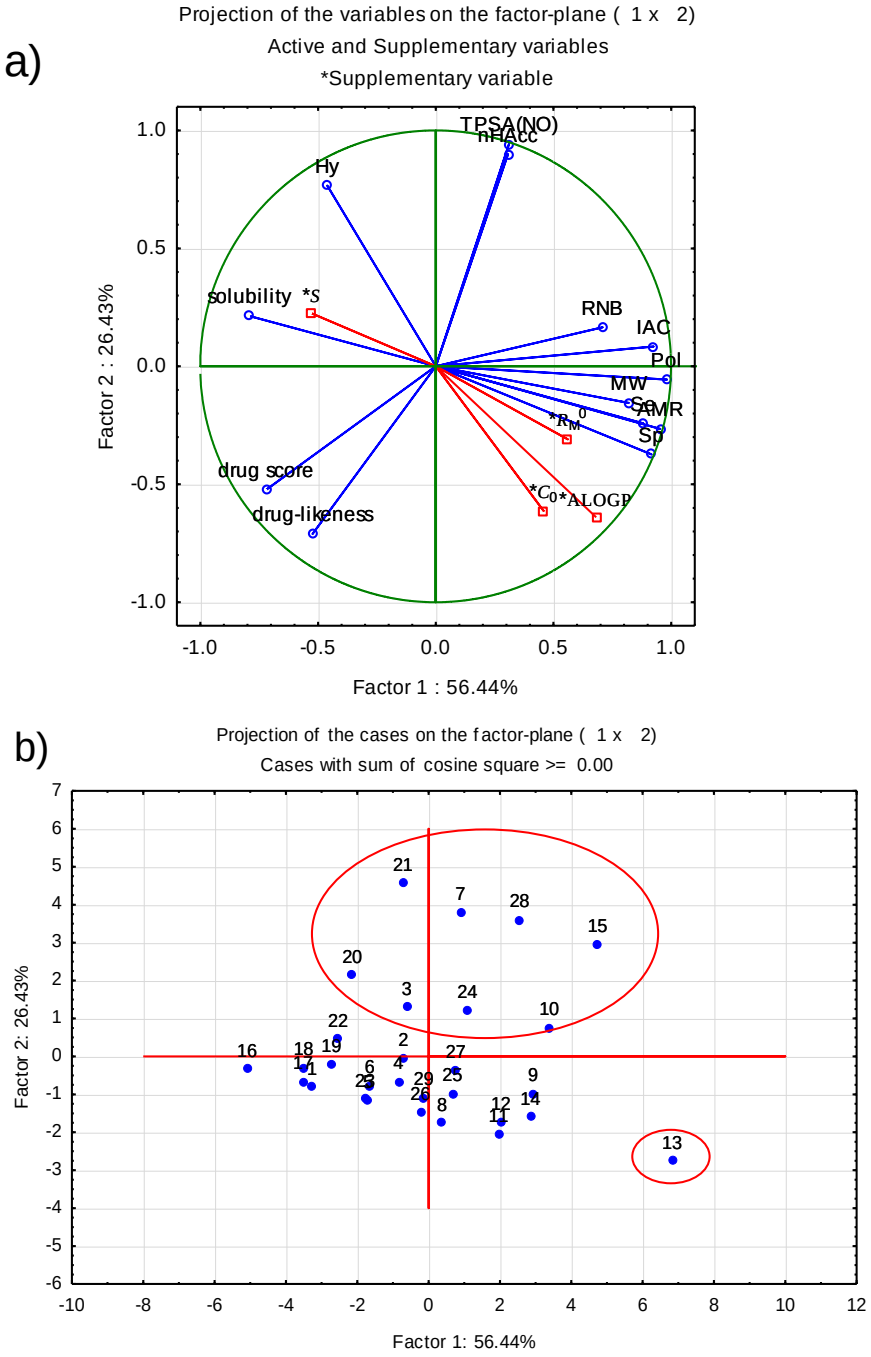

Figure 3. (a) Score and (b) loading plots of PCA.

From Figure 3 it can be seen that C_0 and ALOGP have similar scores. Additionally, similar scores were obtained for Solubility and *S*. According to HCA descriptors TPSA(NO) and nHAcc are classified in two different clusters but have are with very similar PCs.

Loading plot reveal that PC1 (describing 56.44% of the total data variance) reflects the difference between the compound no. **13** (the most lipophilic one) and all the remaining compounds. According to PC2 (describing 26.43% of the total data variance) was constructed mainly due to the differences between the hydantoins with R_2=NO_2 nos. **21**, **7**, **28** and hydantoins with R_2=CN (nos. **20**, **3**, **24** and **10**) from one side and the rest of investigated compounds.

5. Conclusions

Using hierarchical clustering and principal component analysis the molecular descriptors that are important for ADME and of interest for the pharmacokinetics of 29 newly synthesized derivative hydantoins were selected according to similarity. Recognition of the similarity between descriptors as well as between investigated compounds gives a starting point for understanding deeper levels of physiological processes—which may also provide access to new drugs and targets. In addition principal components analysis allows the visualization of hydantoins that are relevant to ADME property prediction using molecular descriptors space proposed by Lipinski.

6. Acknowledgments.

This paper was realized as a part of the project No. 174019 financed by the Ministry of Education, Science and Technological Development of the Republic of Serbia and grant number 114-451-3593/2013-02 supported by the Provincial Secretariat for Science and Technological Development of the Autonomous Province of Vojvodina.

References

[1] Ware E, *Chem. Rev*. 1950; **46**: 403.
[2] Lopez C, Trigo, G, The chemistry of hydantoins. Advances in heterocyclic chemistry; Katritzky AR, Ed.; Academic Press: New York, 1985; **38**, 177.
[3] Scholl S, Koch A, Henning D, Kempter G, Kleinpeter E, *Struct. Chem*. 1999; **10**: 355.
[4] Perucca E, *Br. J. Clin. Pharmacol*. 1996; **42**: 531.
[5] Martínez-López D, Yu M-L, García-Iriepa C, Campos PJ, Frutos LM, Golen JA, Rasapalli S, Sampedro D, *J. Org. Chem.*, 2015, **80**: 3929.
[6] Handzlik J, Bajda M, Zygmunt M, Maciąg D, Dybała M, Bednarski M, Filipek B, Malawska B, Kieć-Kononowicz , *Bioorg. Med. Chem*., 2012 , **20**: 2290.
[7] Ha YM, Kim JA, Park YJ, Park D, Kim JM, Chung KW, Lee EK, Park JY, Lee JY, Lee HJ, Yoon JH, Moon HR, Chung HY, *Biochim. Biophys. Acta*, 2011, **1810**: 612.
[8] Liu Y, Zhong W, Song L, *Chin. Chem. Lett.*, 2012, **23**:133.
[9] Lipinski CA, Lombardo F, Dominy BW, Feeney PJ, *Adv. Drug. Deliv. Rev.* 1997, **23**: 3.
[10] Hmuda SH, Trišović NP, Valentić NV, Uščumlić GS, *J. Solution Chem*. 2011, **40**: 307.
[11] Giaginis C, Tsantili-Kakoulidou A, *J. Liq. Chromatogr. Relat. Technol*. 2008, **31**: 79.
[12] Vieth M, Siegel MG, Higgs RE, Watson IA, Robertson DH, Savin KA, Durst GL, Hipskind PA, *J. Med. Chem.* 2004, **47**: 224.
[13] Proudfoot JR, *Bioorg. Med. Chem. Lett*. 2005, **15**: 1087.
[14] StatSoft Inc. STATISTICA (data analysis software system), v.12; 2012. Available from: www.statsoft.com.