The Hosoya Index, Lucas Numbers, and QSPR

T. M. Westerberg, K. J. Dawson, and K. W. McLaughlin
Department of Chemistry
University of Wisconsin-River Falls

Abstract

The 2-methyl alkanes, the cyclic alkanes, and the 1-methyl bicyclo[X.1.0] alkanes are shown to possess identical values for the Hosoya index. For these specific molecular graphs, this topological index generates the integer sequence known as Lucas numbers. The implication of these results for quantitative structure-property relationship development is illustrated using normal boiling point data.

Introduction

The number of different molecules with just the formula $C_{64}H_{130}$ is more than one hundred million billion times greater that the number of different molecules of all types (and formulas) cataloged in all of human history. Therefore, it is imperative that theoretical chemists develop methods to predict properties of molecules from their structures so that synthetic chemists can identify on which of the enumerable molecular structures they should expend their finite resources. This is one of the primary reasons for developing quantitative structure-property relationships (QSPR).

The Hosoya index (Z) is an example of a graph invariant, called a topological index, which may be calculated directly from the structure of a molecule [1]. Topological indices have proven to be very useful in QSPR models, especially when a physical property such as the normal boiling point is modeled for a specific family of molecular graphs [2]. This is illustrated in Figure 1 where the high correlation between the Hosoya index and the boiling points of the acyclic alkanes is shown [3].
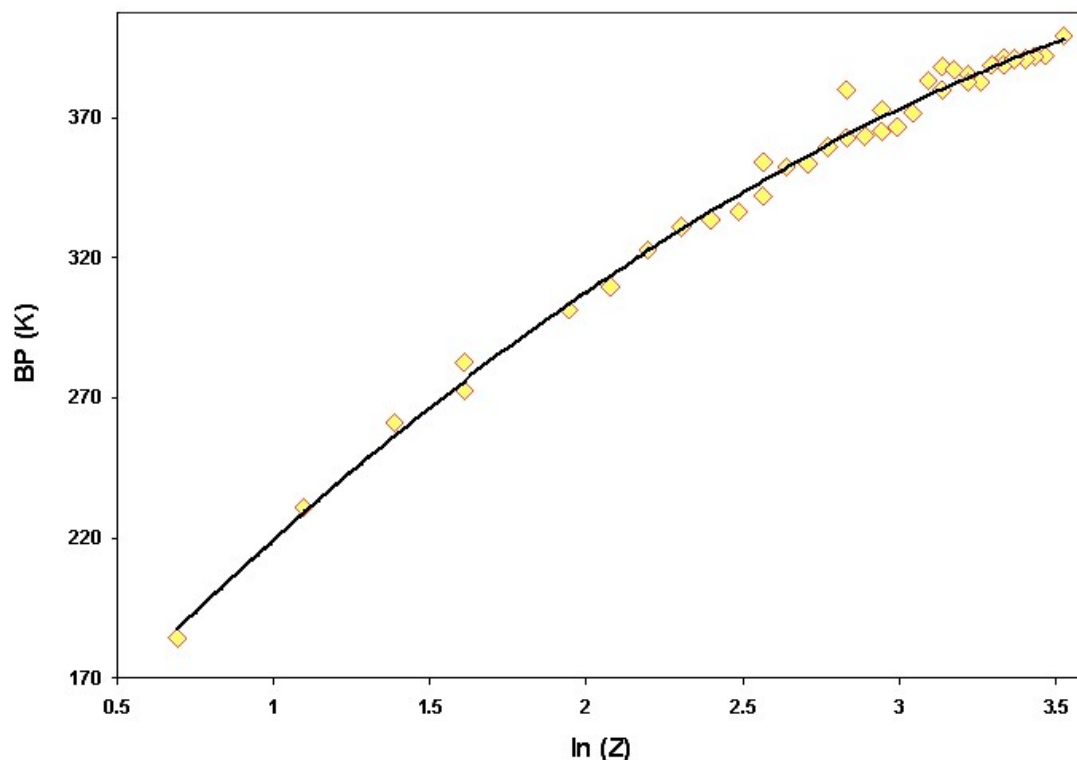
Figure 1. Normal boiling points plotted versus the natural logarithm of the Hosoya index (Z) for all 40 acyclic alkanes from C_2H_6 to C_8H_{18} [3]. The best-fit second-order polynomial is shown as the black line.

Since topological indices represent a mathematical definition arising from the field of graph theory and do not come from the laws of physics which govern the physical properties of molecules, two important questions confronting the theoretical chemistry community have been: "Why do topological indices do such a good job in QSPR?" and "What physical significance do topological indices have?"

As part of a systematic study of the applications of chemical graph theory to macromolecules, we report three families of molecular structures that generate Lucas numbers for the Hosoya index, and what the implication of this result is for QSPR model development.

The Hosoya Index

The alkanes are the most extensively studied class of molecules for QSPR [2]. The hydrogen-depleted molecular graphs are what topological indices are usually applied to. The Hosoya index Z is defined as the number of sets of nonadjacent bonds (edges) in a molecule (graph) [4]. Table 1 shows how the Hosoya index is determined for 2-cyclopropyl butane.
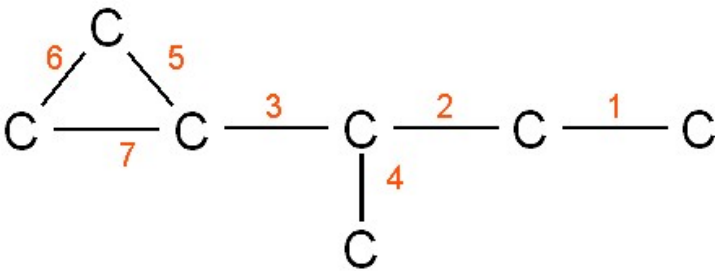

Number of Edges Selected (**t**)	Number of Sets of **t** Non-adjacent Edges
0	1
1	7
2	12
3	4
$\geq$ 4	0
Total	24 = Z

Table 1. Determination of the Hosoya index by direct enumeration for the hydrogen-depleted graph of 2-cyclopropyl butane. To facilitate the enumeration, the dges(bonds) are labeled in red.

From this definition, Z is determined by listing all sets of nonadjacent edges. Thus, in this example, there is one way to select zero edges, seven ways to select one edge, twelve ways to select two nonadjacent edges, and zero ways of selecting four or more nonadjacent edges. The total of all ways, i.e. the value of Z for the hydrogen depleted graph of 2-cyclopropyl butane, is 24.

The method of listing all the combinations of nonadjacent edges in a molecular graph becomes overwhelming for large molecules. Therefore, a second method of evaluating Z is needed. The *Hosoya*

Composition Theorem (HCT) allows Z to be determined from the value of Z for smaller graphs [4]. This theorem states that the Hosoya index for a molecule is equal to the product of the Hosoya index for each connected fragment produced by deleting any single edge plus the product of the Hosoya index for each connected fragment produced by deleting all the edges adjacent to the initial edge that was deleted. Figure 2 illustrates how the Hosoya Composition Theorem works.
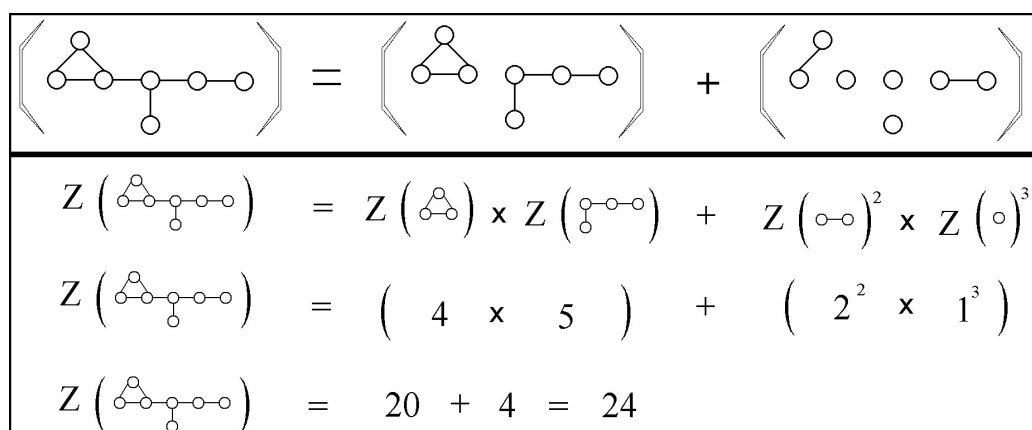

Figure 2. Application of the Hosoya Composition Theorem to 2-cyclopropyl butane.

In this paper, the special brackets in the top line of Figure 2 (Z-notation) will be used as a shorthand notation to represent the product of the Hosoya index for each connected fragment of the initial graph inside the brackets.

The HCT makes it possible to identify recurrence relationships where Z for one molecule equals the sum of the products of the Z's for smaller molecules. The most famous example is that for the linear alkanes which connects the Hosoya index to the Fibonacci numbers [4]. Recall that the Fibonacci numbers are given by the sequence $F_1 = F_2 = 1$, and for n greater than 2, F_n is defined as the sum of the two preceding numbers in the sequence (i.e. $F_{n-1} + F_{n-2}$). For the linear alkanes composed of n carbon atoms,

the HCT produces the following recurrence relation and leads to Z_n equaling the Fibonacci number F_{n+1},

since $Z_0 = Z_1 = 1$.

$$Z_n \;=\; Z_{n-1} \;+\; Z_{n-2} \;=\; F_{n+1} \qquad\qquad (1)$$

In the preceding example and all subsequent equations, the subscript n on Z will indicate the number of

carbon atoms in the parent molecule. The way a molecular graph may grow, one vertex after another, will

be indicated by where the n is placed on the graph. The general notation Z[G(n)] represents the Hosoya

index for a specific family of graphs, while the specific value for n will identify which member of the

family is being considered. The notation Z[G(n)] will be abbreviated as Z_n where the family of graphs

under consideration will be specified in the discussion.

The 2-Methyl Alkanes

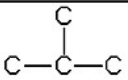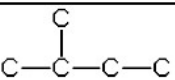   In order to find the recurrence relation arising from the Hosoya index for the 2-methyl alkanes,
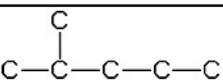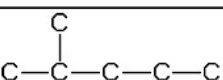
Number of Carbon Atoms	Name	H-depleted Molecular Graph	Z
3	propane	C—C with C above	3
4	2-methyl propane	C—C—C with C above	4
5	2-methyl butane	C—C—C—C with C above	7
6	2-methyl pentane	C—C—C—C—C with C above	11
$(n)_{n-5}$	----	C—C—C—C—C with C above	$F_n + F_{n-2}$

Table 2. The Hosoya index for the first four 2-methyl alkanes and the general case composed of n carbon atoms are shown. The vertices, representing carbon atoms, are designated by the letter C.

the smallest molecular examples, shown in Table 2, were considered. The values of Z shown in the above

table can be quickly obtained for all but the last case by direct enumeration of the sets of nonadjacent edges

in the corresponding molecular graphs. For the nth case, the HCT was used as illustrated below with the

notation introduced in Figure 2.

$$\left(\includegraphics{} \right) = \left(\includegraphics{} \right) + \left(\includegraphics{} \right) \tag{2}$$
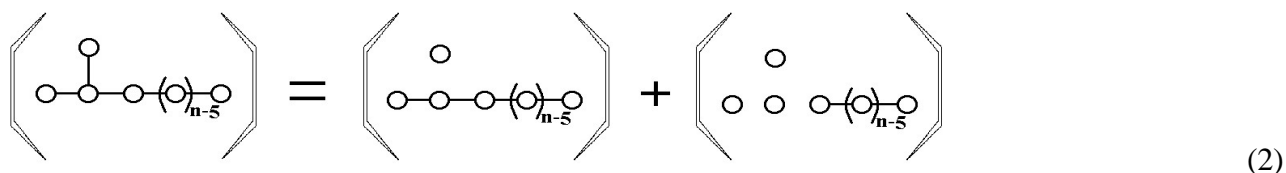
The selection of the first edge to be deleted is critical to the recurrence relation that arises from the HTC.

Equation 2 arises from the deletion of the vertical edge in the original graph. Since the fragments on the

right side of equation 2 are all linear alkanes, they can be replaced with their corresponding Fibonacci

numbers.

$$Z_n \quad = \quad F_n \quad + \quad F_{n-2} \quad = \quad L_{n-1} \tag{3}$$

This produces the well known sequence of Lucas numbers. One of the many ways of defining the Lucas

number sequence is as the sum of the $n+1^{th}$ Fibonacci number plus the $n-1^{th}$ Fibonacci number. Equation

3 produces the L_{n-1} Lucas number for the 2-methyl alkane possessing n carbon atoms.

An alternative application of the HCT, that is selecting the right most (terminal) edge, as the initial

edge to be deleted yields

$$\left(\includegraphics{} \right) = \left(\includegraphics{} \right) + \left(\includegraphics{} \right) \tag{4}$$

and
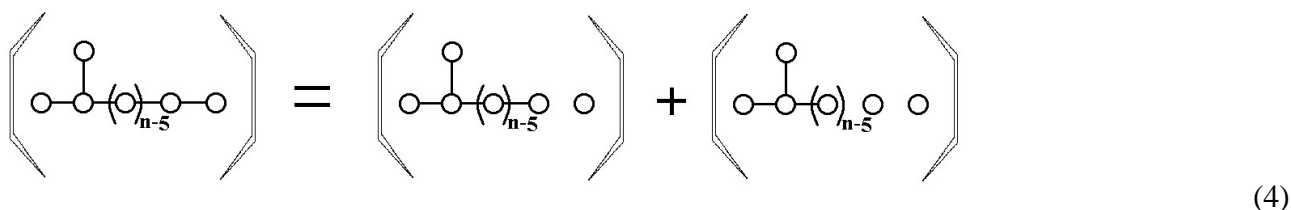
$$Z_n \quad = \quad [\ Z_{n-1} \ \text{x} \ Z(o) \ ] \quad + \quad [\ Z_{n-2} \ \text{x} \ Z(o) \ \text{x} \ Z(o) \ ] \tag{5}$$

Note that the definition of Z_n is the same as in equation 2. Since Z for a lone carbon (vertex) represented

by $Z(o)$ equals one, the equation simplifies to

$$Z_n \quad = \quad Z_{n-1} \quad + \quad Z_{n-2} \tag{6}$$

It was noted, that the analytical solution for this recurrence relation is very accurately approximated by the

last term in equation 7.

$$Z_n = \left(1 - \sqrt{5}\right)^n + \left(1 + \sqrt{5}\right)^n \tag{7}$$

Equation 7 makes it easy to evaluate the Hosoya index for any 2-methyl alkane by just specifying the

number of carbon atoms. It also illustrates that the logarithm of Z_n is linearly related to n.

As shown in Figure 3, the experimentally determined boiling points for the 2-methyl alkanes are

highly correlated with the logarithm of Z_n. The subscript on the Hosoya index has been left off in
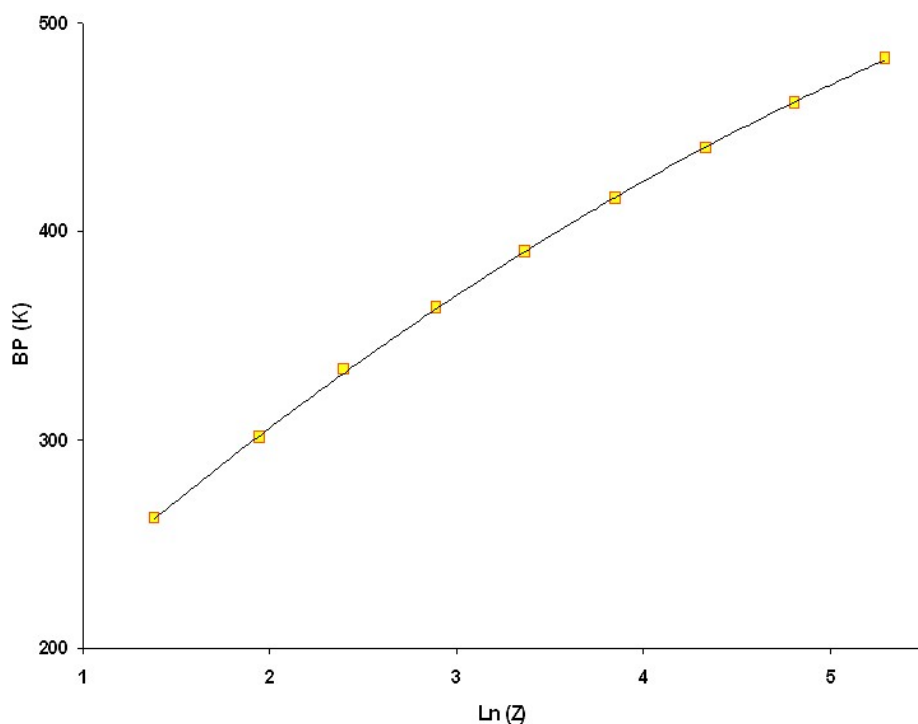

Figure 3. Normal boiling points plotted versus the natural logarithm of the Hosoya index for the 2-methyl alkanes from C_3H_8 to $C_{11}H_{24}$ [3]. The line represents the best fit second-order polynomial.

Figure 3. Of course, the Hosoya index remains a function of n for the specific graph under consideration.

The reason for leaving the index off is that, in subsequent figures where comparisons are being drawn

between different types of molecules, the index n will not always be the same even though the numerical

value of the Hosoya index will be.

The Cyclic Alkanes

Another family of structurally related molecules is the unsubstituted monocyclic alkanes. Table 3

shows the first few molecules in this series. Direct enumeration gives the value of Z for
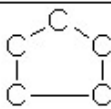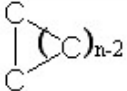
n	Name	Structure of Molecule	Z
3	Cyclopropane		4
4	Cyclobutane		7
5	Cyclopentane		11
6	Cyclohexane		18
n	----		$F_{n+1} + F_{n-1}$

Table 3. The Hosoya index for the first four cyclic alkanes and the general case composed of n carbon atoms.

the smallest members and suggest that the same integer sequence is being produced as for the 2-methyl

alkanes. Application of the HCT to the nth cycloalkane, choosing any edge, results in

$$\left(\triangle \right)_{n-2} = \left(\triangle \right)_{n-2} + \left(\circ \circ \right)_{n-2} \tag{8}$$
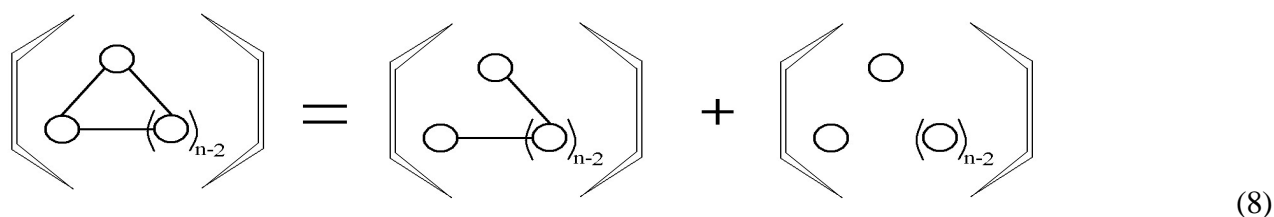
Since linear fragments are generated, equation 8 simplifies to

$$Z_n = F_{n+1} + F_{n-1} = L_n \tag{9}$$

The recurrence relation given by equation 8 generates the Lucas numbers. For this class of molecules, Z_n

generates the nth Lucas number, as previously reported [5]. Note that Z_n now represents the Hosoya index

for the monocycloalkane composed of n carbon atoms.

1-Methyl Bicyclo[x.1.0] Alkanes

So far, one family of acyclic alkanes and one family of monocyclic alkanes have been

n	Name	Structure of Molecule	Z_n
5	1-methyl-[3.1.0] bicyclobutane		11
6	1-methyl-[4.1.0] bicyclopentane		18
7	1-methyl-[5.1.0] bicyclohexane		29
8	1-methyl-[6.1.0] bicycloheptane		47
n	---		$F_{n+1} + F_{n-1}$

Table 4. The Hosoya index for the first four 1-methyl bicyclo[x.1.0] alkanes and the general case composed of n carbon atoms.

shown to generate Lucas numbers for the Hosoya index. One family of bicyclic alkanes has been

found which is relevant to this study, the 1-methyl bicyclo[n-3.1.0] alkanes. Table 4 shows the first few
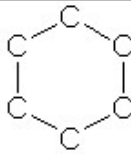
molecules in this series. Direct enumeration produces Lucas numbers for the Hosoya index for the smallest

of these bicyclic molecules. Application of the HCT to the nth case by first deleting the internal edge

yields:

$$(11)$$
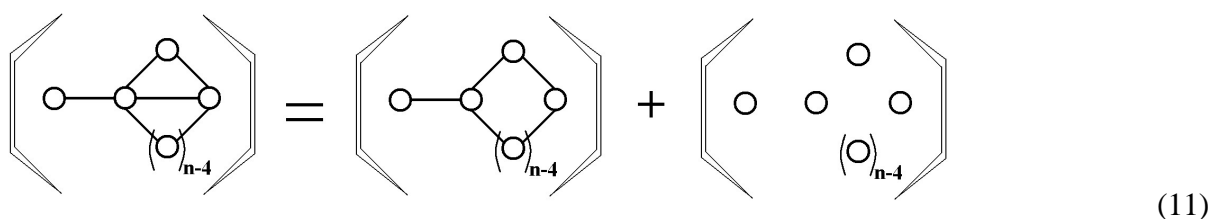
The second term on the right hand side of equation 11 is a linear alkane, which can be replaced by the

Fibonacci number F_{n-3}. Applying the HCT to the first term on the right in equation 11 can produce a linear

alkane which will result in

$$Z_n = (F_{n+1} + F_{n-2}) + F_{n-3} \tag{12}$$

Using the recurrence relation for Fibonacci numbers, equation 12 simplifies to

$$Z_n = F_{n+1} + F_{n-1} = \mathbf{L}_n \tag{13}$$

which again generates the Lucas numbers, now for a family of bicyclic alkanes, where n on the Hosoya

index continues to define the number of carbon atoms.

Boiling Point Analysis

We now have a triply degenerate set of molecules for the Hosoya index (i.e. three distinct graphs all

possessing the same value of Z) coming from an acyclic, a monocyclic and a bicyclic family of alkanes.

This allows the normal boiling point for three different alkanes to be compared when each has the same

value for their Hosoya index, and this comparison can be extended as each member of the respective alkane

family is allowed to grow one carbon at a time. Since the Hosoya index is known to be highly correlated

with boiling point for all molecules composed of ten or fewer carbons as well as in many other cases, one

might expect molecules having the same value of Z to have similar boiling points (i.e. varying by a degree

or two based on Figure 1). This is not the case when comparing molecules from the three families of

molecular structures under consideration here.   As Figure 4 illustrates, the three families of molecular
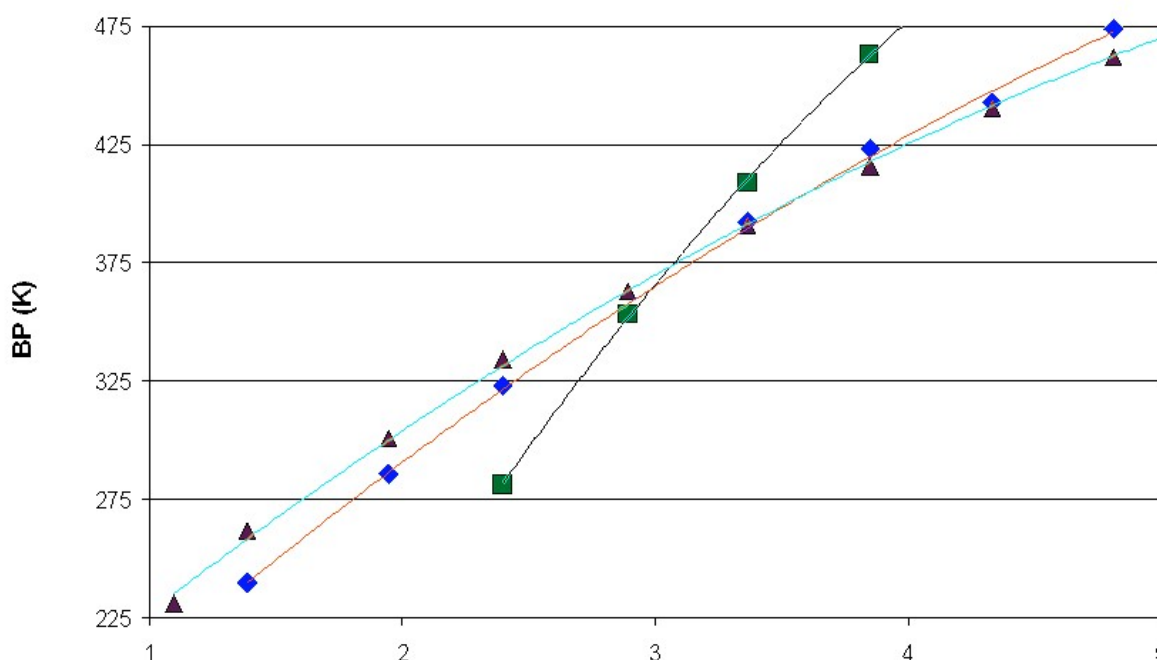
structures do not exhibit the same behavior.

Figure 4.   The normal boiling points in Kelvin versus the natural logarithm of the Hosoya index for each

member of the three families of molecular graphs: purple triangles = acyclic alkanes, blue diamonds =

monocyclic alkanes, and green squares = bicyclic alkanes.

In fact, typically the sets of three distinct molecular graphs having the same Hosoya index do not have the

same boiling points, as can be seen by looking at any vertical line passing through the data points in Figure

4. Furthermore, there are differences in the trends in the boiling point as a function of the Hosoya index and

therefore as a function of n. The trend lines in Figure 4 are just there to help illustrate the differences

between the three families of alkanes.

   In Figure 5, the difference between the normal boiling points for unsubstituted, monocyclic alkanes

and the 2-methyl alkanes having the same value of Z are plotted. In this figure, not only is the magnitude

of the difference between the boiling points visible, but so is a trend in this difference between boiling

points.
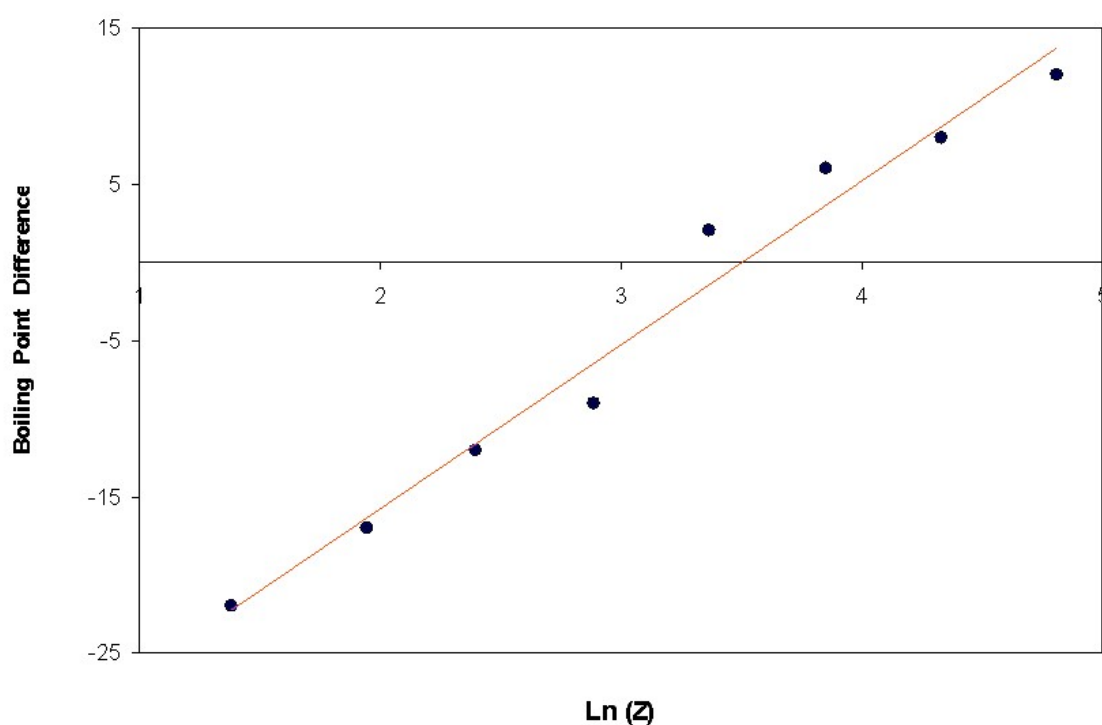
Figure 5.   Plot of the normal boiling point difference in Kelvin between the n carbon cycloalkane and the

n+1 carbon 2-methyl alkane. The best fit line with a regression factor (r^2) of 0.97 is shown in red to

illustrate the correlation.

   Note that not only are the differences between the pairs of molecules with the same Hosoya index

not small, but the differences change spanning a range of $34^{o}C$. The data taken from the NIST WebBook

includes all boiling points measured after 1900 [3]. One data point, from 1887, which exhibited a large but abnormal difference had never been duplicated and was therefore omitted.

Even more pronounced trends exist for the differences in the boiling points for the other two combinations of molecules possessing identical values of Z: 2-methyl alkanes with 1-methyl bicyclo[X.1.0] alkanes, and cycloalkanes with 1-methyl bicyclo[X.1.0] alkanes. As can be seen from Table 5, the difference ranges over 80°C for the boiling points for the four pairs of compounds in the latter series (monocyclic and bicyclic), while the four pairs of compounds in the former series (acyclic and bicyclic) span nearly $100\,^{\circ}$C. These are remarkably pronounced differences considering that only three carbons are added across each family.

Boiling Points (K)					
Z	2-methyl alkane		cycloalkane		1-methyl bicyclo [X.1.0] alkane	
3	C_3H_8	231.1	---	---	---	---
4	C_4H_{10}	262	C_3H_6	240	---	---
7	C_5H_{12}	301.1	C_4H_8	284	---	---
11	C_6H_{14}	334	C_5H_{10}	322.4	C_5H_8	281.6
18	C_7H_{16}	363.2	C_6H_{12}	353.9	C_6H_{10}	353
29	C_8H_{18}	392	C_7H_{14}	392.0	C_7H_{12}	409
47	C_9H_{20}	416.1	C_8H_{16}	422	C_8H_{14}	463
76	$C_{10}H_{22}$	440.1	C_9H_{18}	448	---	---
123	$C_{11}H_{24}$	462.3	$C_{10}H_{20}$	474.4	---	---
199	$C_{12}H_{26}$	483.2	---	---	---	---

Table 5. The normal boiling points for the three families of alkanes discussed in this paper. The boiling points for the 2-methyl alkanes and the cycloalkanes were taken from reference 3, while the boiling points for the bicycloalkanes were taken from reference 6.

These results suggest that a relationship may exist between the boiling points of these three families of molecules which is both significant and unaccounted for by the Hosoya index. The unaccounted for trends

in the differences of the boiling points correspond to "the more cyclized a molecule is the more pronounced this trend becomes," a factor which would need to be built into any QSPR model involving boiling point.

Conclusions

The Hosoya index for the 2-methyl alkanes, the cycloalkanes, and the 1-methyl bicyclo[X.1.0] alkanes is degenerate. This topological index generates Lucas numbers for all three families of alkanes. Normal boiling point comparisons show that within each family, the correlation with Z_n is high and similar. However, comparison of molecules possessing the same value for the Hosoya index shows significant differences in the normal boiling points. Furthermore, these differences show a trend as higher molecular analogs are examined. From the boiling point data available for these three families of alkanes, it appears as though the range of differences in the normal boiling point relative to the acyclic family of alkanes becomes greater the more cyclized a molecule becomes. This hypothesis is currently being tested with other families of alkanes, i.e. those possessing values of Z other than the Lucas numbers.

Finally, this work suggests a series of questions for future consideration. First, how many other types of graphs whose Hosoya index generate Lucas numbers are there? The following tetracyclic graph which does not match any known molecular structure (the reason it was left out of the preceding boiling
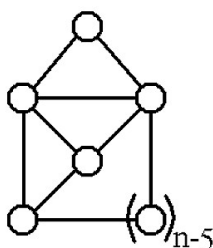
point discussion) is the only other "Lucas number" example we have found so far. The existence of this non-molecular graph, raises a second question: If there are more types of Lucas number generating graphs,

how many of them are there and how many have molecular counterparts? Also, since all four of the known

types have different numbers of cycles, is there a three-cycle family which generates Lucas numbers?

References

1. Trinajstic, N., Chemical Graph Theory, 2nd Ed., **1992**, CRC Press, London, Ch. 10.

2. Mihalic, Z. and Trinajstic, N., *J. Chem. Ed.*, **1992**, 69(9), 701-712.
    "A Graph-Theoretical Approach to Structure-Property Relationships"

3. Boiling point data was taken from the NIST Chemistry WebBook
    http://webbook.nist.gov/chemistry/

4. Hosoya, H., *Bull. Chem. Soc. Jpn*., **1971**, 44, 2332-2339.
    "A Newly Proposed Quantity Characterizing the Topological Nature of Structural
     Isomers of Saturated Hydrocarbons"

5. Hosoya, H., *Fib. Quart*., **1973**, 11(3), 255-266.
    "Topological Index and Fibonacci Numbers with Relation to Chemistry"

6. Boiling point data for the bicyclic molecules was taken from Beilstein:
    http://www.mdli.com/products/knowledge/crossfire_beilstein/index.jsp