

**TOPOLOGICAL INDICES DERIVED FROM THE G(a,b,c) MATRIX,
USEFUL AS PHYSICOCHEMICAL PROPERTY INDICES****Anton Perdih,* Branislav Perdih***Mala vas 12, SI-1000 Ljubljana, Slovenia**Received 05-07-2004***Abstract**

From the $G(a,b,c)$ matrix at least 20 groups of topological indices can be derived. Each of these groups contains an infinite number of possible indices. Some of these indices correlate by $\text{abs}(r)_{\max} > 0.99$, and a number of them by $\text{abs}(r)_{\max} > 0.9$ with particular tested 29 resp. 31 physicochemical properties of alkanes. Therefore, their use in QSPR/QSAR models is recommended. The indices of these groups are not useful to predict the physicochemical properties: dc, Zc, A, Cohesive Energy Density, and Solubility Parameter.

Key words: alkane, index, matrix, QSAR, QSPR, topological

Introduction

Mathematical topological methods occupy an eminent place in the field of prediction of properties and activities of chemical compounds, and even materials. These methods, known under the acronym QSPR/QSAR (quantitative-structure-property or structure-activity relationship) are normally, but not always, based on graph-theoretical descriptors, where molecules are seen as chemical graphs, i.e. as a set of vertices attached to each other by a set of non-metrical connections.¹ These descriptors are known also as topological indices. They are the simplest means of describing the structure of a molecule, characterizing it by a simple number.² A huge number of topological indices are known^{3,4} but in spite of that, interest in topological indices has grown remarkably during recent years.

A substantial part of topological indices is derived from one or another matrix associated with molecular structure. Estrada⁵ developed a matrix that enables the derivation of an infinite number of indices. We⁶ presented a type of matrices, i.e. the generalized vertex-degree vertex-distance matrices, that enable the derivation of an infinite number of indices, too, and we have shown that these matrices represent a step in unification of several matrices which have been used to derive topological indices, i.e. of the adjacency matrix, the distance matrix, the reciprocal distance matrix, etc. The

characteristics of some groups of indices derived by means of the generalized vertex-degree vertex-distance matrices have been studied, namely:

The summation derived $W(m,n)$ indices,⁶ which besides the Wiener index⁷ W also include some Ivanciuc indices.⁸ The summation derived "mean degree of vertices" indices,⁹ which include the Randić index χ .¹⁰ The summation derived $V_{ij}(m,n)$ indices,¹¹ which besides the Wiener index⁷ and several Ivanciuc indices^{8,12-16} also include the Randić index χ .¹⁰ The susceptibilities for branching of the $W(m,n)$ indices,¹⁷ and the difference derived indices.¹⁸ The largest eigenvalues of the matrices from which the $V_{ij}(m,n)$ indices¹¹ have been derived. One index of this group has a long tradition, namely λ_1 .^{19,20} It is also a member of the indices²¹⁻²³ which are the largest eigenvalues of the matrices from which the $W(m,n)$ indices⁶ or the "mean degree of vertices" indices⁹ are derived. The multiplication derived indices, which include the Gutman index π .²⁴ They are derived by multiplication of the non-diagonal elements of the matrix.²⁵

The aim of the present study is to give a survey of the large number of indices groups that can be derived from the above mentioned generalized vertex-degree vertex-distance matrices and especially to indicate where in these groups of indices, consisting of infinite numbers of them, are to be expected the best indices, useful for the prediction of properties or activities of chemical compounds or materials. These indices could then be used in methods known under the acronym QSPR/QSAR (quantitative-structure-property or structure-activity relationship). The aim of the present study is not to look for the best QSPR/QSAR models but to indicate, using a set of physicochemical properties of alkanes, how and where the best indices useful for that can be looked for.

Data and notations

Physicochemical properties (PP)

The data for the boiling point (BP), density (d), the critical data T_c , P_c , V_c , Z_c , α_c , and d_c , the standard enthalpy of formation for the ideal gas (ΔH_f°), the enthalpy of vaporisation (ΔH_v), the Antoine constants A , B , and C , as well as the Pitzer's acentric factor (ω) and the refractive index (n_D) were taken from the CRC Handbook²⁶ or from Lange's Handbook²⁷. The data for the liquid molar volume (V_m), the intrinsic molar volume (V_i), the intermolecular volume (V''), the ratios T_c^2/P_c and T_c/P_c used instead of the van der Waals parameters a_0 and b_0 , the ratio BP/ T_c (reduced BP), the molar

refraction (MR), cohesive energy density (CED) and its square root, the solubility parameter (Sol. par.) were calculated from data presented in the handbooks. The data for Octane Numbers (BON, MON, RON) were taken from: Pogliani,²⁸ Balaban and Motoc,²⁹ Gutman et al.,²⁴ and Morley,³⁰ those for vapour pressure (logVP) from Goll and Jurs,³¹ and those for the entropy (S) and quadratic mean radius (R^2) from Ren.³² Surface tension (ST) data were taken from Li.³³

Notations

The structures of alkanes are presented in shorthand, e.g. Hp is n-heptane, Oct is n-octane, 223M5 is 2,2,3-trimethylpentane, 3E2M5 is 3-ethyl-2-methylpentane, etc., cf. also Fig. 1.

Matrix and indices

Let there be the matrix $\mathbf{G}(a,b,c)$, composed of elements $\mathbf{g} = v_i^a \times v_j^b \times d_{ij}^c$, where v_i and v_j is the degree of vertex i and j , respectively (in alkanes it is the number of C-C bonds the carbon in question is involved in) and d_{ij} is the shortest distance from vertex i to vertex j (in alkanes it is the smallest number of bonds between the carbons in question).

2,3-dimethyl pentane:	$ \begin{array}{c} 6 \quad 7 \\ \quad \\ 1-2-3-4-5 \equiv 23M5 \end{array} $
-----------------------	--

v_i	distance matrix	$\mathbf{G}(a,b,c)$ matrix
1	0 1 2 3 4 2 3	0 3^b $3^b 2^c$ $2^b 3^c$ 4^c 2^c 3^c
3	1 0 1 2 3 1 2	3^a 0 $3^a 3^b$ $3^a 2^b 2^c$ $3^a 3^c$ 3^a $3^a 2^c$
3	2 1 0 1 2 2 1	$3^a 2^c$ $3^a 3^b$ 0 $3^a 2^b$ $3^a 2^c$ $3^a 2^c$ 3^a
2	3 2 1 0 1 3 2	$2^a 3^c$ $2^a 3^b 2^c$ $2^a 3^b$ 0 2^a $2^a 3^c$ $2^a 2^c$
1	4 3 2 1 0 4 3	4^c $3^b 3^c$ $3^b 2^c$ 2^b 0 4^c 3^c
1	2 1 2 3 4 0 3	2^c 3^b $3^b 2^c$ $2^b 3^c$ 4^c 0 3^c
1	3 2 1 2 3 3 0	3^c $3^b 2^c$ 3^b $2^b 2^c$ 3^c 3^c 0
v_j	1 3 3 2 1 1 1	

Figure 1. 2,3-dimethyl pentane: its formula, its label, its distance matrix composed of distances d_{ij} , its vectors of degrees of vertices, v_i and v_j , as well as the $\mathbf{G}(a,b,c)$ matrix derived from them.

v_i - vector of degrees of starting vertex, v_i

v_j - vector of degrees of target vertex, v_j

normal script: matrix elements, *common* to both halves of the matrix

bold: matrix elements, **not** common to both halves of the matrix

Vertex i is the starting vertex whereas vertex j is the target vertex; \mathbf{a} , \mathbf{b} , and \mathbf{c} are the exponents. For the purpose of this study, the value of diagonal elements in the matrix $\mathbf{G}(\mathbf{a},\mathbf{b},\mathbf{c})$ is set to $\mathbf{g}_{ii} = 0$, since $\mathbf{d}_{ii} = 0$ and the contribution of hydrogen atoms is set to zero, whereas the contribution of all other atoms is set to one, i.e. only the contribution of the structure of the molecule in question to the value of the matrix and its elements is considered.

As an example, the $\mathbf{G}(\mathbf{a},\mathbf{b},\mathbf{c})$ matrix of the 2,3-dimethyl pentane structure is presented in Figure 1. To derive the indices from the $\mathbf{G}(\mathbf{a},\mathbf{b},\mathbf{c})$ matrix, one or both halves of the matrix or only some parts of them can be used. Indices derived by summation or multiplication of non-diagonal elements of the $\mathbf{G}(\mathbf{a},\mathbf{b},\mathbf{c})$ matrix are considered here. The definitions of indices studied here are presented in Table 1.

Table 1. Definitions of indices studied in the present paper.

$V_L(\mathbf{a},\mathbf{b},\mathbf{c})$	$= \sum(v_i^{\mathbf{a}} \times v_j^{\mathbf{b}} \times d_{ij}^{\mathbf{c}})_{\text{left}}$
$V_D(\mathbf{a},\mathbf{b},\mathbf{c})$	$= \sum(v_i^{\mathbf{a}} \times v_j^{\mathbf{b}} \times d_{ij}^{\mathbf{c}})_{\text{right}}$
$V(\mathbf{a},\mathbf{b},\mathbf{c})$	$= V_L(\mathbf{a},\mathbf{b},\mathbf{c}) + V_D(\mathbf{a},\mathbf{b},\mathbf{c})$
$\Delta V(\mathbf{a},\mathbf{b},\mathbf{c})$	$= V_L(\mathbf{a},\mathbf{b},\mathbf{c}) - V_D(\mathbf{a},\mathbf{b},\mathbf{c})$
$PV(\mathbf{a},\mathbf{b},\mathbf{c})$	$= V_L(\mathbf{a},\mathbf{b},\mathbf{c}) \times V_D(\mathbf{a},\mathbf{b},\mathbf{c})$
$QV(\mathbf{a},\mathbf{b},\mathbf{c})$	$= V_L(\mathbf{a},\mathbf{b},\mathbf{c}) / V_D(\mathbf{a},\mathbf{b},\mathbf{c})$
$VC(\mathbf{a},\mathbf{b},\mathbf{c})$	$= \sum(v_i^{\mathbf{a}} \times v_j^{\mathbf{b}} \times d_{ij}^{\mathbf{c}}$ elements, <i>common</i> to both halves of the matrix)
$VN_L(\mathbf{a},\mathbf{b},\mathbf{c})$	$= \sum(v_i^{\mathbf{a}} \times v_j^{\mathbf{b}} \times d_{ij}^{\mathbf{c}}$ elements, found only on the <i>left</i> side of the main diagonal of the matrix)
$VN_D(\mathbf{a},\mathbf{b},\mathbf{c})$	$= \sum(v_i^{\mathbf{a}} \times v_j^{\mathbf{b}} \times d_{ij}^{\mathbf{c}}$ elements, found only on the <i>right</i> side of the main diagonal of the matrix)
$VN(\mathbf{a},\mathbf{b},\mathbf{c})$	$= VN_L(\mathbf{a},\mathbf{b},\mathbf{c}) + VN_D(\mathbf{a},\mathbf{b},\mathbf{c})$
$PVN(\mathbf{a},\mathbf{b},\mathbf{c})$	$= VN_L(\mathbf{a},\mathbf{b},\mathbf{c}) \times VN_D(\mathbf{a},\mathbf{b},\mathbf{c})$
$QVN(\mathbf{a},\mathbf{b},\mathbf{c})$	$= VN_L(\mathbf{a},\mathbf{b},\mathbf{c}) / VN_D(\mathbf{a},\mathbf{b},\mathbf{c})$
$P_L(\mathbf{a},\mathbf{b},\mathbf{c})$	$= \prod(v_i^{\mathbf{a}} \times v_j^{\mathbf{b}} \times d_{ij}^{\mathbf{c}})_{\text{left}}$
$P_D(\mathbf{a},\mathbf{b},\mathbf{c})$	$= \prod(v_i^{\mathbf{a}} \times v_j^{\mathbf{b}} \times d_{ij}^{\mathbf{c}})_{\text{right}}$
$M(\mathbf{a},\mathbf{b},\mathbf{c})$	$= P_L(\mathbf{a},\mathbf{b},\mathbf{c}) \times P_D(\mathbf{a},\mathbf{b},\mathbf{c})$
$QP(\mathbf{a},\mathbf{b},\mathbf{c})$	$= P_L(\mathbf{a},\mathbf{b},\mathbf{c}) / P_D(\mathbf{a},\mathbf{b},\mathbf{c})$
$VP(\mathbf{a},\mathbf{b},\mathbf{c})$	$= P_L(\mathbf{a},\mathbf{b},\mathbf{c}) + P_D(\mathbf{a},\mathbf{b},\mathbf{c})$
$\Delta P(\mathbf{a},\mathbf{b},\mathbf{c})$	$= P_L(\mathbf{a},\mathbf{b},\mathbf{c}) - P_D(\mathbf{a},\mathbf{b},\mathbf{c})$
$PC(\mathbf{a},\mathbf{b},\mathbf{c})$	$= \prod(v_i^{\mathbf{a}} \times v_j^{\mathbf{b}} \times d_{ij}^{\mathbf{c}}$ elements, <i>common</i> to both halves of the matrix)
$PN_L(\mathbf{a},\mathbf{b},\mathbf{c})$	$= \prod(v_i^{\mathbf{a}} \times v_j^{\mathbf{b}} \times d_{ij}^{\mathbf{c}}$ elements, found only on the <i>left</i> side of the main diagonal of the matrix)
$PN_D(\mathbf{a},\mathbf{b},\mathbf{c})$	$= \prod(v_i^{\mathbf{a}} \times v_j^{\mathbf{b}} \times d_{ij}^{\mathbf{c}}$ elements, found only on the <i>right</i> side of the main diagonal of the matrix)
$PN(\mathbf{a},\mathbf{b},\mathbf{c})$	$= PN_L(\mathbf{a},\mathbf{b},\mathbf{c}) \times PN_D(\mathbf{a},\mathbf{b},\mathbf{c})$
$VPN(\mathbf{a},\mathbf{b},\mathbf{c})$	$= PN_L(\mathbf{a},\mathbf{b},\mathbf{c}) + PN_D(\mathbf{a},\mathbf{b},\mathbf{c})$
$\Delta PN(\mathbf{a},\mathbf{b},\mathbf{c})$	$= PN_L(\mathbf{a},\mathbf{b},\mathbf{c}) - PN_D(\mathbf{a},\mathbf{b},\mathbf{c})$

Expression of indices

The values of indices derived from the $G(a,b,c)$ matrix can be expressed analytically, Table 2. The expressions for multiplication-derived indices are in general simple, whereas for the summation-derived ones they are more complicated.

Table 2. Analytical expressions of some indices derived from the $G(a,b,c)$ matrix for the structure of 2,3-dimethyl pentane (23M5).

Summation-derived indices

$$V_L(a,b,c) = 2^a 3^c + 2^b(1+2^c+3^c) + 3^a(1+2^c) + 3^b(2+3 \times 2^c+3^c) + 2^a 3^b(1+2^c) + 3^a 3^b + 2^c + 3 \times 3^c + 2 \times 4^c$$

$$V_D(a,b,c) = 2^a(1+2^c+3^c) + 2^b 3^c + 3^a(2+3 \times 2^c+3^c) + 3^b(1+2^c) + 3^a 2^b(1+2^c) + 3^a 3^b + 2^c + 3 \times 3^c + 2 \times 4^c$$

$$V(a,b,c) = (2^a+2^b)(1+2^c+2 \times 3^c) + (3^a+3^b)(3+4 \times 2^c+3^c) + (2^a 3^b+3^a 2^b)(1+2^c) + 2(3^a 3^b+2^c+3 \times 3^c+2 \times 4^c)$$

$$\Delta V(a,b,c) = (2^b-2^a)(1+2^c) + (3^b-3^a)(1+2 \times 2^c+3^c) + (2^a 3^b-3^a 2^b)(1+2^c)$$

$$VC(a,b,c) = 3^c(2^a+2^b) + (3^a+3^b)(1+2^c) + 3^{a+b} + 2^c + 3 \times 3^c + 2 \times 4^c$$

$$VN_L(a,b,c) = 2^b(1+2^c) + 3^b(1+2 \times 2^c+3^c) + 2^a 3^b(1+2^c)$$

$$VN_D(a,b,c) = 2^a(1+2^c) + 3^a(1+2 \times 2^c+3^c) + 3^a 2^b(1+2^c)$$

$$VN(a,b,c) = (2^a+2^b)(1+2^c) + (2^a 3^b+3^a 2^b)(1+2^c) + (3^a+3^b)(1+2 \times 2^c+3^c)$$

Multiplication-derived indices

$$P_L(a,b,c) = 2^{3a+3b+11c} 3^{3(a+3b+2c)}$$

$$P_D(a,b,c) = 2^{3a+3b+11c} 3^{3(3a+b+2c)}$$

$$M(a,b,c) = 2^{2(3a+3b+11c)} 3^{12(a+b+c)}$$

$$VP(a,b,c) = 2^{3a+3b+11c} 3^{3(a+b+2c)} (3^{6b}+3^{6a})$$

$$\Delta P(a,b,c) = 2^{3a+3b+11c} 3^{3(a+b+2c)} (3^{6b}-3^{6a})$$

$$QP(a,b,c) = 3^{6(b-a)}$$

$$PC(a,b,c) = 2^{a+b+7c} 3^{3a+3b+5c}$$

$$PN_L(a,b,c) = 2^{2a+2b+4c} 3^{6b+c}$$

$$PN_D(a,b,c) = 2^{2a+2b+4c} 3^{6a+c}$$

$$PN(a,b,c) = 2^{4(a+b+2c)} 3^{6a+6b+2c}$$

$$VPN(a,b,c) = 2^{2a+2b+4c} 3^c (3^{6b}+3^{6a})$$

$$\Delta PN(a,b,c) = 2^{2a+2b+4c} 3^c (3^{6b}-3^{6a})$$

Characteristics of indices

The characteristics of several groups of these indices, as well as the presentation which known indices belong to which group of the indices presented here, have been given elsewhere.^{6,9,11,17,18,21-23,25}

Secondary indices

From the indices considered above, several groups of secondary indices can be derived. We used the following ones, exemplified by those derived from the whole-matrix index $V(a,b,c)$:

$V(a,b,c)/(N^2-N)$, which is the average value of the non-diagonal element of the matrix, where N is the number of vertices in the graph, i.e. the carbon number of the alkane in question; $V(a,b,c)/N$, which is the mean row-sum or the mean column sum of

the non-diagonal matrix elements; original indices raised to several exponents, e.g. $V^2(a,b,c)$, $V^{1/2}(a,b,c)$, etc., furthermore $V^{N/\Sigma v}(a,b,c)$, where v is the degree of vertex in the graph, $V^{1/N}(a,b,c)$, or $V^{1/(N^2-N)}(a,b,c)$, as well as $\ln V(a,b,c)$. Note: (N^2-N) in the exponent means (N^2-N) .

Among the multiplication derived indices, exemplified here by the whole-matrix index $M(a,b,c)$, we tested their logarithm, $\ln M(a,b,c)$, the exponent $1/N$, which gives the geometric mean of the row products of the matrix, the exponent $1/(N^2-N)$, which gives the geometric mean of the non-diagonal elements of the matrix, and when feasible, several other exponents, e.g. 2, $1/2$, $1/3$, $1/10$, etc., or even $1/10^{10}$.

To derive secondary indices from the half-matrix indices, we used as a rule the exponents in the form $2/k$ or $k/2$, as applicable, instead of $1/k$ or k , where k means any number, including N or N^2-N .

Among the indices derived from other parts of the matrix, besides the above-mentioned ways to derive the secondary indices, the following approaches were used:

$VC(a,b,c)/NC$ is the average value of the non-diagonal elements common to both halves of the matrix, where NC is the *number* of elements *common* to both halves of the matrix.

$$NC = VC(0,0,0)$$

$PC^{1/NC}(a,b,c)$, on the other hand, is the geometric mean of the non-diagonal elements common to both halves of the matrix. Similarly, $VN_L(a,b,c)/NN$ is the average value of the non-diagonal elements not common to both halves of the matrix, where NN is the *number* of elements *not common* to both halves of the matrix.

$$NN = VN_L(0,0,0) = VN_D(0,0,0)$$

$PN_L^{1/NN}(a,b,c)$ is the geometric mean of the non-diagonal elements not common to both halves of the matrix.

Usefulness of indices

The aim of present paper is to screen the 20 groups of indices, each containing an infinite number of possible indices, and the secondary indices derived from them, for most promising indices and not to find the best QSPR models. Therefore, the usefulness of the indices derived from the $G(a,b,c)$ matrix was tested using a simple linear model of the form $PP = m \times I(a,b,c) + n$, where PP means the physicochemical property in question, and $I(a,b,c)$ means the index in question. To keep the results simple and

comparable, only one criterion of the goodness of relation is used, i.e. the correlation coefficient r . In most our cases maxima of the correlation coefficients coincide with minima of the standard error, therefore it is not given. The values of $\text{abs}(r) > 0.99$ are considered to be potentially useful.² For screening purposes, also $\text{abs}(r) > 0.9$ is indicated in some cases.

Results

Labelling the indices

Labelling the indices derived from the $\mathbf{G}(a,b,c)$ matrix is not a simple task. Some well-known indices, e.g. the Wiener index,⁷ W , or the Randić index,¹⁰ χ , can be unambiguously labelled by a single letter since their value is determined by the structure taken into account and by the way of their derivation, which is well defined. The indices derived from the $\mathbf{G}(a,b,c)$ matrix, on the other hand, form several groups, each one consisting of an infinite number of indices. The indices of each group are derived in a different way and/or from different parts of the source matrix and on their derivation several data are used, i.e. the degrees of those two vertices which are in question at a particular time, the distance between them, as well as the three exponents to which the degrees of vertices and the distance between them are raised. Altogether, eight data are used besides the structure and to identify such an index unambiguously, at least a four-item label is needed, which presents that index as a function of exponents **a**, **b**, and **c**. In fact, four- to six-item labels have to be used to identify these indices unambiguously.

Survey of indices that might be useful to predict the values of particular tested physicochemical properties

One-digit values of exponents a, b, and c

For screening the most promising indices, we used one-digit values of exponents **a**, **b**, and **c**, i.e. whole numbers from -20 (which gives results close to those using $-\infty$) to 5, as well as + or -: 0.1 to 0.9. In correlations we used all available data. Lacking data, i.e. for all alkanes except octanes among R^2 and S, for propane and butanes at ST, for propane at BON and MON, for 2,3-dimethyl butane (23M4) at RON, as well as for 2,2,3,3-tetramethyl butane (2233M4) at ST, BON, MON, RON, logVP, and VP/Pc, were not considered in correlations.

Table 3 indicates that when the size *and* branching of alkanes are considered and the best combination of tested one-digit values of exponents **a**, **b**, and **c** is used, then the best results, $\text{abs}(r_{\text{max}}) > 0.999$, are observed among the van der Waals parameters a_0 and b_0 , represented here by Tc^2/Pc and Tc/Pc . They are followed by MR, $\Delta Hf^\circ g$, ω , BP, ΔH_v , VP/Pc, BP/Tc, Vc, Tc, Pc, and logVP, where $\text{abs}(r_{\text{max}}) > 0.99$ is observed.

Table 3. The highest values of correlation coefficients observed using data of physicochemical properties (PP) of alkanes from **propane** to **octanes** inclusive, as well as one-digit values of exponents **a**, **b**, and **c**, to obtain the values of the original indices derived from the **G(a,b,c)** matrix (left) or also of secondary indices (right)

PP	Index	a	b	c	r_{max}	r_{max}	Sec. index	a	b	c	
Tc^2/Pc	V_L	0	-0.6	-3	0.998	<	0.9997	$V^{4/7}$	0.5	0.5	0.4
Tc/Pc	V_L	0.9	-0.4	$-\infty$	0.999	<	0.999	$V^{N/\Sigma v}$	-0.5	0.1	1
MR	V	-0.5	0	-5	0.999	<	0.999	$V^{3/7}$	-0.5	0	0.5
$\Delta Hf^\circ g$	V	-0.6	-0.4	$-\infty$	0.996	<	0.998	$V^{2/7}$	3	0.1	0.9
ω	V_L	0.3	-1	$-\infty$	0.993	<	0.998	$P_L^{2/(N2-N)}$	1	2	1
BP	V	-0.5	-0.4	-5	0.995	<	0.998	$V^{N/\Sigma v}$	-0.6	-0.6	-4
ΔH_v	V	-0.7	-0.6	-5	0.996	<	0.997	$V^{N/\Sigma v}$	-0.5	-0.5	-5
VP/Pc	P_L	-0.6	-1	0.5	0.988	<	0.997	$P_L^{2/(N2-N)}$	-0.5	-5	-4
BP/Tc	V_L	0.5	-1	-5	0.993	<	0.995	$P_L^{2/(N2-N)}$	0	0.4	1
Vc	V_L	1	-0.4	-4	0.992	<	0.994	$PV^{1/7}$	-1	3	2
Pc	V_L	1	-0.8	-6	-0.986	<	-0.994	$V_L^{1/3}$	-1	-0.8	$-\infty$
Tc	V	-1	-0.5	-4	0.987	<	0.992	$V^{2/7}$	-1	-0.6	-4
logVP	V	-0.5	-0.5	-6	-0.990	=	-0.990	V	-0.5	-0.5	-6
Vm	V_L	1	-0.5	-6	0.987	<	0.988	$V_L^{1/2}$	1	-0.4	$-\infty$
B	V	-1	-0.4	-5	0.982	<	0.984	$V^{3/7}$	-1	-0.5	-5
ST	V	-2	-0.9	-5	0.956	<	-0.977	$V^{2/N}$	$-\infty$	5	3
C	V_L	0.4	-0.9	-4	-0.971	=	-0.971	V_L	0.4	-0.9	-4
αc	V_L	0.4	-0.9	-4	0.952	<	0.954	$PV^{1/7}$	2	0.1	2
BON	VC	-2	0.2	-4	-0.924	<	-0.950	$P_L^{2/(N2-N)}$	4	2	0.6
RON	VC	0.2	-2	-6	0.936	<	0.945	$P_L^{1/(N2-N)}$	11	4	0.9
d	V_L	2	0.9	-2	0.940	<	0.942	V_L/N	2	1	-0.9
MON	VC	-3	0.1	-6	-0.918	<	0.940	V_L/N	-3	-0.9	-5
n_D	V_L	3	2	-0.2	0.925	=	0.925	V_L	3	2	-0.2
V''	V_L	0.9	-1	$-\infty$	0.905	<	0.910	$V_L^{N/\Sigma v}$	1	-0.9	$-\infty$
dc	V	-4	-1	-5	0.849	<	0.852	$V^{N/\Sigma v}$	-4	-1	-5
CED	P_L	4	-5	0.5	0.778	<	0.839	$PV/(N^2-N)$	-2	-2	$-\infty$
Sol.par.	P_L	4	-5	0.5	0.766	<	0.833	$PV/(N^2-N)$	-2	-2	$-\infty$
A	P_L	-2	4	-3	-0.756	<	-0.789	$2V_L/(N^2-N)$	3	-0.5	-4
Zc	P_L	-0.3	-0.2	0.2	0.753	<	-0.769	$P_L^{2/(N2-N)}$	2	4	-0.7

< and = refer to non-rounded data

In exponent, $(N2-N)$ stands for (N^2-N)

In most cases the summation-derived original indices derived from the $G(a,b,c)$ matrix give the best result (22 out of 29 cases). The rest represent the multiplication-derived indices. When also tested secondary indices are considered, then in 18 cases the summation-derived indices give the best result, in 5 cases the multiplication-derived ones, and in 6 cases the mixed, summation-and-multiplication-derived indices give the best result. Among them prevail by far the whole-matrix and half-matrix indices.

The values of exponents **a** or **b**, giving rise to $\text{abs}(r)_{\text{max}} > 0.99$ are between -1 and +1. The values of exponent **c** are mostly between $-\infty$ and +1.

Table 4 illustrates the best observed correlation coefficients of original indices derived from the $G(a,b,c)$ matrix using one-digit values of exponents **a**, **b**, and **c** and the physicochemical properties Tc/Pc, BP, and the Solubility Parameter, which is formally equal to $(\Delta H_v/V_m)^{1/2}$. In former cases, the whole-matrix and half-matrix summation-derived indices (V_L and V), as well as multiplication-derived indices (P_L and M) give rise to best results, whereas the mixed type indices (PV and VP) and indices derived from other parts of the matrix (PC and VC) are not so good. In the case of Solubility Parameter only low values of correlation coefficients are observed. Anyway, these are provisional conclusions that have to be rechecked by fine-tuning of exponents **a**, **b**, and **c**. For this reason, the values of exponents **a**, **b**, and **c** are not given in Table 4.

The observed "best ten" types of secondary indices are:

$$\begin{aligned} \text{Tc/Pc:} & \quad V^{N/\Sigma v}, V^{3/7}, V_L^{1/3}, V^{1/3}, V^{2/7}, PV^{1/7}, V^{1/2}, V_L^{1/2}, V_L^{N/\Sigma v}, V/N \\ \text{BP:} & \quad V^{N/\Sigma v}, V^{3/7}, V^{1/2}, V_L^{1/2}, V^{1/3}, V_L^{1/3}, V^{4/7}, V^{2/7}, PV^{1/7}, V^{1/7} \end{aligned}$$

Table 4. The best correlation coefficients observed using one-digit values of exponents **a**, **b**, and **c**, as well as data for alkanes from **propane** to **octanes** inclusive and the original indices derived from the $G(a,b,c)$ matrix

Tc/Pc		BP		Sol.Par.	
0.999	V_L	0.995	V_L	0.766	P_L
0.994	V	0.995	V	-0.753	V_L
-0.991	P_L	-0.988	P_L	0.702	PC
-0.991	M	-0.983	PC	-0.679	VC
0.987	PV	-0.981	M	0.658	M
-0.977	PC	0.981	PV	-0.647	PV
0.972	VC	-0.957	VP	0.643	VP
-0.968	VP	0.952	VC	-0.641	V

Table 5. The highest values of correlation coefficients observed using data of physicochemical properties of octanes as well as one-digit values of exponents **a**, **b**, and **c** to obtain the values of the original indices derived from the $G(a,b,c)$ matrix (left) or also secondary indices (right)

P.P.	Index	a	b	c	r_{\max}		r_{\max}	Sec. index	a	b	c
Tc/Pc	PV	1	1	-0.8	-0.999	=	-0.999	V^2, V_L^2	1	1	-0.8
BP/Tc	PV	1	0.9	-2	-0.998	=	-0.998	PV	1	0.9	-2
ω	V_L	-0.1	-0.2	-1	-0.997	=	-0.997	V_L	-0.1	-0.2	-1
RON	V_L	-2	-0.6	-4	0.993	<	0.995	$V_L^{1/4}$	-2	-0.5	-4
MON	V_L	1	0	2	-0.993	<	-0.993	$VP^{1/2}$	0.1	0.1	0.4
Tc^2/Pc	P_L	0.1	0.2	0.09	0.987	<	0.992	$P_L^{1/2}$	0.2	0.3	0.1
BON	V_L	0.6	0.1	1	-0.988	<	0.989	$\ln V$	-4	-0.6	-4
C	PV	-0.6	2	0	0.984	<	0.984	PV^2	2	0	-0.1
Pc	V_L	1	0.4	-2	0.970	<	0.979	$M^{1/10}$	2	0.5	-3
n_D	P_L	0.5	0.03	-2	0.976	<	0.972	$P_L^{1/3}$	0.1	0.03	-3
d	P_L	0.5	0.2	-2	0.963	<	0.967	$P_L^{1/4}$	1	-1	-5
$\Delta H_f^{\circ g}$	V	-2	4	2	0.961	<	0.967	$V^{1/7}$	-2	4	2
S	V_L	-0.4	-1	-1	-0.966	<	-0.971	VC/NC	-0.4	4	-0.4
R^2	V	-0.5	0	-5	0.960	<	0.960	$\ln V$	-0.5	0	-5
ΔH_v	PV	-0.4	0.4	-8	-0.961	<	-0.961	$\ln PV$	-0.5	0.6	$-\infty$
V''	P_L	0.6	0.3	-2	-0.954	<	-0.957	$P_L^{1/3}$	0.9	-0.4	-4
V_m	P_L	0.6	0.3	-2	-0.955	<	-0.957	$P_L^{1/4}$	1	-0.7	-5
ST	V	5	3	2	-0.940	<	-0.940	$V^{6/7}$	5	3	2
BP	V_L	$-\infty$	4	2	-0.932	=	-0.932	V_L	$-\infty$	4	2
MR	V	3	1	2	0.894	<	0.921	$\ln V$	5	3	5
Tc	V_L	4	5	3	-0.881	=	-0.881	V_L	4	5	3
A	VP	4	2	-0.9	0.823	<	-0.852	$P_L^{1/4}$	3	5	-1
B	ΔP	0.5	-2	2	0.842	=	-0.842	ΔP	0.5	-2	2
V_c	VP	-0.2	-0.1	0.3	0.851	<	0.864	$\ln V_L$	5	2	4
d_c	VP	-0.2	-0.1	0.3	-0.837	<	-0.859	$\ln V_L$	5	2	4
a_c	P_L	0.1	0.3	-0.2	-0.849	<	-0.851	$VP^{1/8}$	3	1	-2
Z_c	VC	$-\infty$	4	0.9	0.831	<	0.844	VC^2	$-\infty$	4	1
$\log VP$	VN	$-\infty$	2	1	0.775	<	0.779	VC/NC	-4	-4	2
CED	P_L	0	-4	-0.5	0.731	<	0.736	$VPN^{1/2}$	-2	0.5	-1
Sol.par.	P_L	0	-4	-0.5	0.721	<	0.727	PN_L^2	-0.3	0.1	-0.3
VP/Pc	VN	$-\infty$	3	2	0.645	<	0.668	$\ln PV$	7	5	2

< and = refer to non-rounded data

Table 5 indicates that when only branching of alkanes is considered and the best combination of tested one-digit values of exponents **a**, **b**, and **c** is used, then there is not observed any result having $\text{abs}(r_{\max}) > 0.999$. The best result, $\text{abs}(r_{\max}) > 0.99$, is observed testing the van der Waals parameter b_0 , represented here by Tc/Pc. It is

followed by the result of BP/Tc, ω , RON, MON, and the van der Waals parameter a_0 , represented here by Tc^2/Pc . In other cases, $\text{abs}(r_{\text{max}}) < 0.99$.

In most cases the summation-derived original indices derived from the $\mathbf{G}(a,b,c)$ matrix give the best result (12 out of 31 cases). The rest represent the multiplication-derived indices (8 cases), the mixed type indices (9 cases) and the indices derived from other parts of the matrix (2 cases). When also tested secondary indices are considered, then in 12 cases the summation-derived indices give the best result, in 7 cases the multiplication-derived ones, in 7 cases the mixed, summation-(or subtraction)-and-multiplication-derived or multiplication-(or division)-and-summation-derived indices give the best result, whereas in 5 cases the indices derived from other parts of the matrix do.

The values of exponents \mathbf{a} or \mathbf{b} , giving rise to $\text{abs}(r)_{\text{max}} > 0.99$ are again in most cases between -1 and +1. The values of exponent \mathbf{c} are mostly between -5 and +2.

The best correlation coefficients observed using data of octanes and the primary indices derived from the $\mathbf{G}(a,b,c)$ matrix using one-digit values of exponents \mathbf{a} , \mathbf{b} , and \mathbf{c} with Tc/Pc are: PV (-0.998) > V_L (-0.998) > V (-0.998) > M (-0.994) > P_L (-0.992) > VP (-0.992) > VC (0.966) > ΔP (0.928) > QV (-0.918) > ΔV (0.912) > VN_L (-0.895) > VN (-0.892) > PN_L (0.889) > ΔPN (0.873) > VPN (-0.800) > PN (0.774) > QP (-0.752) > QVN (-0.750) > PC (0.743) > PVN (-0.709). The observed "best ten" types of secondary indices are $PV^{4/7}$, PV, V_L^2 , V^2 , $PV^{1/2}$, V_L , V, $V^{6/7}$, $V^{5/7}$, and $PV^{1/3}$.

The best observed indices giving rise to $\text{abs}(r)_{\text{max}} > 0.99$, derived using two-digit values of exponents \mathbf{a} , \mathbf{b} , and \mathbf{c}

Some indices which give rise to the highest $\text{abs}(r)_{\text{max}} > 0.99$ using two-digit values of exponents \mathbf{a} , \mathbf{b} , and \mathbf{c} , are presented in Table 6 for alkanes from propane to octanes inclusive as well as for octanes only.

We can see that, as a rule, the secondary indices derived from indices originating from the $\mathbf{G}(a,b,c)$ matrix give rise to the best correlation coefficients. Some of them are only slightly better than those obtained using one-digit values of exponents \mathbf{a} , \mathbf{b} , and \mathbf{c} . In some cases, several local maxima of similar value are observed.

The first digit in the value of exponents \mathbf{a} , \mathbf{b} , and \mathbf{c} , used to derive an index from the $\mathbf{G}(a,b,c)$ matrix, defines in most tested cases the first three decimals of the

correlation coefficient. The second digit in the exponents **a**, **b**, and **c** improves in most tested cases the value of the third to fifth decimal, depending on how far from the best value of the exponent is its one-digit approximation. The third digit in the exponents **a**, **b**, and **c** improves the value of the fifth or higher decimal of the correlation coefficient. For our purpose, five decimals in the value of the correlation coefficient are considered sufficient, therefore only two digits in the value of exponents **a**, **b**, and **c** are needed.

Table 6. The highest values of correlation coefficients observed using data of physicochemical properties BP/Tc and Tc/Pc of alkanes from **propane** to **octanes** inclusive resp. of **octanes** only, as well as one-digit and two-digit values of exponents **a**, **b**, and **c** to obtain the values of the best indices derived from the **G(a,b,c)** matrix

C ₃ -C ₈					C ₈								
No. of digits	Index	a	b	c	r	No. of digits	Index	a	b	c	r		
BP/Tc	1	P _L ^{2/(N2-N)}	0	0.4	1	0.99522	BP/Tc	1	PV	1	0.9	-2	-0.99755
	2	P _L ^{2/(N2-N)}	0.010	0.37	0.92	0.99953	2	VP ^{1/(N2-N)}	2.1	2.1	-12.6	-0.99879	
							2	P _L ^{2/(N2-N)}	0.97	1.03	-6.1	-0.99849	
Tc/Pc	1	V _L ^{N/Σv}	-0.2	0	1	0.99933	2	M	-0.164	0.20	-0.111	-0.99848	
	2	V _L ^{N/Σv}	-0.27	-0.0091	1.06	0.99936	2	M ^{1/(N2-N)}	1.5	0.0	-5.6	-0.99834	
	2	V _L ^{N/Σv}	-0.33	-0.20	1.14	0.99936	2	P _L ^{1/3}	0.056	0.030	-0.51	-0.99805	
	1	V ^{N/Σv}	-0.5	0.1	1	0.99915	2	PV	1.26	0.82	-1.63	-0.99764	
	2	V ^{N/Σv}	-0.32	-0.32	1.15	0.99931	Tc/Pc	1	V ² , V _L ²	1	1	-0.8	-0.99851
	2	V ^{N/Σv}	-0.51	-0.023	1.08	0.99930	2	V _L ²	1.01	0.99	-0.88	-0.99859	
	2	V ^{N/Σv}	-0.29	-0.30	1.10	0.99930	2	V ²	1.04	1.00	-0.80	-0.99855	
	2	V ^{N/Σv}	1.34	-0.26	0.88	0.99920	2	PV ^{4/7}	1.03	1.03	-0.86	-0.99807	
						2	V	1.03	1.04	-0.87	-0.99793		
						2	V ^{4/7}	1.05	1.05	-0.88	-0.99737		
						2	lnV _L	1.04	0.76	-2.1	-0.99736		

Discussion

The **G(a,b,c)** matrix

The **G(a,b,c)** matrix is in general non-symmetric, i.e. it is composed of two non-equal halves (cf. Fig. 1). It is symmetric only in special cases, i.e. when **a** = **b**, as well as when *n*-alkanes are considered.

There are also other features in the **G(a,b,c)** matrix. On the one hand, if we use a positive value of exponent **c**, we have a usual form of the vertex-degree vertex-distance

matrix, which includes also the simple distance matrix. If, on the other hand, we use negative values of exponent c , we have to do with a reciprocal type of the vertex-degree vertex-distance matrix, which includes besides several other reciprocal types of matrix also the adjacency matrix ($a = b = 0, c = -\infty$).⁶

The primary indices derived from the $G(a,b,c)$ matrix

The asymmetry of the $G(a,b,c)$ matrix enables deriving a great number of index groups. They can be derived by summation of its non-diagonal elements: $V(a,b,c)$, $V_D(a,b,c)$, $V_L(a,b,c)$, $VC(a,b,c)$, $VN(a,b,c)$, $VN_D(a,b,c)$, $VN_L(a,b,c)$; by multiplication of them: $M(a,b,c)$, $P_D(a,b,c)$, $P_L(a,b,c)$, $PC(a,b,c)$, $PN(a,b,c)$, $PN_D(a,b,c)$, $PN_L(a,b,c)$; by multiplication of summation derived indices: $PV(a,b,c)$, $PVN(a,b,c)$; by summation of multiplication derived indices: $VP(a,b,c)$, $VPN(a,b,c)$, etc.

They can be the highest eigenvalues of the matrix, $L(a,b,c)$ (not considered here), or they can be derived by other ways one can imagine, e.g. the difference between the values of the indices derived from two halves of the matrix: $\Delta V(a,b,c)$, $\Delta P(a,b,c)$, $\Delta PN(a,b,c)$, the quotients of them $QV(a,b,c)$, $QVN(a,b,c)$, $QP(a,b,c)$, etc.

The indices $V_L(a,b,c)$, $V_D(a,b,c)$, $V(a,b,c)$, $PV(a,b,c)$, $VC(a,b,c)$, $P_L(a,b,c)$, $P_D(a,b,c)$, $M(a,b,c)$, $VP(a,b,c)$, and $PC(a,b,c)$ represent the size as well as branching of the molecule. They thus belong to the BI_M -type³⁴ of indices. They were used to test data of all alkanes from propane to octanes inclusive, as well as of octanes only.

The indices $\Delta V(a,b,c)$, $QV(a,b,c)$, $VN_L(a,b,c)$, $VN(a,b,c)$, $PVN(a,b,c)$, $QVN(a,b,c)$, $QP(a,b,c)$, $\Delta P(a,b,c)$, $PN_L(a,b,c)$, $PN(a,b,c)$, $VPN(a,b,c)$, and $\Delta PN(a,b,c)$ do not reflect the size but only branching of the molecule. Therefore, they belong to the BI_A -type³⁴ of indices. For this reason they were used to index the physicochemical properties of octanes only. These facts are reflected in Tables 3, 5, 7, and 8.

It has been reasonably expected that the characteristics of these indices would vary widely, which proved to be true. There is also an additional source of variation of characteristics of these indices. It is the exponents a , b , and c . Here we have an infinite number of possibilities to choose the value of each of these exponents, as well as of the combinations of them. Using these indices, we have at our disposal at least 20 groups of them, each group containing an infinite number of possible indices.

The secondary indices derived from the G(a,b,c) matrix

From these groups of, let us say, primary indices derived from the $G(a,b,c)$ matrix, we can derive an infinite number of groups of, let us say, secondary indices. We can derive them by division of values of primary indices using a variable factor like N (the number of vertices in the graph, which is equal to the number of rows or columns in the matrix), or N^2-N (the number of non-diagonal elements of the matrix), or some other variable that one finds plausible, e.g. NC (the number of elements common to both halves of the matrix), NN (the number of elements not common to both halves of the matrix), etc., etc. We can derive them by raising the values of primary indices to an appropriate exponent, which can be any real number or a variable like N , N^2-N , $N/\sum v$ (v is the degree of vertex in the graph), etc., etc.

We can also draw the logarithms of the values of primary indices. Here we have an interesting situation. Namely, the logarithms of the multiplication-derived indices are in fact summation-derived indices since $\log_x(\prod(v_i^a \times v_j^b \times d_{ij}^c)) = \sum(a \times \log_x v_i + b \times \log_x v_j + c \times \log_x d_{ij})$. Because $\log_x I = \log_y I / \log_y x$, the base of the logarithm does not influence the correlation coefficient in this case. Due to this fact, the natural logarithm (ln) was chosen here.

Usefulness of indices derived from the G(a,b,c) matrix

Then another question arises: Are all of these indices useful? And if not, then which of them are useful? To obtain some indication to the answer, correlation of the values of a number of these indices with the values of 29 physicochemical properties of alkanes from propane to octanes inclusive, as well as of 31 physicochemical properties of octanes, was performed. A wide spread of values of correlation coefficients was observed. Since only the best correlations are interesting, the indices giving rise to them are presented in Table 3 for alkanes from propane to octanes inclusive, and in Table 5 for octanes only. As a rule, the maxima of correlation coefficients coincide with minima of standard error data and for this reason the data of standard error are not given in these tables.

From Table 3 it can be seen that when the influence of the size and branching of alkanes is to be considered, then a number of physicochemical properties can be well²

correlated with some of the tested indices. Less well expressed is the correlation when only the influence of branching of alkanes is considered; see Table 5.

From the data, which were the basis for construction of Table 3 and 5 we can deduce a lot of information:

1. Separate data for *n*-alkanes are not given in these tables, but as a rule of thumb the best correlations between the tested physicochemical properties and indices of alkanes are observed when considering *n*-alkanes; lower correlations are observed when all alkanes are considered and the lowest ones when only octanes are considered. Thus, this rule of thumb is: $\text{abs}(r)_{\text{max}}: n\text{-alkanes} \gg \text{all alkanes} > \text{octanes}$.

2. Which group of indices is the most promising to index well a physicochemical property?

According to the criterion $\text{abs}(r)_{\text{max}} > 0.99$ for the correlation with any one of the tested physicochemical properties,² these are the following indices:

a. *Testing data of all alkanes, from propane to octanes inclusive:*

Original indices and/or some tested secondary indices derived from them: V(a,b,c), V_L (a,b,c), PV(a,b,c), P_L (a,b,c), M(a,b,c), and VP (a,b,c);

b. *Testing data of octanes:*

Original indices and/or some tested secondary indices derived from them: V(a,b,c), V_L (a,b,c), PV(a,b,c), M(a,b,c), P_L (a,b,c), and VP(a,b,c);

c. Other tested indices [VC(a,b,c), PC(a,b,c), ΔV (a,b,c), QV(a,b,c), VN_L (a,b,c), VN(a,b,c), PVN(a,b,c), QVN(a,b,c), QP(a,b,c), ΔP (a,b,c), PN_L (a,b,c), PN(a,b,c), VPN(a,b,c), ΔPN (a,b,c) and the tested secondary indices derived from them] do *not* give rise to $\text{abs}(r)_{\text{max}} > 0.99$.

3. From the $G(a,b,c)$ matrix can be derived good (original and/or secondary) indices for the following physicochemical properties, since $\text{abs}(r)_{\text{max}} > 0.99$:

a. *Testing data of all alkanes, from propane to octanes inclusive:* Tc/Pc as well as:

V_L (a,b,c): MR, Tc^2/Pc , BP, ΔH_f° , ΔH_v , Vc, Tc, BP/Tc, logVP, Pc

V(a,b,c): MR, Tc^2/Pc , BP, ΔH_f° , ΔH_v , Vc, Tc, BP/Tc, logVP, ω ,

PV(a,b,c): MR, Tc^2/Pc , BP, ΔH_f° , ΔH_v , Vc, Tc, logVP

P_L (a,b,c): MR, Tc^2/Pc , BP, ΔH_f° , ΔH_v , Vc, BP/Tc, ω , VP/Pc

M(a,b,c): MR, Tc^2/Pc , BP, ΔH_v

VP(a,b,c): Pc

b. *Testing data of octanes*: Tc/Pc, BP/Tc, ω , as well as:

V_L(a,b,c): RON, MON
 V(a,b,c): RON
 PV(a,b,c): MON
 P_L(a,b,c): MON, Tc²/Pc
 M(a,b,c): MON
 VP(a,b,c): MON, Tc²/Pc

4. The indices derived from the **G**(a,b,c) matrix which *might* be interesting as indices for the physicochemical properties, since $0.9 < \text{abs}(r)_{\text{max}} < 0.99$:

- Testing data of alkanes from propane to octanes inclusive: All groups of tested indices.
- Testing data of octanes: All groups of them, even PVN(a,b,c) and PC(a,b,c), which are the worst cases.

5. The original and the tested secondary indices derived from the **G**(a,b,c) matrix are *not* good indices for the following physicochemical properties, since $\text{abs}(r)_{\text{max}} < 0.9$:

a. *Testing data of all alkanes, from propane to octanes inclusive*: dc, Zc, A, Cohesive energy density, Solubility parameter, as well as:

V(a,b,c): MON, V"
 PV(a,b,c): BON, V"
 VC(a,b,c): VP/Pc, V"
 P_L(a,b,c): n_D, d
 M(a,b,c): BON, MON, RON, V", n_D, d
 VP(a,b,c): V", n_D, d
 PC(a,b,c): ST, BON, MON, RON, V", n_D, d

b. *Testing data of octanes*: logVP, VP/Pc, Tc, Vc, dc, α c, Zc, A, B, CED, Sol.par., as well as

PV(a,b,c): BP
 VC(a,b,c): ΔH_f° , MR, BP, ST
 ΔV (a,b,c): MR
 QV(a,b,c): MR, n_D, d, Vm, V", ΔH_f° , BP, ST
 VN_L(a,b,c): MR, Pc, Tc/Pc
 VN(a,b,c): MR, ST, Vm
 PVN(a,b,c): All of them except BON, MON, RON, Tc²/Pc
 QVN(a,b,c): All of them except n_D, d, Vm, V"

M(a,b,c): BP, MR
 VP(a,b,c): BP, MR, ST
 PC(a,b,c): All of them except BON, MON, RON, Tc²/Pc
 QP(a,b,c): All of them except RON, n_D, d, Vm, V"
 ΔP (a,b,c): R², MR, ST
 PN_L(a,b,c): BP, MR, ST, Tc/Pc, ΔH_v , ΔH_f° , Pc, C
 PN(a,b,c): All of them except n_D, d, Vm, V"
 VPN(a,b,c): All of them except BON, RON, S, BP/Tc, ω , n_D, d, Vm, V"
 APN(a,b,c): All of them except BON, MON, RON, S, BP/Tc, Tc²/Pc, ω , C, n_D, d, Vm, V"

Evidently, the indices derived from the $G(a,b,c)$ matrix are not good indices of some physicochemical properties of alkanes, namely of dc , Zc , A , *Cohesive energy density*, and *Solubility parameter*. Well predicted are, referring to alkanes of different size and branching, Tc^2/Pc , Tc/Pc , MR , ΔH_f° , ω , BP , ΔH_v , VP/Pc , BP/Tc , Vc , Tc , Pc , and $\log VP$, whereas referring to alkanes of the same size are well predicted Tc/Pc , BP/Tc , ω , RON , MON , and Tc^2/Pc . Other physicochemical properties are predicted worse than one would desire.

6. The whole-matrix as well as the half-matrix indices are in general good indices, whereas the indices derived from elements common to both halves of the matrix as well as the indices derived from elements not common to both halves of the matrix are in general not. They are to be tested for their usefulness as the second or the third index in multiparameter relationships.

Using tested indices derived from the $G(a,b,c)$ matrix and optimizing the values of exponents a , b , and c one can arrive, considering particular physicochemical properties of alkanes, to quite good correlations. This can be achieved by proper selection of the type of the index, of values of exponents a , b , and c , as well as by proper selection of modification of the index (secondary indices). Using these indices, the maximum of correlation coefficient in the space of exponents a , b , and c is in most cases flat and in several cases there are observed several local maxima of similar value. Looking for the best QSPR/QSAR model and using one or more of these indices, one should test the values of indices at these local maxima as well as in vicinity of them.

We compared the best observed indices derived from the $G(a,b,c)$ matrix using two-digit values of exponents a , b , and c with a selection of known indices: χ ,¹⁰ D (the largest eigenvalue of the distance matrix),³⁴ EA ,³⁵ ID ,³⁶ J ,³⁷ λ_1 ,^{19,20} $\lambda\lambda_1$,³⁸ MTI ³⁹ (the values were taken from ref.⁴⁰), $Sch-S$,⁴¹ $Sch-TF$,⁴¹ Xu ,⁴² W ,⁷ and Z .⁴³ This comparison is presented in Table 7 for alkanes from propane to octanes inclusive and in Table 8 for octanes only.

When alkanes from propane to octanes inclusive are taken into consideration, Table 7, Tc^2/Pc , Tc/Pc , MR , BP , ΔH_v , and Vc are well predicted ($r > 0.99$) by $\lambda\lambda_1$, Xu , ID , Xu , χ , and $\lambda\lambda_1$, respectively, but the best indices derived from the $G(a,b,c)$ matrix are better in this respect. There is no case that an index from the above mentioned selection would be better than the best observed index derived from the $G(a,b,c)$ matrix.

When only octanes are considered, Table 8, no index of the above mentioned selection gives rise to $\text{abs}(r)_{\text{max}} > 0.99$, whereas a number of indices derived from the $\mathbf{G}(a,b,c)$ matrix do. In most cases, the indices derived from the $\mathbf{G}(a,b,c)$ matrix are appreciably better than those from the above mentioned selection.

Table 7. The highest values of correlation coefficients and standard errors observed using data of physicochemical properties (PP) of alkanes from **propane** to **octanes** inclusive, as well as with the best observed indices derived from the $\mathbf{G}(a,b,c)$ matrix using two-digit values of exponents **a**, **b**, and **c** (left) and with a selection of known indices (right)

PP	Index(a,b,c)	$s \times 10^x$	r_{max}	Index	$s \times 10^x$	r_{max}
Tc ² /Pc	$V^{4/7}(0.49, 0.50, 0.41)$	6.67	0.9997	$\lambda\lambda_1$	20.80	0.997
Tc/Pc	$V^{N/\Sigma v}(-0.56, 0.039, 1.08)$	1.34	0.999	Xu	3.10	0.996
MR	$V^{3/7}(-0.52, 0.00, 0.50)$	3.09	0.999	ID	4.61	0.997
ΔH_f°	$V^{2/7}(3.0, 0.090, 0.91)$	4.68	0.998	Sch-S	14.86	0.979
ω	$P_L^{2/(N^2-N)}(1.12, 2.1, 0.99)$	3.84	0.998	D	18.38	0.949
BP	$V^{N/\Sigma v}(-0.62, -0.60, -3.9)$	2.76	0.998	Xu	5.36	0.992
ΔH_v	$V^{N/\Sigma v}(-0.47, -0.50, -5.3)$	7.07	0.997	χ	10.92	0.993
VP/Pc	$P_L^{2/(N^2-N)}(-0.83, -5.5, -3.1)$	3.00	0.997	Xu	26.00	-0.762
BP/Tc	$P_L^{2/(N^2-N)}(0.029, 0.40, 0.89)$	1.70	0.995	D	4.16	0.972
Vc	$PV^{1/7}(0.087, 3.0, 1.64)$	7.45	0.995	$\lambda\lambda_1$	10.41	0.990
Pc	$V_L^{1/3}(0.98, -0.86, -\infty)$	4.32	-0.994	Xu	9.75	-0.968
Tc	$V^{2/7}(-1.37, -0.75, -3.8)$	5.77	0.994	ID	9.44	0.983
logVP	$V(-0.50, -0.50, -5.7)$	9.46	-0.990	Xu	10.89	-0.987
Vm	$V_L^{1/2}(1.05, -0.46, -\infty)$	3.33	0.988	ID	4.31	0.980
ST	$V^{2/N}(-\infty, 4.8, 3.0)$	3.10	-0.986	χ	7.16	0.921
B	$V^{3/7}(-1.42, -0.52, -5.1)$	2.42	0.985	Xu	3.13	0.975
C	$V_L(0.56, -0.67, -\infty)$	2.14	-0.972	D	2.54	-0.960
αc	$V_L(0.094, -1.48, -3.1)$	8.49	0.956	D	9.47	0.945
BON	$P_L^{2/(N^2-N)}(4.3, 2.1, 0.61)$	1.04	-0.951	D	2.66	-0.603
RON	$VC^2(-2.4, -0.74, -5.0)$	1.05	0.948	D	2.82	-0.513
d	$V_L(2.3, 1.36, -0.54)/N$	1.53	0.947	Sch-S	2.05	0.904
MON	$V_L(-2.2, -0.79, -4.7)/N$	0.97	0.945	EA	2.43	0.577
n _D	$V_L(2.8, 1.61, -0.53)$	8.58	0.929	Sch-S	11.37	0.871
V''	$V_L^{N/\Sigma v}(0.99, -0.87, -\infty)$	3.34	0.917	Sch-TF	3.70	0.897

In exponent, (N^2-N) stands for (N^2-N)

^x: Integer to adjust the value of s to the same order of magnitude

If we compare the correlation coefficients observed here with those reported by some other authors using more than one variable in their equations, we can see that in

several cases, by using one index derived from the $G(a,b,c)$ matrix, we can obtain results which are comparable to or not much worse than those obtained by using models with several (2 to 5) other indices at once, Table 9.

Table 8. The highest values of correlation coefficients observed using data of physicochemical properties of octanes, as well as the best observed indices derived from the $G(a,b,c)$ matrix using two-digit values of exponents **a**, **b**, and **c** (left), and with a selection of known indices (right)

P.P.	Index(a,b,c)	r_{\max}	Index	r_{\max}
Tc/Pc	$V_L^2(1.03, 1.01, -0.84)$	-0.999	Xu	0.917
BP/Tc	$PV(1.07, 1.00, -0.77)$	-0.999	$\lambda\lambda_1$	0.927
ω	$V_L(-1.08, -0.22, -1.20)$	-0.997	$\lambda\lambda_1$	0.956
RON	$V_L^{1/4}(-2.0, -0.60, -4.0)$	0.995	ID	-0.964
BON	$\ln V(-\infty, -1.37, -3.6)$	0.994	Xu	-0.961
MON	$VP^{1/2}(0.128, 0.128, 0.33)$	-0.993	D	-0.941
Tc ² /Pc	$P_L^{1/2}(0.20, 0.30, 0.111)$	0.992	D	0.962
C	$PV^2(1.75, -0.35, -0.071)$	0.986	ID	-0.920
Pc	$M^{1/10}(0.20, 0.45, -3.0)$	0.979	J	0.793
n_D	$P_L^{1/3}(1.01, -0.58, -4.9)$	0.977	Sch-TF	-0.794
S	$VC(-0.27, 4.6, -0.29)/NC$	-0.975	$\lambda\lambda_1$	0.942
$\Delta H_f^\circ g$	$V^{1/7}(-2.1, 4.1, 2.0)$	0.968	ID	-0.709
d	$P_L^{1/4}(1.00, -1.08, -5.0)$	0.967	Sch-TF	-0.745
R^2	$\ln V(0.47, 0.00, -5.0)$	0.963	ID	0.902
ΔH_v	$\ln PV(-0.40, 0.48, -\infty)$	-0.962	ID	0.886
V''	$P_L^{1/3}(0.90, -0.45, -3.9)$	-0.958	Sch-TF	0.717
V_m	$P_L^{1/4}(1.00, -0.93, -4.7)$	-0.958	D	-0.941
ST	$V^{6/7}(5.1, 3.1, 2.2)$	-0.944	Z	0.453
BP	$V_L(-\infty, 3.9, 2.1)$	-0.937	Z	0.742
MR	$\ln V(4.9, 2.9, 5.1)$	0.921	J	-0.629

Table 9. Comparison of correlation coefficients and standard errors observed by some other authors and our data.

No. of variables	Li ³³		Ivanciuc ⁴⁴		Ivanciuc ⁴⁵		Table 7, here			
	5	5	3	2	1	1	1			
Indices	χ^{**}		W ^{**}		various		N, IB		I(a,b,c)	
	r	s	r	s	r	s	r	s	r	s
BP	0.997	3.31	0.998	3.07	0.994	2.79	0.984	4.94	0.998	2.76
Tc	0.996	4.76	0.998	4.54					0.992	5.77
Pc	0.992	0.50	0.990	0.55					-0.994	0.43
ST	0.998	0.22	0.989	0.29					-0.977	0.31
ΔH_v					0.990	0.63	0.975	0.81	0.997	0.30
n_D					0.984	0.0025	0.960	0.0038	0.929	0.0086
d					0.990	3.73	0.961	7.30	0.947	15.3

** : Molecular connectivity indices, resp. Extended Wiener indices

Thus, in several instances a properly selected index of the type described here gives as good a correlation as if several classic indices are used at once. Naturally, there are also exceptions.

On the other hand, when using the Randić¹⁰ index χ one can expect several good results since $\chi = V_L(-1/2, -1/2, -\infty) = 1/2 V(-1/2, -1/2, -\infty)$, which is close to many good indices derived from the $G(a,b,c)$ matrix. The exponents which are used to derive the Wiener⁷ index, $W = V_L(0,0,1) = 1/2 V(0,0,1)$ are not so close to the exponents of the best observed indices as there are exponents to be used to derive the Randić index χ . This is the reason why the Randić index χ is often a better index than the Wiener index.

Extending the idea of the $G(a,b,c)$ matrix and the diversity of indices derived from it to other sources of indices would open additional possibilities to find good indices for particular needs.

Conclusion

The indices derived from the $G(a,b,c)$ matrix in the ways described here are promising indices of several physicochemical properties of alkanes. Therefore, their use in QSPR/QSAR models is recommended.

References

1. L. Pogliani, *Indian J. Chem.* **2003**, 42A, 1347–1353.
2. Z. Mihalić, N. Trinajstić, *J. Chem. Educ.* **1992**, 69, 701–712.
3. M. Randić, C. Basak, M. Pompe, M. Novič, *Acta Chim. Slov.* **2001**, 48, 169–180.
4. R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley-CH, New York 2000.
5. E. Estrada, *Chem. Phys. Lett.* **2001**, 336, 248–252.
6. A. Perdih, B. Perdih, *Acta Chim. Slov.* **2002**, 49, 67–110.
7. H. Wiener, *J. Am. Chem. Soc.* **1947**, 69, 17–20.
8. O. Ivanciuc, *Rev. Roum. Chim.* **1989**, 34, 1361–1368.
9. A. Perdih, B. Perdih, *Acta Chim. Slov.* **2002**, 49, 291–308.
10. M. Randić, *J. Am. Chem. Soc.* **1975**, 97, 6609–6615.
11. A. Perdih, B. Perdih, *Acta Chim. Slov.* **2003**, 50, 95–114.
12. O. Ivanciuc, *Rev. Roum. Chim.* **1999**, 44, 519–528.
13. O. Ivanciuc, *Rev. Roum. Chim.* **2000**, 45, 587–596.
14. A. T. Balaban, D. Mills, O. Ivanciuc, S. C. Basak, *Croat. Chem. Acta.* **2000**, 73, 923–941.
15. O. Ivanciuc, T. Ivanciuc, A.T. Balaban *ACH-Models Chem.* **2000**, 137, 57–82.
16. O. Ivanciuc, *J. Chem. Inf. Comput. Sci.* **2000**, 40, 1412–1422.
17. A. Perdih, B. Perdih, *Acta Chim. Slov.* **2002**, 49, 497–514.
18. A. Perdih, B. Perdih, *Acta Chim. Slov.* **2003**, 50, 83–94.
19. L. Lovasz, J. Pelikan, *Period. Math. Hung.* **1973**, 3, 175–182.
20. D. M. Cvetković, I. Gutman, *Croat. Chem. Acta* **1977**, 49, 115–121.
21. A. Perdih, B. Perdih, *Acta Chim. Slov.* **2002**, 49, 309–330.

22. A. Perdih, B. Perdih, *Acta Chim. Slov.* **2002**, *49*, 467–482.
23. A. Perdih, B. Perdih, *Acta Chim. Slov.* **2003**, *50*, 161–184.
24. I. Gutman, W. Linert, I. Lukovits, Ž. Tomović, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 113–116.
25. A. Perdih, B. Perdih, *Acta Chim. Slov.* **2003**, *50*, 513–538.
26. D. R. Lide, *CRC Handbook of Chemistry and Physics*, 76th Ed., CRC Press, Boca Raton 1995–96.
27. J. A. Dean, *Lange's Handbook of Chemistry*, McGraw-Hill, New York 1985.
28. L. Pogliani, *J. Phys. Chem.* **1995**, *99*, 925–937.
29. A. T. Balaban, I. Motoc, *MATCH (Commun. Math. Chem.)* **1979**, *5*, 197–218.
30. C. Morley, *Combust. Sci., Technol.* **1987**, *55*, 115–123.
31. E. S. Goll, P.C. Jurs, *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1081–1089.
32. B. Ren, *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 139–143.
33. X. H. Li, *Chem. Phys. Lett.* **2002**, *365*, 135–139.
34. A. Perdih, M. Perdih, *Acta Chim. Slov.* **2000**, *47*, 231–259.
35. Y.-Q. Yang, L. Xu, C.-Y. Hu, *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1140–1145.
36. M. Randić, *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 164–175.
37. A.T. Balaban, *Chem. Phys. Lett.* **1982**, *89*, 399–404.
38. M. Randić, *Acta Chim. Slov.* **1997**, *44*, 57–77.
39. H.P. Schultz, *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 227–228.
40. Z. Mihalić, S. Nikolić, N. Trinajstić, *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 28–37.
41. H.P. Schultz, T.P. Schultz, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 107–112.
42. B. Ren, *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 139–143.
43. H. Hosoya, *Bull. Chem. Soc. Japan* **1971**, *44*, 2332–2339.
44. O. Ivanciuc, T. Ivanciuc, A.T. Balaban, *Internet Electron. J. Mol. Des.* **2002**, *1*, 467–487.
45. O. Ivanciuc, T. Ivanciuc, D. Cabrol-Bass, A.T. Balaban, *Internet Electron. J. Mol. Des.* **2002**, *1*, 319–331.

Povzetek

Iz matrike tipa $G(a,b,c)$ lahko izvedemo vsaj 20 skupin topoloških indeksov. Vsaka od teh skupin ima neskončno mnogo teh indeksov. Številni od njih korelirajo z nekaterimi od 29 odnosno 31 upoštevanih fizikokemijskih lastnosti alkanov z $\text{abs}(r)_{\text{max}} > 0.99$, in še več od njih z $\text{abs}(r)_{\text{max}} > 0.9$. Zato priporočamo njihovo uporabo v modelih za QSPR/QSAR. Neprimerni pa so za opis fizikokemijskih lastnosti d_c , Z_c , A , gostote kohezivne energije in topnostnega parametra.