

Figure 5. Complexity comparison of linear and convergent syntheses.

lessened, i.e., $Q\eta \lg \eta$, with $2 > Q > 1$.

The application of complexity calculations to assess synthesis efficiency has been illustrated by Bertz, who has used the sum of complexities of intermediates as a comparative measure for the efficiency of synthesis.¹ In comparisons of syntheses for the same target, however, this will generally simply prefer the synthesis with the fewest steps, hence fewest intermediates to sum. However, in a comparison of two syntheses with the same starting fragments and the same number of steps, this measure should show a preference for one. This amounts to two different orders of assembly of the same units and is characteristic of the difference between a convergent and a linear synthesis.⁶ If we apply this method to the assembly of a linear target skeleton of 16 atoms from 8 2-carbon starting units, the results

for the two orders are shown in Figure 5 (functional groups and the minor symmetry terms are not included). The results show a clear preference for the convergent order, as expected from other considerations.⁶ This procedure points up another interesting conclusion about synthesis. In a real convergent synthesis with added refunctionalization reactions, the total complexity of intermediates will be much less if these added reactions precede the final coupling of the two skeletal halves. The conclusion for synthesis planning is clear: the final coupling of intermediates should come near the end rather than the beginning of a convergent synthetic route. In more general terms, any step that exhibits a large increase in complexity is more efficiently positioned near the end rather than the beginning of a synthetic sequence.

The calculation of molecular complexity is rendered much easier to carry out by hand when the variations in eq 1, 3, 5, and 7 are used and the values from Tables I and II are applied. The method presented here, used by hand or computer, yields the same complexity values as the Bertz method. Also, it is less liable to error than identifying and counting bond connectivities, when used by hand, and much more amenable to computerization in this way as well. The program CPXCAL⁷ is available to anyone with an interest in comparing molecular complexities in any molecular families of interest.

ACKNOWLEDGMENT

We are grateful to the National Science Foundation for a grant (CHE-9102972) that has generously supported this work.

REFERENCES AND NOTES

- (a) Bertz, S. H. *Chem. Commun.* **1981**, 818. (b) Bertz, S. H. *J. Am. Chem. Soc.* **1981**, 103, 3599; **1982**, 104, 5801. (c) Bertz, S. H. *Bull. Math. Biol.* **1983**, 45, 849. (d) Bertz, S. H. In *Chemical Applications of Topology and Graph Theory*; King, R. B., Ed.; Elsevier: New York, 1983; p 206.
- Hendrickson, J. B.; Grier, D. L.; Toczko, A. G. *J. Am. Chem. Soc.* **1985**, 107, 5228.
- The three-atom skeleton extracted may be of any three linked, non-hydrogen atoms, i.e., C-C-C, C-O-C, C-N-S, etc.
- In eq 4, h is the number of attached hydrogens plus the number of unshared electron pairs, regarded as the conjugate bases of potentially attached hydrogens.
- Hendrickson, J. B.; Toczko, A. G. *J. Chem. Inf. Comput. Sci.* **1983**, 23, 171.
- Hendrickson, J. B. *J. Am. Chem. Soc.* **1977**, 99, 5439.
- CPXCAL is written for a DEC VAX computer with an input module for direct graphic input from a Tektronix graphics terminal.

An Algorithm To Identify and Count Coplanar Isomeric Molecules Formed by the Linear Fusion of Cyclopentane Modules

SEYMOUR B. ELK

Elk Technical Associates, New Milford, New Jersey 07646

Received September 18, 1985

Because each of the various possible isomers formed by the coplanar linear fusion of cyclopentane modules may be represented by a binary sequence, the reverse technique of examining each binary sequence of a specified length underlies the formation of an algorithm to identify and count such isomeric molecules. The algorithm involves the specification of a set of three binary operations—which correspond to allowable physical transformations. This may be expressed in the form of a formal algebraic table of operations. Application of this algorithm, with Patterson's drawing convention, produces a canonical representation for each such isomer.

Despite the presence of a certain amount of noncoplanarity in the cyclopentane molecule (which is caused by the relieving of the strain that would result if the hydrogen atoms were

allowed to remain eclipsed),¹ the simplified geometrical skeleton model formed by the successive "straight line"² concatenation of regular pentagonal modules gives a fairly good

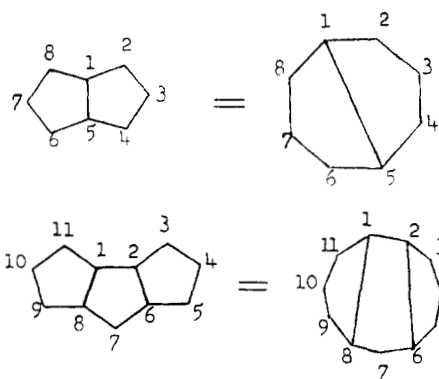


Figure 1.

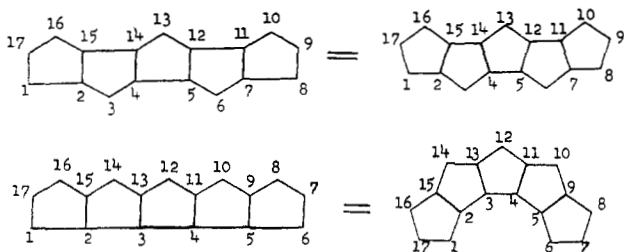


Figure 2.

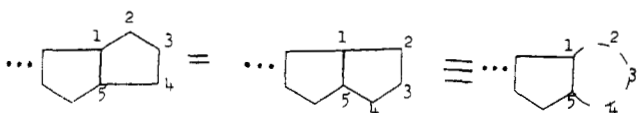


Figure 3.

description of an interesting class of fused pentagonal rings. The first two members of this set (Figure 1) are

bicyclo[3.3.0]octane

tricyclo[6.3.0.0^{2,6}]undecane

The description of this class of molecules is enhanced through the use of the drawing convention prescribed by Patterson³ for the pentagonal module—inasmuch as such a convention assigns a unique orientation that is readily machine codable. In fact, with this convention, the idea of “straight” is fulfilled by an alternation of pentagons in the “up” vs. “down” position, whereas the intuitive alignment (represented by a straight line in the drawing along the top or bottom of a connected set of pentagons) produces a molecule that is, in actuality, circular (Figure 2).

The representation of polyfused pentagonal rings using Patterson's format yields a sequence of rings pointing up or down—which may be machine coded as “0” or “1”, respectively. The algorithm to manipulate each string of 0's and 1's is based on the following three principles: (A) The two end digits of each binary number are immaterial; i.e., 0's and 1's may be interchanged at will. From a chemistry perspective, this is equivalent to saying that there is a plasticity to each end ring (Figure 3); i.e., an end ring may be represented in either format without changing the incidence relations (vertices 1 and 5 have incidence = 3, while vertices 2–4 have incidence = 2). (B) Each binary number is equal to its 1's complement. From a chemistry perspective, this is equivalent to a horizontal reflection (Figure 4). (C) Each binary number is equal to its reverse-order number. From a chemistry perspective, this is equivalent to a vertical reflection (Figure 5).

From the vantage point of a chemist, all three operations must be examined inasmuch as this is the standard method for drawing such molecules. However, from the viewpoint of an information scientist, Ryon⁴ has demonstrated that the final form of the algorithm may be simplified by noting that an arc

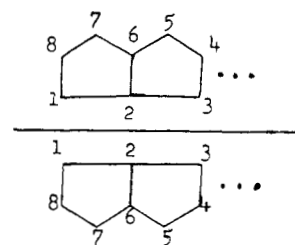


Figure 4.

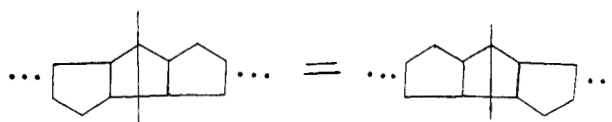


Figure 5.

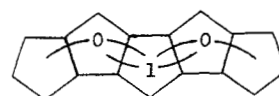


Figure 6.

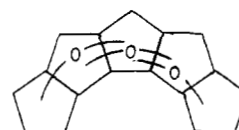


Figure 7.

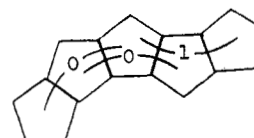


Figure 8.

through each sequence of three rings will point in the same direction as the center ring. In other words, for the five-ring sequence shown in Figure 2, instead of having to examine $2^5 = 32$ numbers and then ignoring the two end rings (i.e., the two end digits), by considering the direction of bend of the arc through each sequence of three rings there is the need to examine only $2^3 = 8$ binary numbers—rather than 32—in order to find the set of all of the different pentacycloheptadecane isomers. Even with this set of eight numbers, however, there is still multiple redundancy; namely, there are only three distinct isomers (Figures 6–8). In other words, the final form of the algorithm will need to consider the set of $r - 2$ rings (without the end rings) and the use of operations B and C only. Furthermore, we may note that the application of first B and then C is equal to applying first C and then B; also, both are geometrically equivalent to rotating the configuration through 180° . This latter transformation may be represented by the combined operation BC. Also, applying either B or C twice gives the original configuration. This is expressed algebraically as $BB = CC = I$, where I is the identity operation ($IX = XI = X$, for every X).

Since geometrically there are three operations (two reflections and one rotation), there exists a corresponding algebraic table of operations—containing four rows (and columns), one for each operation plus an additional one for the identity operation—which describes the manipulations that may be performed on each sequence of the $r - 2$ digits:⁴

	I	B	C	BC
I	I	B	C	BC
B	B	I	BC	C
C	C	BC	I	B
BC	BC	C	B	I

Table I^a

			comment
tricyclo	0	1 \xrightarrow{B} 0	exactly one isomer for three rings—see Figure 9
tetracyclo	00 = 0 01 = 1	10 \xrightarrow{B} 01 = 1 11 \xrightarrow{B} 00 = 0	exactly two isomers (0 and 1) for four rings—see Figure 10
pentacyclo	000 = 0 001 = 1	010 = 2 011 \xrightarrow{BC} 001 = 1 100 through 111 \xrightarrow{B} 3 through 0, respectively	exactly three isomers (0, 1, and 2) for five rings—see Figure 11
hexacyclo	0000 = 0 0001 = 1 0010 = 2 0011 = 3	0100 \xrightarrow{C} 0010 = 2 0101 = 5 0110 = 6 0111 \xrightarrow{BC} 0001 = 1 1000 through 1111 \xrightarrow{B} 7 through 0, respectively	exactly six isomers (0, 1, 2, 3, 5, and 6) for six rings—see Figure 12
heptacyclo	canonical C transformed BC transformed	0, 1, 2, 3, 4, 5, 6, 9, 10, and 14 8 (=2) and 12 (=6) 7 (=3), 11 (=5), 13 (=9), and 15 (=1)	exactly ten isomers for seven rings

^aNote that the second half of each set of numbers is always equivalent to the first half in the reverse order—by operation B —and thus may be omitted, as it cannot describe a new isomer.

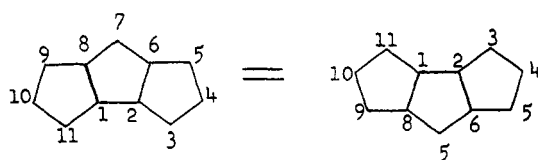
tricyclo[6.3.0.0^{2,6}]undecane

Figure 9.

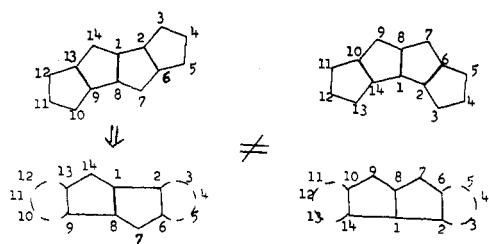
tetracyclo[6.6.0.0^{2,6}.0^{9,13}]tetradecane tetracyclo[6.6.0.0^{2,6}.0^{10,14}]tetradecane

Figure 10.

In order to describe each of the different ways for forming isomers having r rings, consider the set of binary numbers from 0 to $2^r - 1$. Let n designate the given binary number without the end digits, $B(n)$ be the binary number formed by taking the 1's complement of n , $C(n)$ be the binary number formed by writing n in the reverse order, and $BC(n)$ be the binary number formed by taking the 1's complement of the digits of n in the reverse order. The smallest of the numbers $[n, B(n), C(n), \text{ and } BC(n)]$ is the canonical representation of each of the different isomers.

In practice, however, if we are cataloging the entire set of r rings, rather than just examining a single arbitrary configuration, the process may be greatly simplified. The formal algorithm is as follows: Apply the three operations $B(n)$, $C(n)$, and $BC(n)$ to each of the binary numbers formed starting with a string of $r - 2$ 0's and progressing through $r - 2$ 1's (see Table I),⁵ as follows:

(1) If the first digit of n is a "0", proceed to step 2; otherwise, apply $B(n)$ to get a smaller equivalent binary representation (i.e., set this new number equal to n) and then proceed to step 2.

(2) If the final digit is a "0", apply $C(n)$. The canonical representation is $\min [n, C(n)]$. Otherwise, the final digit must

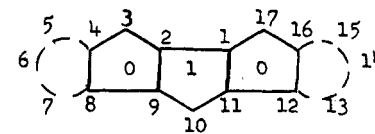
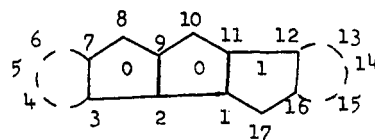
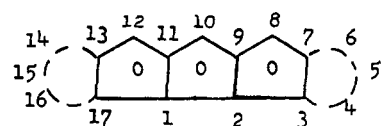


Figure 11.

x0000x = hexacyclo[9.9.0.0^{2,9}.0^{3,7}.0^{13,20}.0^{15,19}]icosane
 x0001x = hexacyclo[9.9.0.0^{2,9}.0^{3,7}.0^{13,20}.0^{14,18}]icosane
 x0010x = hexacyclo[9.9.0.0^{2,9}.0^{3,7}.0^{12,19}.0^{14,18}]icosane
 x0011x = hexacyclo[9.9.0.0^{2,9}.0^{3,7}.0^{12,19}.0^{13,17}]icosane
 x0101x = hexacyclo[9.9.0.0^{2,9}.0^{4,8}.0^{12,19}.0^{14,18}]icosane
 x0110x = hexacyclo[9.9.0.0^{2,9}.0^{4,8}.0^{13,20}.0^{14,18}]icosane

Figure 12.

be "1". Apply $BC(n)$. The canonical representation is $\min [n, BC(n)]$.

Note that applying any of the other operations will give a larger number than the starting one or else an isomer already included in the set.

REFERENCES AND NOTES

- (1) Greenberg, A.; Liebman, J. F. "Strained Organic Molecules"; Academic Press: New York, 1978; p 65.
- (2) Elk, S. B. "Refinement of Systematic Nomenclature for Polybenzenes and Its Extension to Systems of General Arenes". *Math. Chem.* **1982**, *13*, 261.
- (3) Patterson, A. M. "Proposed International Rules for Numbering Organic Ring Systems". *J. Am. Chem. Soc.* **1925**, *47*, 547.
- (4) Ryon, J. W., Professor of Computer Science, New Jersey Institute of Technology, personal communication.
- (5) Figures 9–12 give the pictorial representations and the corresponding IUPAC nomenclature⁶ for each of the different tricyclo through hexacyclo compounds.
- (6) International Union of Pure and Applied Chemistry "Nomenclature of Organic Chemistry: Section A"; Pergamon Press: Oxford, England, 1979.