J Math Chem (2008) 44:559–568
DOI 10.1007/s10910-007-9329-4

ORIGINAL PAPER

Kekulé structures of hexagonal chains—some unusual connections

Haruo Hosoya · Ivan Gutman

Received: 2 August 2007 / Accepted: 10 October 2007 / Published online: 8 December 2007

Abstract It is known since 1977 that the number K of Kekulé structures of a hexagonal chain is equal to the topological Z-index of a pertinently constructed "caterpillar" tree. Based on this relation we now connect K with some of other, seemingly unrelated, concepts: continuants (from number theory) and matchings of the path–graph (further related to Fibonacci numbers). We also arrive at a tridiagonal determinantal expression for K.

Keywords Hosoya index · Topological index · Z-index · Hexagonal chain · Kekulé structure

1 Introduction

In 1971 one of the present authors [1] introduced a molecular-graph-based structure descriptor that he named "*topological index*" and denoted by Z. In this paper we call this quantity the Z-index.[1] A legion of chemical applications and mathematical properties of the Z-index has been discovered; for details see the recent surveys [2,3] and some of the newest papers published in this area [4–9].

The Z-index is defined as follows. Let G be a molecular graph and let $p(G, j)$, $j = 2, 3, \ldots$, be the number of selections of j disjoint (i.e., mutually non-touching)

[1] In the current literature Z is usually referred to as the *Hosoya index*.

H. Hosoya
Faculty of Science, Ochanomizu University, 112-8610, Tokyo, Japan
e-mail: hosoya@is.ocha.ac.jp

I. Gutman (✉)
Faculty of Science, University of Kragujevac, P.O. Box 60, 34000 Kragujevac, Serbia
e-mail: gutman@kg.ac.yu

edges in G. In addition, $p(G, 1)$ is equal to the number of edges of G, and $p(G, 0) = 1$. Then

$$Z = Z(G) = \sum_{j \geq 0} p(G, j). \tag{1}$$

Two immediate properties of the Z-index [1] will be needed in the following considerations.

Isolated vertices have no effect on the value of Z. Thus, if G' is obtained from G by adding to it any number of isolated vertices, then

$$Z(G') = Z(G). \tag{2}$$

If the graph G consists of disconnected components $G_1, G_2, \ldots, G_t$, then

$$Z(G) = Z(G_1) \times Z(G_2) \times \cdots \times Z(G_t). \tag{3}$$

The enumeration of the Kekulé structures in benzenoid molecules is a traditional and extensively elaborated field of mathematical chemistry [10,11]. Methods for determining the Kekulé structure count K of hexagonal chains were discovered already in the pioneering days [12], and were re-iterated from time to time [13–17].

In 1977 one of the present authors [18] discovered a curious relation between the sextet polynomial of a hexagonal chain and the matching polynomial of a caterpillar tree (for details see below). As a special case of this result, the Kekulé structure count of a hexagonal chain was shown to be equal to the Z-index of the corresponding caterpillar.

A *hexagonal chain* or *unbranched catacondensed benzenoid system* is a benzenoid system in which no hexagon has more than two neighbors. An example of a hexagonal chain is given in Fig. 1.

Denote the number of hexagons of a hexagonal chain by h. The hexagons of a hexagonal chain may be annelated in only three ways [11]: Each chain possesses exactly two terminal hexagons (L_1) whereas all other hexagons are annelated either linearly (L_2) or angularly (A), see Fig. 2.

To each hexagonal system a string of h symbols L and A can be associated, indicating the mode of annelation of the consecutive hexagons, starting from a terminal hexagon; the modes L_1 and L_2 are not distinguished. This string is referred to as the LA-sequence [11,18]. For instance, the LA-sequence of the chain depicted in Fig. 1 is

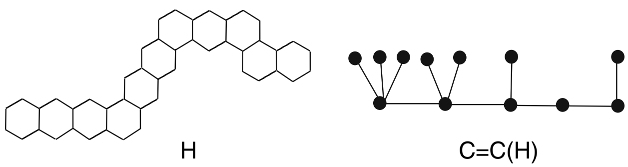

Fig. 1 A hexagonal chain H with 11 hexagons, and the corresponding caterpillar tree $C = C(H)$ with 11 edges

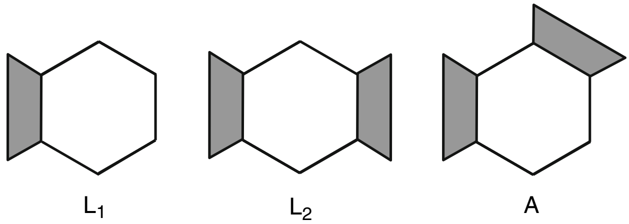

Fig. 2 Annelation modes of hexagons that occur in hexagonal chains

$LLLALLALAAL$. Abbreviating LL by L^2, LLL by L^3, etc. the latter LA-sequence is written as $L^3AL^2ALAL^0AL$.

The general form of the LA-sequence of a hexagonal chain in which there are $n-1$ angularly annelated hexagons is

$$L^{k_1}AL^{k_2}AL^{k_3}A\cdots L^{k_{n-1}}AL^{k_n}$$

where k_i is the number of L-mode hexagons lying between the $(i-1)$th and ith angularly annelated hexagon, $i = 2, \ldots, n-1$, whereas k_1 and k_n are, respectively, the number of the L-mode hexagons before the first and after the last A-mode hexagon. Therefore,

$$k_1, k_n \geq 1, \quad k_2, \ldots, k_{n-1} \geq 0$$

and

$$k_1 + k_2 + \cdots + k_n + (n-1) = h.$$

Let P_n denote the n-vertex path, and let its vertices be labelled consecutively by $1, 2, \ldots, n$, see Fig. 3.

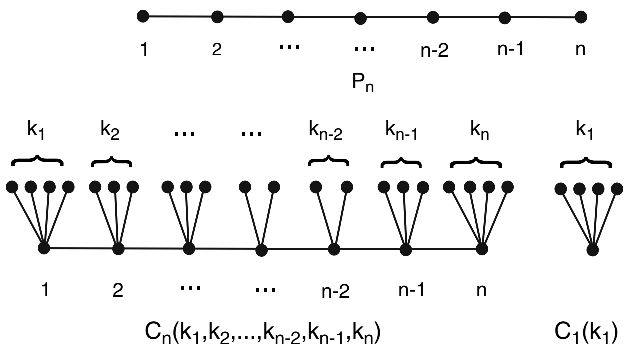

Fig. 3 The n-vertex path P_n and the caterpillar tree C with parameters $k_1, k_2, \ldots, k_n$. In the case $n = 1$ the caterpillar tree $C_1(k_1)$ is just the (k_1+1)-vertex star. Recall that the Z-index of such a star is equal to $k_1 + 1$

Then a *caterpillar tree* C with parameters $k_1, k_2, \ldots, k_n$ is obtained by attaching k_i pendent vertices to the ith vertex of P_n for $i = 1, 2, \ldots, n$. This caterpillar tree will be denoted by $C_n(k_1, k_2, \ldots, k_n)$. It has

$$k_1 + k_2 + \cdots + k_n + n$$

vertices and

$$k_1 + k_2 + \cdots + k_n + (n-1)$$

edges. An example of a caterpillar tree, namely $C_5(3, 2, 1, 0, 1)$, is given in Fig. 1. The general form of a caterpillar tree is displayed in Fig. 3.

The main result obtained in [18] is the following:

Theorem 1 *If H is a hexagonal chain whose LA-sequence is $L^{k_1}AL^{k_2}A\cdots L^{k_{n-1}}AL^{k_n}$, then the jth coefficient $s(H, j)$ of its sextet polynomial is equal to the number $p(C, j)$ of selections of j disjoint edges in the caterpillar tree $C = C_n(k_1, k_2, \ldots, k_n)$.*

Recall that the sextet polynomial of a benzenoid system H is defined as $\sigma(H, x) = \sum_{j\geq 0} s(H, j)\, x^j$, where $s(H, j)$ is equal to the number of generalized Clar formulas of H with exactly j aromatic sextets; for more details on this matter (which are not needed for the present considerations) see [11,18–20]. Knowing that the sum of the coefficients of the sextet polynomial is equal to the Kekulé structure count [11,19], and bearing in mind Eq. 1, we arrive at [18]:

Theorem 2 *If H is a hexagonal chain whose LA-sequence is $L^{k_1}AL^{k_2}A\cdots L^{k_{n-1}}AL^{k_n}$, then the number $K(H)$ of its Kekulé structures is equal to the Z-index of the caterpillar tree $C_n(k_1, k_2, \ldots, k_n)$.*

For the caterpillar tree associated (in the sense of Theorems 1 and 2) with the hexagonal chain H we will write $C(H)$. Then Theorem 2 is tantamount to the equality

$$K(H) = Z(C(H)). \tag{4}$$

Formula (4) is a simple, but quite unusual connection between two such seemingly unrelated concepts as the Kekulé structure count and the Z-index. Its applications were discussed in a number of subsequent papers [21–26].[2]

[2] The caterpillar $C(H)$ is sometimes referred to as the *Gutman tree*.

2 An excursion to classical mathematics

Let p and q be two mutually prime integers, $p < q$. Then the rational number $Q = p/q$ can be presented in the form of a continued fraction [27]:

$$Q = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\cdots + \cfrac{1}{a_{n-1} + \cfrac{1}{a_n}}}}} \tag{5}$$

In what follows we will be concerned only with finite continued fractions, such as (5).

Leonhard Euler (whose 300th anniversary of the birth is just in this year) posed and solved the inverse problem: Suppose that we know the numbers $a_1, a_2, \ldots, a_n$. How can we (in an efficient manner) compute Q?

For this, Euler introduced the so-called *continuants* (or *continuant polynomials*). These are defined recursively via [28]

$$L_n(x_1, x_2, \ldots, x_n) = x_n\, L_{n-1}(x_1, x_2, \ldots, x_{n-1}) + L_{n-2}(x_1, x_2, \ldots, x_{n-2}) \tag{6}$$

with initial conditions

$$\begin{aligned} L_0(*) &= 1 \\ L_1(x_1) &= x_1 \\ L_2(x_1, x_2) &= x_1\, x_2 + 1. \end{aligned}$$

Then,

$$Q = \frac{L_{n-1}(a_2, a_3, \ldots, a_n)}{L_n(a_1, a_2, a_3, \ldots, a_n)}.$$

The continuant satisfies the following tridiagonal determinantal expression:

$$L_n(x_1, x_2, \ldots, x_n) = \begin{vmatrix} x_1 & 1 & 0 & 0 & \cdots & 0 \\ -1 & x_2 & 1 & 0 & \cdots & 0 \\ 0 & -1 & x_3 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & x_{n-1} & 1 \\ 0 & 0 & \cdots & 0 & -1 & x_n \end{vmatrix}.$$

3 On the Z-index of caterpillar trees

If G is any graph, and if e is its edge connecting the vertices u and v, then the Z-index of G can be recursively calculated by means of the relation [1]

$$Z(G) = Z(G - e) + Z(G - u - v).$$

Consider now a caterpillar tree $C_n(k_1, k_2, \dots, k_n)$ and its edge e_{n-1} connecting the vertices labelled by $n-1$ and n (see Fig. 3). We denote these two vertices by v_{n-1} and v_n. Then the subgraph $C_n(k_1, k_2, \dots, k_n) - e_{n-1}$ consists of the union of the caterpillar tree $C_{n-1}(k_1, k_2, \dots, k_{n-1})$ and the $(k_n + 1)$-vertex star. Further, $C_n(k_1, k_2, \dots, k_n) - v_{n-1} - v_n$ consists of the union of the caterpillar tree $C_{n-2}(k_1, k_2, \dots, k_{n-2})$ and $k_{n-1} + k_n$ isolated vertices.

In view of Eqs. 2 and 3 and the fact that the Z-index of the t-vertex star is equal to t, we have

$$Z(C_n(k_1, k_2, \dots, k_n) - e_{n-1}) = Z(C_{n-1}(k_1, k_2, \dots, k_{n-1}) \times (k_n + 1)$$
$$Z(C_n(k_1, k_2, \dots, k_n) - v_{n-1} - v_n) = Z(C_{n-2}(k_1, k_2, \dots, k_{n-2}))$$

from which follows

$$Z(C_n(k_1, k_2, \dots, k_n)) = (k_n + 1)\, Z(C_{n-1}(k_1, k_2, \dots, k_{n-1})) + Z(C_{n-2}(k_1, k_2, \dots, k_{n-2})). \tag{7}$$

The initial conditions for this recursion relation are obtained by direct and easy calculation:

$$Z(C_0(*)) = 1$$
$$Z(C_1(k_1)) = k_1 + 1$$
$$Z(C_2(k_1, k_2)) = (k_1 + 1)(k_2 + 1) + 1$$

By comparing (7) and its initial conditions, with the recursion relation (6) and its initial conditions, we immediately recognize that the Z-index of the caterpillar trees coincides with Euler's continuant. In particular,

$$Z(C_n(k_1, k_2, \dots, k_n)) = L_n(k_1 + 1, k_2 + 1, \dots, k_n + 1). \tag{8}$$

This remarkable connection between the Z-index of caterpillar trees and continuants was recently discovered by one of the present authors [29].

By combining (8) with Eq. 4 we arrive at some bizarre, hitherto not reported, identities for the Kekulé structure count of hexagonal chains.

Before doing this we establish here another expression for the Z-index of caterpillar trees.

Consider the n-vertex path P_n and the caterpillar tree $C = C_n(k_1, k_2, \ldots, k_n)$, shown in Fig. 3. Denote by e_i the edge of both P_n and C, connecting the vertices labelled by i and $i + 1$.

Let M be a selection of some disjoint edges of P_n and let $\mathcal{M}(P_n)$ be the set of all such selections, including the selections of a single edge and the (unique) "selection" of no edges. Clearly, $Z(P_n) = |\mathcal{M}(P_n)|$. In mathematics, $\mathcal{M}(P_n)$ is called the set of matchings of P_n. Its size, $|\mathcal{M}(P_n)|$, is equal to the nth Fibonacci number [30].

For example, in the case of P_{10}, a selection of disjoint edges may be $M^* = \{e_2, e_5, e_9\}$. In the case of P_5, the set $\mathcal{M}(P_5)$ consist of the following eight elements: $\emptyset$, $\{e_1\}$, $\{e_2\}$, $\{e_3\}$, $\{e_4\}$, $\{e_1, e_3\}$, $\{e_1, e_4\}$, $\{e_2, e_4\}$.

By definition (1), the Z-index of the caterpillar tree $C = C_n(k_1, k_2, \ldots, k_n)$ is equal to the total number of selections of its disjoint edges. These selections can be grouped according to the edges $e_1, e_2, \ldots, e_{n-1}$ which they contain. Each such group corresponds to an element $M \in \mathcal{M}(P_n)$, and therefore

$$Z(C) = \sum_{M \in \mathcal{M}(P_n)} Z(C - [M]) \tag{9}$$

where $C - [M]$ denotes the subgraph obtained by deleting from C all edges $e_1, e_2, \ldots, e_{n-1}$, those vertices that are endpoints of the edges from M, and the pendent vertices attached to them. It is easily seen that $C - [M]$ is a union of stars consisting of vertices $1, 2, \ldots, n$ that are not endpoints of the edges from M, and the pendent vertices attached to them.

An illustrative example is depicted in Fig. 4. In this example $n = 10$ and $M^* = \{e_2, e_5, e_9\}$. The endpoints of the edges from M^* are 2, 3, 5, 6, 9, 10. Therefore $C - [M^*]$ consists of four stars, involving the vertices 1, 4, 7, 8, and

$$Z(C - [M^*]) = (3 + 1)(5 + 1)(3 + 1)(2 + 1)$$

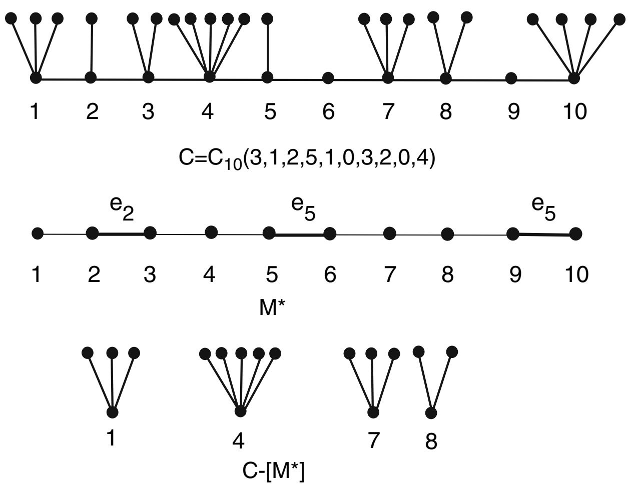

Fig. 4 An example illustrating formulas (9)–(11). For details see text

because these stars possess 3, 5, 3, and 2 pendent vertices, i.e., $k_1 = 3, k_4 = 5, k_7 = 3, k_8 = 2$.

Because for a star with k pendent vertices, $Z = k + 1$, from (9) follows

$$Z(C) = \sum_{M \in \mathcal{M}(P_n)} \prod_{i \notin M} (k_i + 1) \tag{10}$$

with $i \notin M$ indicating that the vertex $i \in \{1, 2, \ldots, n\}$ is not an endpoint of any of the edges contained in M. Another form of the same identity is:

$$Z(C) = \left(\prod_{i=1}^{n} (k_i + 1) \right) \sum_{M \in \mathcal{M}(P_n)} \prod_{e_i \in M} \frac{1}{(k_i + 1)(k_{i+1} + 1)}. \tag{11}$$

In order to illustrate formulas (10) and (11) we apply them for the case $n = 5$. The eight elements of $\mathcal{M}(P_5)$ have been specified above. Using these in the earlier given order we get:

$$\begin{aligned} &Z(C(k_1, k_2, k_3, k_4, k_5)) \\ &\quad = (k_1 + 1)(k_2 + 1)(k_3 + 1)(k_4 + 1)(k_5 + 1) + (k_3 + 1)(k_4 + 1)(k_5 + 1) \\ &\qquad + (k_1 + 1)(k_4 + 1)(k_5 + 1) + (k_1 + 1)(k_2 + 1)(k_5 + 1) \\ &\qquad + (k_1 + 1)(k_2 + 1)(k_3 + 1) + (k_5 + 1) + (k_3 + 1) + (k_1 + 1) \end{aligned}$$

and

$$\begin{aligned} &Z(C(k_1, k_2, k_3, k_4, k_5)) \\ &\quad = (k_1 + 1)(k_2 + 1)(k_3 + 1)(k_4 + 1)(k_5 + 1) \left[1 + \frac{1}{(k_1 + 1)(k_2 + 1)} \right. \\ &\qquad + \frac{1}{(k_2 + 1)(k_3 + 1)} + \frac{1}{(k_3 + 1)(k_4 + 1)} + \frac{1}{(k_4 + 1)(k_5 + 1)} \\ &\qquad + \frac{1}{(k_1 + 1)(k_2 + 1)(k_3 + 1)(k_4 + 1)} + \frac{1}{(k_1 + 1)(k_2 + 1)(k_4 + 1)(k_5 + 1)} \\ &\qquad \left. + \frac{1}{(k_2 + 1)(k_3 + 1)(k_4 + 1)(k_5 + 1)} \right]. \end{aligned}$$

This, of course, is just another way of writing the continuant-based formula (8).

4 Formulas for Kekulé structure count of hexagonal chains

Combining the relation (4) with the above stated expressions for the Z-index of caterpillar trees we arrive at a series of identities for $K(H)$.

Theorem 3 *If H is a hexagonal chain whose LA-sequence is $L^{k_1} A L^{k_2} \ldots A L^{k_n}$, then the number of Kekulé structures of H is equal to the continuant $L_n(k_1 + 1, k_2 + 1, \ldots, k_n + 1)$.*

Theorem 4 *If H is a hexagonal chain whose LA-sequence is $L^{k_1}AL^{k_2}\dots AL^{k_n}$, then the number of Kekulé structures of H is equal to*

$$\begin{vmatrix} k_1+1 & 1 & 0 & 0 & \cdots & 0 \\ -1 & k_2+1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & k_3+1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & k_{n-1}+1 & 1 \\ 0 & 0 & \cdots & 0 & -1 & k_n+1 \end{vmatrix}.$$

Theorem 5 *In the notation specified in Eqs. 10 and 11, the number of Kekulé structures of a hexagonal chain whose LA-sequence is $L^{k_1}AL^{k_2}\dots AL^{k_n}$, is equal to*

$$\sum_{M\in\mathcal{M}(P_n)}\prod_{i\notin M}(k_i+1)$$

or

$$\left(\prod_{i=1}^{n}(k_i+1)\right)\sum_{M\in\mathcal{M}(P_n)}\prod_{e_i\in M}\frac{1}{(k_i+1)(k_{i+1}+1)}.$$

We do not claim that the above relations are of great practical value for enumerating the Kekulé structures of hexagonal chains, but these certainly shed light on some concealed connections between various, seemingly unrelated, fields of mathematics and mathematical chemistry.

References

1. H. Hosoya, Bull. Chem. Soc. Jpn. **44**, 2332 (1971)
2. H. Hosoya, Internet Electr. J. Mol. Des. **1**, 428 (2002)
3. H. Hosoya, Croat. Chem. Acta **80**, 239 (2007)
4. G.G. Cash, J. Math. Chem. **37**, 117 (2005)
5. A. Yu, X. Lv, J. Math. Chem. **41**, 34 (2007)
6. H. Hosoya, J. Chem. Inf. Model. **47**, 744 (2007)
7. S.G. Wagner, MATCH Commun. Math. Comput. Chem. **57**, 221 (2007)
8. X. Pan, C. Yang, M.J. Zhou, MATCH Commun. Math. Comput. Chem. **57**, 235 (2007)
9. H. Liu, X. Yan, Z. Yan, MATCH Commun. Math. Comput. Chem. **57**, 371 (2007)
10. S.J. Cyvin, I. Gutman, *Kekulé Structures in Benzenoid Hydrocarbons* (Springer-Verlag, Berlin, 1988)
11. I. Gutman, S.J. Cyvin, *Introduction to the Theory of Benzenoid Hydrocarbons* (Springer-Verlag, Berlin, 1989)
12. M. Gordon, W.H.T. Davison, J. Chem. Phys. **20**, 428 (1952)
13. D. Cvetković, I. Gutman, Croat. Chem. Acta **46**, 15 (1974)
14. S.J. Cyvin, Monatsh. Chem. **114**, 13 (1983)
15. T.P. Živković, N. Trinajstić, Chem. Phys. Lett. **136**, 141 (1987)
16. A.T. Balaban, C. Artemi, I. Tomescu, MATCH Commun. Math. Chem. **22**, 77 (1987)
17. R. Tošić, Stud. Phys. Theor. Chem. **63**, 123 (1989)
18. I. Gutman, Theor. Chim. Acta **45**, 309 (1977)
19. H. Hosoya, T. Yamaguchi, Tetrahedron Lett. 4659 (1975)
20. I. Gutman, B. Furtula, A.T. Balaban, Polycycl. Aromat. Compds. **26** 17 (2006)

21. I. Gutman, S. El-Basil, Z. Naturforsch. **40a**, 923 (1985)
22. S. El-Basil, J. Chem. Soc. Faraday Trans. **2**(82), 299 (1986)
23. S. El-Basil, J. Math. Chem. **1**, 153 (1988)
24. S. El-Basil, M. Randić, J. Chem. Soc. Faraday Trans. **2**(84), 1875 (1988)
25. S. El-Basil, Topics Curr. Chem. **153**, 273 (1990)
26. H. Zhang, Discr. Math. **172**, 163 (1997)
27. A.Y. Khintchine, *Continuous Fractions* (Noordhoff, Groningen, 1963)
28. L. Graham, D.E. Knuth, O. Patashnik, *Concrete Mathematics. A Foundation for Computer Science* (Addison-Wesley, Reading, 1989), p. 287
29. H. Hosoya, Nat. Sci. Rep. Ochanomizu Univ., in press
30. H. Hosoya, Fibonacci Quart. **11**, 255 (1973)