STUDIA UBB. CHEMIA, LV, 4, 2010

# CORRELATING STUDY OF NEW MOLECULAR GRAPH DESCRIPTORS

**MONICA L. POP[a], MIRCEA V. DIUDEA[a*] AND ALEKSANDAR ILIĆ[b]**

**ABSTRACT.** A correlating study, using as independent variables the topological indices newly developed within the design of the super-index Cluj-Niš *CJN*, is performed on the set of octane isomers. Among the modeled properties, the boiling point BP, entropy S and total surface area TSA, gave best scores, in mono- to three-variable regressions, with respect to our novel descriptors. The most important result was the monovariate description of octane boiling points, which is the best result so far published in literature.

***Keywords:*** *Octanes, topological indices, QSPR.*

## INTRODUCTION

Alkanes represent an interesting class of compounds as a starting point for the application of molecular modeling procedures. Many properties of alkanes vary function of molecular mass or branching, and alkanes can be described by using a single type of (carbon) atom. There are properties well accounted by a single molecular descriptor, e.g., octane number MON, entropy S, molar volume MV, molar refraction MR, etc. Other properties, such as boiling point BP, heat of vaporization HV, total surface area TSA, partition coefficient LogP, density DENS, critical temperature CT, critical pressure CP, and heat of formation DHF, are notable exceptions, being not well modeled by a single parameter [1,2].

The purpose of the present report is to evaluate the relative performances of new topological indices developed within the design of super-index Cluj-Niš *CJN* [3] in relating the hydrocarbon molecular structures to a set of physical properties.

## LAYER MATRICES

Layer matrices have been proposed in connection with sequences of metrics: *DDS* (Distance Degree Sequence), *PDS* (Path Degree Sequence), and *WS* (Walk Sequence) [4-8]. They are built up on layer partitions in a graph.

[a] *Faculty of Chemistry and Chemical Engineering, "Babes-Bolyai" University, 400028 Cluj-Napoca, Romania, mpop@chem.ubbcluj.ro; diudea@gmail.com*

[b] *Faculty of Sciences and Mathematics, University of Niš, Višegradska 33, 18000 Niš, Serbia, aleksandari@gmail.com*

Let $G(v)_k$ be the $k^{th}$ layer of vertices $v$ lying at distance $k$, in the partition $G(i)$:

$$G(v)_k = \{v \mid d_{i,v} = k\} \quad (1)$$

$$G(i) = \{ G(v)_k \; ; k \in [0,1,..,ecc_i] \} \quad (2)$$

with $ecc_i$ being the *eccentricity* of $i$. The entries in the layer matrix (of a vertex property) **LM**, are defined as

$$\left[\mathbf{LM}\right]_{i,k} = \sum_{v \mid d_{i,v}=k} p_v \quad (3)$$

The layer matrix is a collection of the above defined entries:

$$\mathbf{LM}(G) = \{ [\mathbf{LM}]_{i,k} ; i \in V(G); k \in [0,1,..,d(G)] \} \quad (4)$$

with $d(G)$ standing for the diameter of the graph (*i.e.*, the largest distance in $G$

## SHELL MATRICES

The entries in the shell matrix **ShM** (of a vertex pair property) are defined as follows [8]:

$$\left[\mathbf{ShM}\right]_{i,k} = \sum_{v \mid d_{i,v}=k} [\mathbf{M}]_{i,v} \quad (5)$$

The shell matrix is a collection of the above defined entries:

$$\mathbf{ShM}(G) = \{ [\mathbf{ShM}]_{i,k} ; i \in \mathrm{V}(G); k \in [0,1,..,d(G)] \} \quad (6)$$

A shell matrix **ShM**($G$) will partition the entries of a square matrix according to the vertex (distance) partitions in the graph. It represents a true decomposition of the property collected by the info square matrix according to the contributions brought by vertex pairs pertaining to shells located at distance $k$ around each vertex.

The shell matrices were recently used as the basis in the calculation of super-index super index Cluj-Niš[3] *CJN* [3].

## SHELL-DEGREE-DISTANCE POLYNOMIALS

The Cramer product of the diagonal matrix of vertex degrees **D** with the distance **DI** matrix provides the matrix of degree distances [9] denoted **DDI**.

$$\mathbf{D}(G) \times \mathbf{DI}(G) = \mathbf{DDI}(G) \quad (7)$$

The above Cramer product is equivalent (gives the same half sum of entries) with the pair-wise (Hadamard) product of the vectors “row sum” *RS* in the Adjacency **A** and Distance **DI** matrices, respectively.

$$RS(\mathbf{A}) \bullet RS(\mathbf{DI}) = RS(\mathbf{DDI}) \tag{8}$$

Next, by applying the Shell operator, we obtain the matrix **ShDDI**, of which column half sums are just the coefficients of the corresponding Shell-polynomial.

$$P(\mathbf{ShDDI}, x) = \sum\nolimits_k p(G,k) \cdot x^k \tag{9}$$

An index, called *Cluj-Tehran CT*(**ShM**,*G*), with specified **M,** is defined as

$$CT(\mathbf{ShM}, G) = P'(\mathbf{ShM}, 1) + (1/2)P''(\mathbf{ShM}, 1) \tag{10}$$

where P' and P" refers to the polynomial firs and second derivative, respectively (here calculate dat x=1) [9].

## ECCENTRIC CONNECTIVITY INDEX

Sharma *et al*. [10] introduced a distance-based molecular structure descriptor, the eccentric connectivity index, which is defined as:

$$\xi^c(G) = \sum_{v \in V} \deg(v) \cdot \mathrm{ecc}(v) \tag{11}$$

The eccentric connectivity index was successfully used for mathematical models of biological activities of diverse nature. The index $\xi^c$ has been shown to give a high degree of predictability of pharmaceutical properties, and provide leads for the development of safe and potent anti-HIV compounds. The investigation of its mathematical properties started only recently, and has so far resulted in determining the extremal values and the extremal graphs, and also in a number of explicit formulas for the eccentric connectivity index of various products of graphs, several families of benzenoid graphs, zigzag and armchair hexagonal belts, nanotubes and nanotori [11].

## MODELING OCTANE PROPERTIES

To test the correlating ability of the descriptors derived from the degree-distance matrices (**ShDDI**$_k$) and Shell-polynomials, we focused attention to the set of octanes, as one of the benchmark-sets [12,13] in correlating studies by using topological indices. Among several properties tested, three ones (listed in Table 1) provided best scores: boiling point BP, entropy S and total surface area TSA, in mono- to three- variable regressions, by our novel descriptors.

The octane topological descriptors are given in Table 2. They include the numbers derived from the $ShM_1$ (the basic degree-distance matrix, in the Shell-matrix format), the first and second derivatives, $D_1$ and $DD_1$, of the corresponding Shell-polynomial, al level *k*=1 (equivalent to $A_1$) and the same descriptors at *k*=2 (equivalent to $A_2$ – the exponent referring to the remote adjacency rank) and also Cluj-Tehran $CT_1$ index and the eccentric connectivity index, denoted ECC in Table 2. Statistics are presented in Table 3.

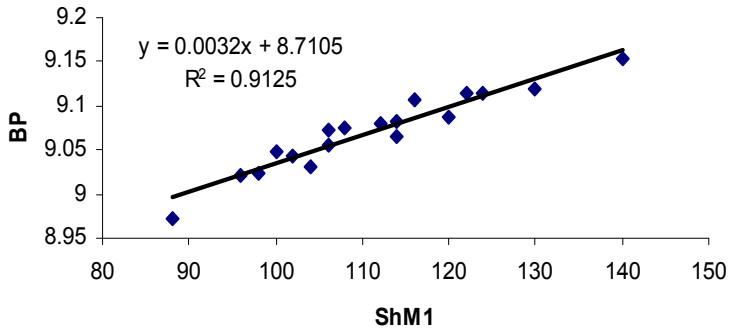
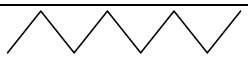
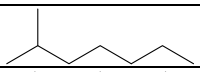
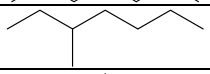
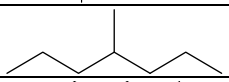
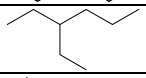
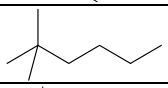
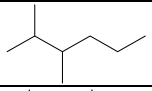
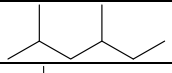
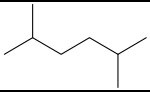



**Figure 1**. The plot of BP vs $ShM_1$ index

The best monovariate model of BP (by the index $ShM_1$) is plotted in Figure 1. Compare our results with the best results reported so far in literature (Table 3, entry 7) and remark the best result ($R^2$= 0.913) for the monovariate model of the octane boiling points. The other results in Table 3 (mono- or tri- variable models) can be considered as good (or acceptable) results.

**Table 1**. Octanes boiling point BP, entropy S and total surface area TSA values

| | Molecule | BP | S | TSA |
|---|---|---|---|---|
| 1 |  | 9.153 | 111.67 | 415.3 |
| 2 |  | 9.120 | 109.84 | 407.85 |
| 3 |  | 9.115 | 111.26 | 397.34 |
| 4 |  | 9.114 | 109.32 | 396.04 |
| 5 |  | 9.108 | 109.43 | 379.04 |
| 6 |  | 9.065 | 103.42 | 405.11 |
| 7 |  | 9.079 | 108.02 | 384.93 |
| 8 |  | 9.082 | 106.98 | 388.11 |
| 9 |  | 9.088 | 105.72 | 395.08 |

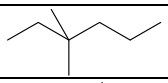
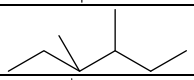
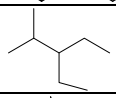
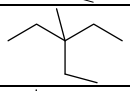
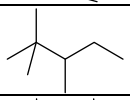
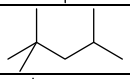
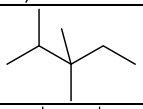
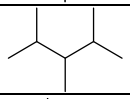
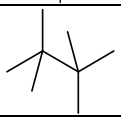
| | Molecule | BP | S | TSA |
|---|---|---|---|---|
| 10 |  | 9.056 | 104.74 | 389.79 |
| 11 |  | 9.074 | 106.59 | 376.91 |
| 12 |  | 9.073 | 106.06 | 368.1 |
| 13 |  | 9.049 | 101.48 | 366.99 |
| 14 |  | 9.023 | 101.31 | 371.75 |
| 15 |  | 9.031 | 104.09 | 392.19 |
| 16 |  | 9.020 | 102.06 | 377.4 |
| 17 |  | 9.044 | 102.39 | 368.93 |
| 18 |  | 8.971 | 93.06 | 390.47 |

**Table 2**. Topological Indices of Octanes

| Molecule | $ShM_1$ | $D_1$ | $DD_1$ | $CT_1$ | $ShM_2$ | $D_2$ | $DD_2$ | ECC |
|---|---|---|---|---|---|---|---|---|
| 1 | 140 | 532 | 1876 | 1470 | 258 | 972 | 3402 | 74 |
| 2 | 130 | 452 | 1406 | 1155 | 263 | 921 | 2900 | 65 |
| 3 | 124 | 404 | 1148 | 978 | 248 | 804 | 2270 | 63 |
| 4 | 122 | 388 | 1066 | 921 | 243 | 765 | 2070 | 61 |
| 5 | 116 | 340 | 808 | 744 | 230 | 668 | 1572 | 54 |
| 6 | 114 | 340 | 854 | 767 | 269 | 815 | 2088 | 56 |
| 7 | 112 | 320 | 746 | 693 | 244 | 704 | 1658 | 54 |
| 8 | 114 | 332 | 790 | 727 | 250 | 740 | 1796 | 54 |
| 9 | 120 | 376 | 998 | 875 | 266 | 856 | 2356 | 56 |
| 10 | 106 | 284 | 602 | 585 | 247 | 663 | 1390 | 52 |
| 11 | 108 | 292 | 620 | 602 | 234 | 636 | 1350 | 52 |
| 12 | 106 | 276 | 538 | 545 | 230 | 606 | 1202 | 43 |
| 13 | 100 | 244 | 432 | 460 | 232 | 568 | 996 | 41 |

| Molecule | $ShM_1$ | $D_1$ | $DD_1$ | $CT_1$ | $ShM_2$ | $D_2$ | $DD_2$ | ECC |
|---|---|---|---|---|---|---|---|---|
| 14 | 98 | 236 | 414 | 443 | 250 | 624 | 1138 | 43 |
| 15 | 104 | 272 | 546 | 545 | 268 | 732 | 1552 | 45 |
| 16 | 96 | 224 | 370 | 409 | 244 | 588 | 1000 | 41 |
| 17 | 102 | 256 | 476 | 494 | 243 | 633 | 1236 | 43 |
| 18 | 88 | 184 | 246 | 307 | 262 | 592 | 876 | 34 |

**Table 3**. Statistics ($R^2$) of QSPR Study on Octanes

| | Descriptors | BP | S | TSA |
|---|---|---|---|---|
| 1 | $ShM_1$ | 0.913 | 0.771 | 0.520 |
| 2 | $D_1$ | 0.863 | | |
| 3 | $CT_1$ | 0.818 | 0.646 | 0.613 |
| 4 | ECC | 0.854 | 0.747 | 0.583 |
| 5 | ShM2&ECC | 0.938 | 0.873 | 0.902 |
| 6 | $ShM_1$&$D_1$&$DD_1$ | 0.987 | 0.933 | 0.914 |
| 7 | $ShM_2$&$D_2$&$DD_2$ | 0.991 | 0.924 | 0.870 |
| 8 | Best in Octanes (monovariable) | $(0.78)^{13}$ $(0.77)^{14}$ | $(0.92)^{13}$ $(0.93)^{14}$ $(0.950)^{1}$ | $(0.72)^{13}$ $(0.92)^{14}$ |

## CONCLUSIONS

A correlating study, using as independent variables the topological indices newly developed within the design of the super-index Cluj-Niš *CJN*, was performed on the set of octane isomers. The modeled properties were: boiling point BP, entropy S and total surface area TSA, which gave best scores, among several properties, in mono- to three-variable regressions, with respect to our novel descriptors. The most important result was the monovariate description of octane boiling points, which is the best result so far published in literature.

### ACKNOWLEDGMENTS

Monica L. Pop thanks for the financial support provided from programs co-financed by The SECTORAL OPERATIONAL PROGRAMME HUMAN RESOURCES DEVELOPMENT, Contract **POSDRU 6/1.5/S/3** – „Doctoral studies: through science towards society". Aleksandar Ilić is grateful for the ESMC support in attending the ICAM-ESMC Conference in Cluj, 2010.

## REFERENCES


1. D. Stevanović, A. Ilić, C. Onisor, and M.V. Diudea, *Acta Chim, Slov*., **2009**, *56*, 410.
2. M.V. Diudea (Ed.), *QSPR/QSAR Studies by Molecular Descriptors*, **NOVA**, New York, 2001.
3. M.V. Diudea, A. Ilić, K. Varmuza, M. Dehmer, *Complexity*, **2010** (in press).
4. M.V. Diudea, I. Gutman; L. Jäntschi, *Molecular Topology*, **NOVA**, New York, 2002.
5. M.V. Diudea, M.S. Florescu, and P.V. Khadikar, *Molecular Topology and Its Applications*, **EFICON**, Bucharest, 2006.
6. M.V. Diudea, *J. Chem. Inf. Comput. Sci*., **1994**, *34*, 1064.
7. M.V. Diudea, M.I. Topan, and A. Graovac, *J. Chem. Inf. Comput. Sci*., **1994**, *34*, 1072.
8. M.V. Diudea and O. Ursu, *Indian J. Chem*., **2003**, *42A*, 1283.
9. M.V. Diudea, Nanomolecules and Nanostructures - Polynomials and Indices, MCM, No. 10, Univ. Kragujevac, Serbia, 2010.
10. V. Sharma, R. Goswami, and A.K. Madan, *J. Chem. Inf. Comput. Sci.,* **1997**, *37*, 273.
11. A. Ilić, in: I. Gutman, B. Furtula, Novel Molecular Structure Descriptors – Theory and Applications II, MCM Vol. 9, University of Kragujevac, 2010.
12. Milano Chemometrics & QSAR Research Group, Molecular Descriptors Dataset (available at http://www.moleculardescriptors.eu/dataset/dataset.htm)
13. http://www.moleculardescriptors.eu/dataset/dataset.htm; **http://www.iamc-online.org/**.
14. D. Vukičević, *Bond Additive Modeling. Adriatic Indices – Overview of the Results,* in I. Gutman and B. Furtula (Eds.), Novel Molecular Structure Descriptors - Theory and Applications II, MCM, Kragujevac, 2010, p. 269.