Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters 15 (2005) 421–425
Bioorganic & Medicinal Chemistry Letters

Novel estimation of lipophilicity using ^{13}C NMR chemical shifts as molecular descriptor

Padmakar V. Khadikar,[a,*] Vimukta Sharma[b] and R. G. Varma[c]

[a]*Research Division, Laxmi Fumigation and Pest Control Pvt Ltd, 3, Khatipura, Indore 452 007, India*
[b]*Department of Pharmacy, Maharaja Ranjeetsingh College, Indore, India*
[c]*Department of Chemistry, PMB Gujarati Science College, Indore, India*

Received 8 July 2004; accepted 21 October 2004

Abstract—This paper describes the use of ^{13}C NMR chemical shift as molecular descriptor (molecular parameter) for modeling lipophilicity ($\log P$). A set of 32 alcohols were chosen for this purpose. The regression analysis of the data showed that ^{13}C NMR chemical shifts of these alcohols can be used as a molecular descriptor (molecular property) for modeling the lipophilicity ($\log P$). Better results are obtained by introducing an indicator parameter.

1. Introduction

Balasubramanian[1] has reviewed applications of combinatory and graph theory to spectroscopy. In the area of NMR spectroscopy these theories are very useful. Randic[2,3] has shown that the graph theoretical techniques could also be used to obtain the chemical shifts of the nuclei. They have developed[4] a computer code for listing equivalent classes of graphs, which works for most of the small, not transitive, and non-isospectral graphs. Furthermore, Duvenbeek[5] has discussed topological and geometral approaches to develop models for the prediction of ^{13}C NMR shifts and used $\sum$ ^{13}C NMR chemical shifts ($\sum$ $^{13}C_n$) as a molecular property.

It is well known that the presence of attached and nearby carbon atoms have a profound effect on ^{13}C NMR chemical shift. That is, topology of the molecule plays a dominant role in estimating ^{13}C NMR chemical shift.

Conversely, we can say that likewise ^{13}C NMR shifts also accounts for the topology of the molecules.

In our earlier communication,[6] while predicting ^{13}C NMR chemical shifts for 2,6-, and 2,7-disubstituted naphthalene, we have shown that $\sum$ ^{13}C can be used as a molecular property, which in turn can be modeled by both Wiener (W)[7] and Szeged (Sz)[8,9] indices. In still another communication[10] we have used PI (Padmakar–Ivan) index for predicting ^{13}C NMR chemical shifts of alkanes and cycloalkanes as well as their coupling constant.[11]

At this stage it is interesting to mention that although individual NMR chemical shifts for different atoms have received wide attention, it is somewhat surprising that there is hardly any study devoted to the collection of NMR chemical shifts for the purpose of drug modeling. It was shown that the average ^{13}C NMR chemical shifts of alkenes display regularity in isomeric variations. Such regulations are analogous to the isomeric variations of numerous thermodynamic properties of alkenes. Our results[6,10,11] also show that individual NMR chemical shifts can also be modeled topologically.

The aforementioned results prompted us that ^{13}C NMR chemical shift can be considered as a molecular property and that it can be used as a molecular descriptor for modeling physicochemical properties and biological activity of organic molecules acting as drugs. The results as discussed below show that this is found to be the case in the present study also.

Keywords: Chemical shift; ^{13}C NMR; Regression analysis; Molecular descriptor; QSAR.
* Corresponding author. Tel.: +91 731 2531906; fax: +91 731 2763618; e-mail: pvkhadikar@rediffmail.com

0960-894X/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2004.10.075

2. Results and discussion

The set of 32 alcohols, their ^{13}C NMR chemical shift, lipophilicity ($\log P$) and the assumed indicator parameters are given in Table 1. The details regarding this are given in the experimental section of this paper.

Application of regression analysis[12] showed that out of this set of 32 compounds, **2**, **6**, and **13** are outliers. They are, therefore, deleted from further regression analysis. At present we cannot provide convincing proof for their deletion and can consider it to be due to the regression procedure.

A comparison of correlation matrices for the set of 32 compounds (Table 2) and that for 29 compounds, after deleting compounds **2**, **6**, and **13** (Table 3), indicates that correlatedness amongst lipophilicity ($\log P$), ^{13}C NMR and indicator parameters (IP_1, IP_2, IP_3) is considerably improved for the new set of 29 compounds. Hence, the following results and discussion pertain to this set of 29 compounds. It is worthy to comment on the indicator parameters used. These are dummy parameters to account for those structural features not covered under the molecular descriptor used. They have only two values, 1 (when that structural feature is present in the structure) or 0 (when that structural feature is absent in the molecule). We have used $IP_1 = 1$ for the presence of a primary alcoholic group, in the absence of which $IP_1 = 0$. Similarly, IP_2 and IP_3 are taken as 1, for the presence of a secondary alcoholic group and a methyl group, respectively. Once again, in the absence of such structural features, both the indicator parameters IP_2 and IP_3 are 0 (zero).

A perusal of Table 3 shows that even in mono-parametric regression ^{13}C NMR chemical shift will be a good molecular descriptor for modeling lipophilicity ($\log P$) of the set of alcohols used. Also, that out of the three indicator parameters, the indicator parameter IP_1 will be more useful in obtaining multiparametric model(s). In view of this we have adopted maximum R^2-method[12] and carried out step-wise regression analysis. We have four correlating parameters: ^{13}C, IP_1, IP_2, and IP_3 and looking to the sample size and in accordance with the 'Rule of Thumb', we can go up to four-parametric (tetra-parametric) regression analysis. The results of simple regression, as well as those of step-wise regressions starting from mono-parametric up to tetra-parameteric regressions are given in Table 4.

In accordance with simple (mono-parametric) regression, the lipophilicity ($\log P$) can be modeled according to the following regression equation:

$$\log P = -3.9700 + 0.0774\ (\pm 0.0021)^{13}C$$
$$n = 29, \quad \mathrm{Se} = 0.5287, \quad R(r) = 0.7576,$$
$$F = 36.72, \quad Q = 1.4239 \qquad (1)$$

Table 1. Alcohols, their $\log P$, ^{13}C values, and indicator parameters (IP_1, IP_2, IP_3)

Compd no.	Name	$\log P$	^{13}C	I_1	I_2	I_3
1	Methanol	−0.764	49.0	1	1	0
2	Ethanol	−0.235	57.0	1	1	0
3	Propanol	0.294	63.6	1	1	0
4	Butanol	0.823	61.4	1	1	0
5	Pentanol	1.352	61.8	1	1	0
6	Hexanol	1.881	61.9	1	1	0
7	Isopropanol	0.154	63.4	1	1	0
8	2-Butanol	0.603	68.7	0	1	0
9	2-Pentanol	1.132	67.0	0	1	0
10	2-Hexanol	1.661	67.2	0	1	0
11	3-Pentanol	1.132	73.8	0	1	0
12	3-Hexanol	1.661	72.3	0	1	0
13	3-Heptanol	2.190	72.6	0	1	0
14	4-Heptanol	2.190	70.6	0	1	0
15	4-Octanol	2.680	70.9	0	1	0
16	5-Nonanol	1.572	71.1	0	0	0
17	Isobutanol	0.805	68.9	1	0	0
18	Tetra-butanol	0.532	68.4	1	0	0
19	Neopentanol	1.664	72.6	1	0	0
20	2-Methyl-pentanol	0.693	66.9	0	0	1
21	3-Methyl-butanol	1.280	60.2	0	0	1
22	3-Methyl-2-butanol	1.280	72.0	0	0	1
23	4-Methyl-2-butanol	1.687	65.2	0	0	1
24	4-Methyl-3-pentanol	1.687	77.3	0	0	1
25	3,3-Dimethyl-butanol	1.808	58.9	0	0	1
26	2,3-Dimethyl-2-butanol	1.529	72.2	0	0	1
27	3,3-Dimethyl-2-butanol	1.480	74.8	0	0	1
28	4,4-Dimethyl-3-butanol	2.154	80.9	0	0	1
29	2,4-Dimethyl-3-pentanol	2.148	80.4	0	0	1
30	2,3,3-Trimethyl-2-butanol	1.996	74.1	0	0	1
31	2,4,4-Trimethyl-3-pentanol	2.615	82.8	0	0	1
32	2,2,4,4-Tetramethyl-3-pentanol	3.082	84.7	0	0	1