EFFECT OF ETHYL GROUPS ON THE RELATION BETWEEN ZENKEVICH AND WIENER INDICES

Ivan Gutman and Boris Furtula

Faculty of Science, P. O. Box 60, 34000 Kragujevac, Serbia & Montenegro

ABSTRACT. The Zenkevich index U provides a measure of the intramolecular energy of an organic molecule. The Wiener index W represents the surface area of the same molecules. Thus, from a physico–chemical point of view, no connection between U and W would be anticipated. Yet, in the case of alkanes, these two quantities depend on the same structural features of the molecular graph and an approximate mathematical relation between them can be established. Within sets of isomeric alkanes, the relation between U and W is linear, the (U, W)-points forming several, mutually parallel, lines. Each such line pertains to a group of isomers possessing a fixed number of methyl and ethyl groups.

INTRODUCTION

In a recent paper [1] we pointed out that in the case of alkanes, the structure–descriptor U, introduced short time ago by Zenkevich [2–4], and the classical Wiener index W [5–7] are mutually related. From a physico–chemical point of view this is a surprising finding, because U is conceived as a (rough) measure of the intramolecular energy, whereas W has been recognized as a (rough) measure of the molecular surface area [8] and intermolecular van der Waals forces [9]. The reason for the existence of a relation between U and W lies in the fact that both quantities depend on the same structural features of the underlying molecular graph [1]. More specifically, if T is the graph representing an alkane molecule (in which case T necessarily is a tree [7]), then

$$W = \sum_e n_1(e|T)\, n_2(e|T) \tag{1}$$

and

$$U = \sum_e \sqrt{\frac{(C + 2H)n + 2H}{[(C + 2H)\, n_1(e|T) + H]\, [(C + 2H)\, n_2(e|T) + H]}} \tag{2}$$

where n is the number of vertices of T, and where H and C stand for the atomic masses of hydrogen and carbon ($H = 1.0$, $C = 12.0$). In Eqs. (1) and (2), $n_1(e|T)$ and $n_2(e|T)$ denote the number of vertices lying on the two sides of an edges e of T, whereas the summations go over all edges of the molecular graph T. Recall that an n-vertex tree possesses $n - 1$ edges, and for each of these edges,

$$n_1(e|T) + n_2(e|T) = n \ . \tag{3}$$

Bearing in mind (3), we can rewrite Eq. (2) as

$$U = A \sum_e \frac{1}{\sqrt{n_1(e|T)\, n_2(e|T) + B}} \tag{4}$$

where the quantities

$$A = \frac{\sqrt{(C + 2H)n + 2H}}{C + 2H} \tag{5}$$

and

$$B = \frac{H\, n}{C + 2H} + \left(\frac{H}{C + 2H}\right)^2 \tag{6}$$

depend only on the number n of vertices, and are independent of any other structural detail of the molecular graph under consideration.

Initially it was believed [1] that the combination of Eqs. (1) and (4) implies the existence of a linear correlation between U and W. However, subsequent studies [10, 11] revealed that the actual relation between U and W is significantly more complicated. If n is sufficiently large (say, $n \geq 10$), then the (U, W)-points form not a single, but a group of nearly parallel, close–lying straight lines. An illustrative example is shown in Figure 1.

THE METHYL GROUP RULE

By a detailed examination of the Zenkevich and Wiener indices and their mutual relation, the following regularity could be established [11], which we refer to as the:

Methyl Group Rule. *The (U, W)-points lying (approximately) on the same line pertain to isomeric alkanes with equal number of methyl groups; (U, V)-points corresponding to isomeric alkanes with different number of methyl groups lie on different lines.*

Figure 1. The Zenkevich index (U) versus the Wiener index (W) for the class of all isomeric alkanes with $n = 11$ carbon atoms.

The *Methyl Group Rule* was first discovered empirically, and was eventually corroborated by a pertinent mathematical analysis of Eqs. (1) and (4). In connection with it the following approximation was deduced [11]:

$$U \approx -\alpha\, W + \beta + \gamma_1\, \nu_1 \tag{7}$$

where ν_1 stands for the number of methyl groups, whereas the coefficients α, β, and γ_1 depend only on n, and are determined via

$$\alpha = \frac{A}{2}\,(\xi + B)^{-3/2} \tag{8}$$

$$\beta = \alpha\,(n-1)(3\,\xi + 2\,B) \tag{9}$$

$$\gamma_1 = \frac{A}{\sqrt{n-1+B}} - \alpha\,(3\,\xi + 2\,B - n + 1) \tag{10}$$

with the auxiliary parameter ξ defined as

$$\xi = \frac{1}{2}\left(2n - 4 + \lfloor n^2/4 \rfloor\right) \ . \tag{11}$$

In Table 1 are given the calculated values for the coefficients occurring in Eq. (7), for the chemically interesting values of n.

THE ETHYL GROUP RULE

Already at $n \geq 10$ an additional, much smaller, splitting of the (U, W)-lines becomes visible, illustrated in Figure 2.

Figure 2. The Zenkevich index (U) versus the Wiener index (W) for the class of all isomeric alkanes with $n = 14$ carbon atoms and with $\nu_1 = 4$ methyl groups.

By empirical checking it was established that, in the case of sufficiently large n, each line described by Eq. (7) is split into several, very closely lying, nearly parallel, lines. In this case the generally valid regularity is the following:

n	$\alpha \cdot 10^3$	β	γ_1
6	12.404	1.635	0.0191
7	9.152	1.868	0.0237
8	6.852	2.070	0.0294
9	5.450	2.280	0.0334
10	4.354	2.466	0.0382
11	3.614	2.659	0.0418
12	3.003	2.832	0.0460
13	2.563	3.010	0.0493
14	2.188	3.172	0.0530
15	1.905	3.338	0.0561
16	1.658	3.490	0.0594
17	1.466	3.647	0.0622
18	1.296	3.790	0.0653
19	1.159	3.938	0.0678
20	1.037	4.074	0.0706
21	0.937	4.215	0.0730
22	0.846	4.345	0.0755
23	0.770	4.478	0.0778
24	0.701	4.603	0.0801
25	0.643	4.731	0.0822
26	0.589	4.850	0.0844
27	0.544	4.973	0.0863
28	0.501	5.088	0.0883
29	0.465	5.206	0.0902
30	0.431	5.317	0.0920
31	0.401	5.431	0.0937
32	0.373	5.539	0.0955
33	0.349	5.649	0.0971
34	0.326	5.753	0.0988
35	0.306	5.859	0.1003
36	0.287	5.960	0.1019
37	0.270	6.063	0.1033
38	0.254	6.162	0.1048
39	0.240	6.262	0.1061
40	0.226	6.358	0.1075

Table 1. The parameters occurring in Eq. (7) for the first few values of n

Ethyl Group Rule. *The (U, W)-points lying (approximately) on the same line pertain to isomeric alkanes with equal number of methyl and ethyl groups; (U, V)-points corresponding to isomeric alkanes with equal number of methyl groups, but different number of ethyl groups lie on different lines.*

Our mathematical analysis (outlined in the subsequent section) leads to the approximate formula

$$U \approx -\alpha\, W + \beta + \gamma_1\, \nu_1 + \gamma_2\, \nu_2 \tag{12}$$

where ν_2 stands for the number of ethyl groups. The coefficients in Eq. (12) conform to the expressions (8)–(10) and (13)

$$\gamma_2 = \frac{A}{\sqrt{2n - 4 + B}} - \alpha\,(3\,\xi + 2\,B - 2n + 4) \tag{13}$$

in which the auxiliary parameter ξ instead of Eq. (11) is given by

$$\xi = \frac{1}{2}\left(3n - 9 + \lfloor n^2/4 \rfloor\right) . \tag{14}$$

In Table 2 are given the calculated values for the coefficients occurring in Eq. (12), for the chemically interesting values of n.

MATHEMATICAL ANALYSIS OF THE ETHYL GROUP RULE

Our starting point will be Eq. (4) in which the summation goes over all edges of the respective molecular graph. Recall that A and B are molecular–structure–independent parameters, defined by Eqs. (5) and (6). Aiming at the (empirically established) *Ethyl Group Rule*, the sum on the right–hand side of (4) is now divided into three parts: one pertaining to the edges of T whose one endpoint is a vertex of degree one (representing the $C\text{–}CH_3$ bond), the second pertaining to the edges of T whose one endpoint is a vertex of degree two adjacent to a vertex of degree one (representing the $C\text{–}C_2H_5$ bond), and the third embracing all other edges. There are ν_1 edges of the first kind, nu_2 edges of the second kind and $n - 1 - \nu_1 - \nu_2$ edges of the third kind.

For the edges of the first kind, $n_1(e|T) = 1$, $n_2(e|T) = n-1$ and their contribution to to right–hand side of (4) is

$$\frac{A\,\nu_1}{\sqrt{n - 1 + B}} .$$

For the edges of the second kind, $n_1(e|T) = 2$, $n_2(e|T) = n-2$ and their contribution to to right–hand side of (4) is

$$\frac{A\,\nu_2}{\sqrt{2n - 4 + B}} .$$

n	$\alpha \cdot 10^3$	β	γ_1	γ_2
6	11.431	1.593	0.0228	0.0010
7	8.076	1.793	0.0296	0.0022
8	5.916	1.973	0.0366	0.0041
9	4.639	2.164	0.0415	0.0056
10	3.684	2.336	0.0467	0.0076
11	3.045	2.514	0.0506	0.0092
12	2.529	2.677	0.0548	0.0111
13	2.158	2.846	0.0582	0.0127
14	1.846	3.000	0.0618	0.0145
15	1.610	3.159	0.0648	0.0161
16	1.405	3.306	0.0679	0.0179
17	1.246	3.456	0.0706	0.0194
18	1.104	3.596	0.0734	0.0211
19	0.991	3.740	0.0758	0.0226
20	0.889	3.873	0.0784	0.0242
21	0.805	4.010	0.0806	0.0256
22	0.729	4.138	0.0829	0.0272
23	0.666	4.269	0.0850	0.0285
24	0.608	4.393	0.0871	0.0300
25	0.559	4.519	0.0890	0.0313
26	0.514	4.637	0.0910	0.0327
27	0.476	4.758	0.0928	0.0340
28	0.440	4.873	0.0946	0.0353
29	0.409	4.990	0.0963	0.0365
30	0.380	5.101	0.0980	0.0378
31	0.354	5.214	0.0996	0.0390
32	0.331	5.321	0.1012	0.0402
33	0.310	5.431	0.1027	0.0413
34	0.290	5.535	0.1042	0.0425
35	0.273	5.641	0.1056	0.0435
36	0.256	5.742	0.1070	0.0446
37	0.242	5.845	0.1083	0.0457
38	0.228	5.944	0.1097	0.0467
39	0.216	6.044	0.1109	0.0477
40	0.204	6.140	0.1122	0.0487

Table 2. The parameters occurring in Eq. (12) for the first few values of n

In view of this,

$$U = \frac{A\,\nu_1}{\sqrt{n-1+B}} + \frac{A\,\nu_2}{\sqrt{2n-4+B}} + A \sum_{e}^{*} \frac{1}{\sqrt{n_1(e|T)\,n_2(e|T)+B}} \qquad (15)$$

where $\sum_{e}^{*}$ indicates summation over the $n-1-\nu_1-\nu_2$ edges of the third type.

The maximum value of the product $n_1(e|T)\,n_2(e|T)$ is

$$\frac{n}{2}\left(n - \frac{n}{2}\right) = \frac{n^2}{4}$$

if n is even, and

$$\frac{n-1}{2}\left(n - \frac{n-1}{2}\right) = \frac{n^2-1}{4}$$

if n is odd. In both cases it is equal to the largest integer not exceeding $n^2/4$, i. e., to $\lfloor n^2/4 \rfloor$. Therefore, for the edges included in the summation $\sum_{e}^{*}$, one has $3(n-3) \le n_1(e|T)\,n_2(e|T) \le \lfloor n^2/4 \rfloor$, i. e., the product $n_1(e|T)\,n_2(e|T)$ assumes values from the interval $(3n-9, \lfloor n^2/4 \rfloor)$.

In order to approximate the third term on the right–hand side of Eq. (15), we consider the function $1/\sqrt{x+B}$ and expand it into a power series around the point ξ given by formula (14). Clearly, ξ is chosen to be the center of the above specified interval $(3n-9, \lfloor n^2/4 \rfloor)$. This may not be the optimal choice for ξ, but for sure is the simplest.

Direct calculation yields

$$\frac{1}{\sqrt{x+B}} = (\xi+B)^{-1/2} - \frac{1}{2}(\xi+B)^{-3/2}(x-\xi) + \text{higher order terms}$$

which, after neglecting the higher order terms, and substitution back into Eq. (15) results in

$$U \approx \frac{A\,\nu_1}{\sqrt{n-1+B}} + \frac{A\,\nu_2}{\sqrt{2n-4+B}} + \frac{A}{2}\sum_{e}^{*}(3\,\xi+2\,B)(\xi+B)^{-3/2}$$

$$- \frac{A}{2}(\xi+B)^{-3/2}\sum_{e}^{*} n_1(e|T)\,n_2(e|T) \ . \qquad (16)$$

Now,

$$\sum_{e}^{*}(3\,\xi+2\,B)(\xi+B)^{-3/2} = (n-1-\nu_1-\nu_2)(3\,\xi+2\,B)(\xi+B)^{-3/2} \qquad (17)$$

and, in view of Eq. (1),

$$\sum_{e}^{*} n_1(e|T)\,n_2(e|T) = \sum_{e} n_1(e|T)\,n_2(e|T) - (n-1)\,\nu_1 - 2(n-2)\,\nu_2$$

$$= W - (n-1)\,\nu_1 - 2(n-2)\,\nu_2 \ . \qquad (18)$$

Substituting (17) and (18) back into (16) and making appropriate rearrangements, we arrive at Eq. (12) in which the coefficients α, β, γ_1, and γ_2 are given by the formulas (8), (9), (10), and (13), respectively, and ξ by formula (14).

BEYOND THE ETHYL GROUP RULE

Using the same type of reasoning as outlined in the preceding section we conclude that at still larger values of n additional splitting of the (U, W)-lines would happen, which could be described by means of Eq. (19)

$$U \approx -\alpha\,W + \beta + \gamma_1\,\nu_1 + \gamma_2\,\nu_2 + \gamma_3\,\nu_3 \tag{19}$$

in which ν_3 stands for the count of propyl and isopropyl groups, and the coefficients are determined by Eqs. (8)–(10), (13) and

$$\gamma_3 = \frac{A}{\sqrt{3n - 9 + B}} - \alpha\,(3\,\xi + 2\,B - 3n + 9)$$

in which

$$\xi = \frac{1}{2}\left(4n - 16 + \lfloor n^2/2 \rfloor\right)\ .$$

Numerical values of the coefficients occurring in Eq. (19), are given in Table 3.

The effect of propyl/isopropyl (and still larger) groups would be very small, would be encountered only at large values of n and would require the restriction of the consideration to sets of isomeric alkanes with fixed number of methyl and ethyl groups. Therefore, in our opinion, the effects that in the (U, W)-correlation go beyond the *Ethyl Group Rule* are not worth being taken into account.

DISCUSSION

As seen from Tables 2 and 3, $\gamma_1 > \gamma_2 > \gamma_3$ holds for all values of n. This implies that the effect of a methyl group is greater (usually 2–3 times) than the effect of an ethyl group, which in turn is greater (usually 2–3 times) as the effect of a propyl or isopropyl group. This is one reason why the effect of ethyl groups was more difficult to recognize than the effect of methyl groups, and why we decided to ignore the effects of propyl/isopropyl and larger groups. The other reason is the following.

It can be shown that the number of methyl groups in an alkane with n carbon atoms cannot be greater than $\lceil 2n/3 \rceil$, i. e.,

$$2 \le \nu_1 \le \left\lceil \frac{2}{3}n \right\rceil\ .$$

n	$\alpha \cdot 10^3$	β	γ_1	γ_2	γ_3
6	12.404	1.635	0.0191	0.0003	0.0002
7	8.076	1.793	0.0296	0.0022	0.0000
8	5.650	1.944	0.0389	0.0051	0.0003
9	4.307	2.112	0.0453	0.0075	0.0007
10	3.360	2.266	0.0514	0.0101	0.0014
11	2.743	2.430	0.0560	0.0122	0.0020
12	2.262	2.581	0.0606	0.0145	0.0029
13	1.920	2.738	0.0643	0.0164	0.0036
14	1.637	2.884	0.0680	0.0184	0.0045
15	1.425	3.034	0.0710	0.0202	0.0053
16	1.243	3.175	0.0742	0.0221	0.0063
17	1.101	3.319	0.0769	0.0237	0.0071
18	0.976	3.453	0.0796	0.0254	0.0081
19	0.876	3.591	0.0820	0.0269	0.0090
20	0.787	3.721	0.0845	0.0285	0.0099
21	0.714	3.854	0.0866	0.0299	0.0108
22	0.647	3.979	0.0888	0.0315	0.0118
23	0.592	4.106	0.0908	0.0328	0.0126
24	0.541	4.227	0.0928	0.0343	0.0136
25	0.499	4.350	0.0946	0.0355	0.0144
26	0.459	4.467	0.0965	0.0369	0.0153
27	0.425	4.586	0.0981	0.0381	0.0162
28	0.394	4.699	0.0999	0.0394	0.0171
29	0.366	4.814	0.1014	0.0405	0.0179
30	0.341	4.923	0.1030	0.0418	0.0188
31	0.319	5.035	0.1045	0.0429	0.0196
32	0.298	5.141	0.1060	0.0440	0.0204
33	0.280	5.249	0.1074	0.0451	0.0212
34	0.262	5.353	0.1088	0.0462	0.0221
35	0.247	5.458	0.1101	0.0472	0.0228
36	0.232	5.558	0.1114	0.0483	0.0236
37	0.220	5.661	0.1127	0.0492	0.0244
38	0.207	5.759	0.1139	0.0502	0.0252
39	0.196	5.859	0.1151	0.0512	0.0259
40	0.186	5.954	0.1163	0.0521	0.0266

Table 3. The parameters occurring in Eq. (19) for the first few values of n

These bounds put a limit on the number of different lines of the form (7); recall that there must exist at least two alkane isomers with equal ν_1 in order that the respective (U, W)-points form a "line". As a consequence of these bounds, in order to be able to visualize the *Methyl Group Rule* one has to consider alkane isomers with sufficiently large number n of carbon atoms, so that the parameter ν_1 can assume sufficiently many different values, and that there are sufficiently many alkane species with the same ν_1-value. In other words, in the case of alkane isomers with smaller n the *Methyl Group Rule* is difficult to be envisaged [1].

The situation with the *Ethyl Group Rule* is analogous. The number of ethyl groups in an alkane with n carbon atoms cannot be greater than $\lceil 2(n-1)/5 \rceil$, or, more preciely:

$$0 \leq \nu_2 \leq \left\lceil \frac{2}{5}(n-1) \right\rceil .$$

Therefore, in order to be able to visualize the *Ethyl Group Rule* one has to consider alkane isomers with larger number n of carbon atoms, so that within groups of alkane isomers with a given value of ν_1, the parameter ν_2 can assume sufficiently many different values, and that there are sufficiently many alkane species with equal ν_1 and equal ν_2. This may be the reason why the *Ethyl Group Rule* was not noticed in the research reported in the paper [11].

References

[1] I. Gutman, I. G. Zenkevich, Wiener index and vibrational energy, *Z. Naturforsch.* **57a** (2002) 824–828.

[2] I. G. Zenkevich, Dependence of chromatographic retention indices on the dynamic characteristics of molecules, *Russ. J. Phys. Chem.* **73** (1999) 797–801.

[3] I. G. Zenkevich, Vzaimnaya korrelyaciya gazohromatograficheskih indeksov uderzhivaniya nepredel'nyh i predel'nyh uglevodorodov s ispol'zovaniem metodov molekulyarnoi dinamiki, *Zh. Anal. Khim.* **55** (2000) 1091–1097.

[4] I. G. Zenkevich, Interpretacijya gazohromatograficheskih indeksov uderzhivaniya dlya ustanovleniya struktur izomernyh produktov radikal'nogo hlorirovaniya alkilarenov, *Zh. Org. Khim.* **37** (2001) 283–293.

[5] H. Wiener, Structural determination of paraffin boiling points, *J. Am. Chem. Soc.* **69** (1947) 17–20.

[6] A. A. Dobrynin, R. Entringer, I. Gutman, Wiener index of trees: theory and applications, *Acta Appl. Math.* **66** (2001) 211–249.

[7] I. Gutman, *Uvod u hemijsku teoriju grafova*, PMF Kragujevac, Kragujevac, 2003.

[8] I. Gutman, T. Körtvélyesi, Wiener indices and molecular surfaces, *Z. Naturforsch.* **50a** (1995) 669–671.

[9] I. Gutman, J. H. Potgieter, Wiener index and intermolecular forces, *J. Serb. Chem. Soc.* **62** (1997) 185–192.

[10] I. Gutman, D. Vidović, B. Furtula, I. G. Zenkevich, Wiener–type indices and internal molecular energy, *J. Serb. Chem. Soc.* **68** (2003) 401–408.

[11] I. Gutman, B. Furtula, B. Arsić, Ž. Bošković, On the relation between Zenkevich and Wiener indices of alkanes *J. Serb. Chem. Soc.* **69** (2004) 000–000.