THE LIMITATIONS OF DESCRIBING CHEMICAL COMPOUNDS BY VIRTUE OF THEIR GRAPH THEORETICAL REPRESENTATION BY A COMPUTER

Seymour B. Elk

Elk Technical Associates
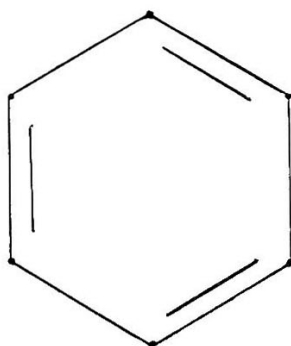
321 Harris Place

New Milford, New Jersey 07646

ABSTRACT: Because the computer is capable of providing a systematic presentation of any given data according to some agreed upon set of pre-selected rules, it is often used in order to describe possible chemical isomers having a given formula. One such computer description focuses attention on the graph theoretical properties of the model being used to represent these compounds. By considering <u>all</u> molecules having the formula $(CH)_6$, a <u>complete</u> listing of <u>all</u> such compounds known to the computer is presented. The inadequacy of ANY such technique is illustrated and expansion of the heuristic set of "reasonable" compounds is shown to occur whenever new compounds are described in either the chemistry laboratory or the literature.
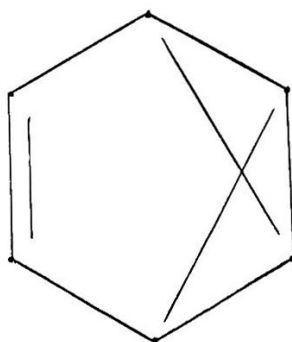
A common problem in mathematical chemistry is the enumeration of isomers having a given formula. Despite the fact that many chemists are willing to concede that the class of all chemical compounds does not have a one-to-one and onto isomorphism with the graph theoretical models of these compounds, there is still the false expectation of the universal applicability of graphs for the enumeration of chemical isomers. This non-isomorphism arises due to the fact that graph theory does not take into consideration differences in topology; attempting instead to give a complete description of all chemical compounds using only "ball and stick" models. When making such an enumeration, there is the implicit assumption that this listing is a complete set. Since the complete listing of all graphs of a given size (V and E) is a matter of simple combinatorics (for example Harary[1] illustrates all 11 graphs possible with V=4; with its built-in implication that E can vary from the completely disconnected graph, E=0, through the complete graph, E=6), it is often erroneously assumed that this is sufficient to guarantee the complete listing of all possible chemical isomers with this formula.

Although this technique is quite useful for delineating the formulas of most compounds that are familiar, as well as quite a few that are chemically unattainable, it contains an inherent bias toward those compounds that are anticipated, with a strong probability that any truly new class of compound will be overlooked by the method. This is seen to be the case in a widely circulated advertisement[2] which states that C_6H_6 is benzene and a whole lot more; namely 217 different compounds. What this particular software manufacturer does not, in fact, CAN NOT give is a complete listing of all possible isomers of the given formula. That such is the logical consequence of any computer algorithm produced to display all compounds having a given formula is shown below.
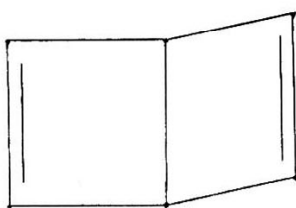
Nevertheless, despite the above stated limitation, to deny the importance of both graph theoretical representations and computer simulation of combinations of 'atoms' into
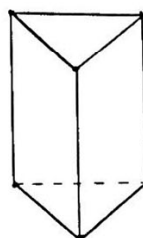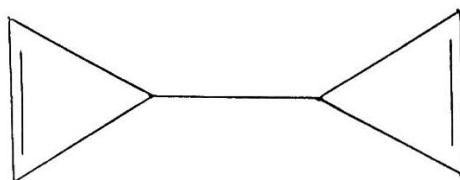
a

b

c

d

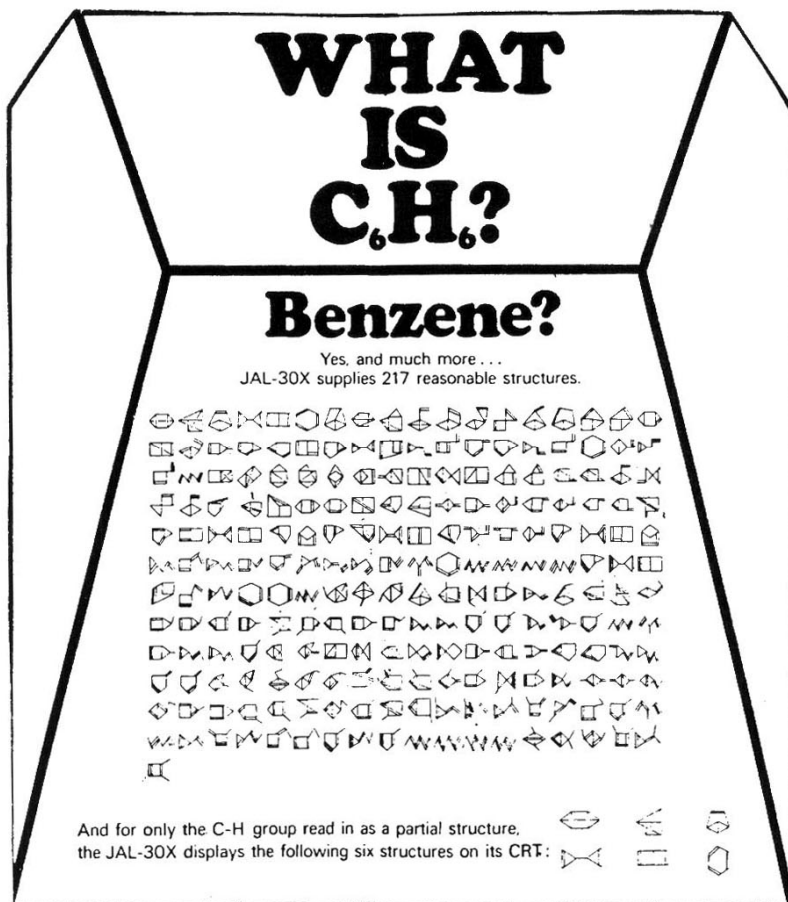e

FIGURE 1

WHAT IS C_6H_6?

Benzene?

Yes, and much more...
JAL-30X supplies 217 reasonable structures.

And for only the C-H group read in as a partial structure,
the JAL-30X displays the following six structures on its CRT:

FIGURE 2

'molecules' would be to deny the obvious. However, over-reliance on either graph theory or the computer can be just as myoptic. For example, it was merely 15 years ago that, in their well-researched descriptive article on $(CH)_n$ compounds, Scott and Jones[3] listed only five such compounds for $n = 6$. (Figure 1).

By contrast, of the 217 structures for C_6H_6 compounds listed as "reasonable" in Reference 2 (Figure 2), six are of the form $(CH)_6$. The sixth such compound (Figure 3), which may be dubbed 'Moebiane' is the logical simplification of the compound created by Walba[4] in 1982. This is equivalent to saying that our concept of "reasonable" is predicated on what is happening in the laboratory or what is of current speculation in the chemical literature, rather than what is mathematically possible. This is notwithstanding the fact that the listing in Reference 2 was formulated in order to include all possible compounds -- without regard to chemical viability. In fact, upon examining the order in which
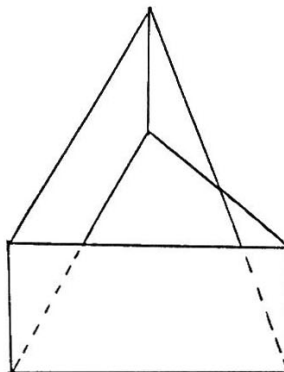

FIGURE 3

these 217 compounds are listed, it seems clear that the only formal limitation built into the system is that the coordination be equal to 4. It should be noted, however, that such a technique makes certain heuristic assumptions -- of which orientability[5] was implicit for Scott and Jones, but was relaxed for the listing in Reference 2.

Now, however, even with the inclusion of selected 'exotic' compounds, such as Figure 3, the listing in Reference 2 is seen to be incomplete; namely, Figure 4, which is a large picture of the 'compound' presented in the eighth row, next to last column in Figure 2 corresponds not to a single structure, but rather to a pair of geometrical isomers (Figures 5 and 6).
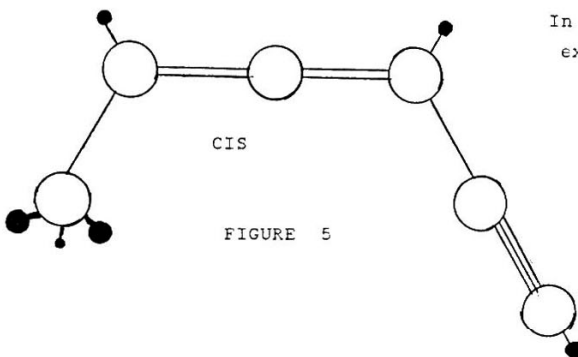
FIGURE   4

CIS

FIGURE   5

In other words,
examination of
'chemical'
structure
by consid-
eration of
only
constitution
using an
ordered
set of
incidence

relationships in a graph can not be counted on to be complete.

In   a
similar way,
not only is
the general
class   of
C_6H_6
compounds
inscrutable
using only
applications
of   graph
theory and
the
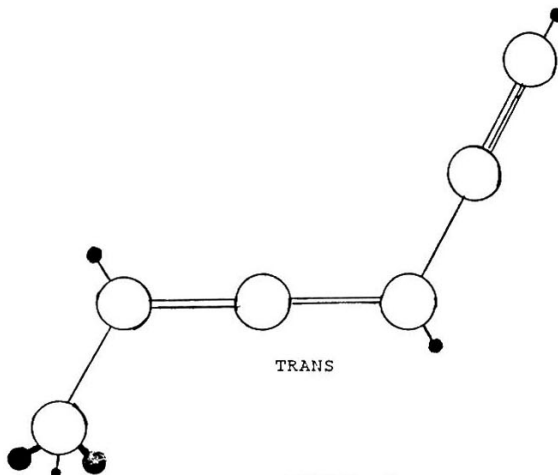computer,
even   the
more limited
class of

TRANS

FIGURE   6

(CH)$_6$ compounds is incomplete. For example, additional (CH)$_6$
compounds can be described if one is willing to relax another
requirement of 'simple' topology[6] -- whatever that means. In
fact, determination of exactly what is a 'simple' figure has
eluded mathematicians for millenia. Every proposed line of
delineation has both included point sets not desired
intuitively while simultaneously omitting others that
heuristically one might wish to include.

One such omission of the listing of possible (CH)$_6$
compounds is the union of two smaller compounds that are
topologically linked together[7] -- either with or without a
graph theoretical connection. For
example, the knot (Figure 7) formed by
interdigiting the two 'triangles' of
Figure 1e is described by the identical
graph as is the topologically simpler
compound that is 'unknotted'.
Additionally, using ball and stick
models with their admitted
oversimplification of chemical
molecules, it is possible to produce
rotaxanes[8] by constraining the free

FIGURE 7

movement of an ethyne molecule by 'threading' it through a
cyclobutadiene structure (Figure 8) and through 'tetrahedrane'
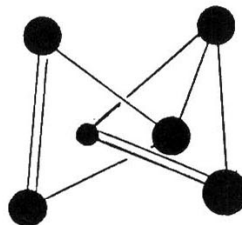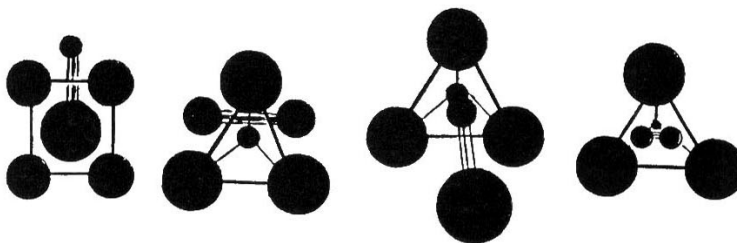

FIGURE 8   FIGURE 9   FIGURE 10   FIGURE 11

in three different ways (Figures 9, 10 and 11). These are
but a few of the further examples of the need for an intuitive
appreciation of the geometry that is not presently included in

computer simulations. The possibility of forming different rotaxanes, which differ only in how the 'interior part' is 'caged' effectively eliminates any graph theoretical study from guaranteeing that the entire field will be covered by this technique. Admittedly, the ability to constrain an ethyne molecule in precisely the orientation required to keep it so that either 0, 1 or 2 of the carbon atoms are inside the tetrahedral 'cage' is highly speculative. What is not speculative is that all catenanes presently known are very much larger than cyclopropane; however, the same could be said of the 'compound' pictured in Figure 3.

In a similar manner, let us allow the metric distortion of angles; i.e., consider an 'allenic' structure having a 60°, instead of the traditional 180°, angle. This now allows for formation of a catenane having C_3H_0 (i.e., a geometrically possible, but energetically VERY unlikely, allotropic form of carbon in the form of a three membered ring whose carbon atoms are joined by double bonds) threaded through a cyclopropane ring (Figure 12).
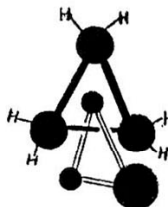

FIGURE 12

From the above, we can properly conclude that we are at the mercy of our imagination -- and the better our imagination the more likely that we may discover a 'new' class of compounds that had not been anticipated. This is NOT to imply that the computer can not simulate compounds once they have been described, but rather to illustrate that the computer can only be programmed to search for what is considered (by the programmer) to be a reasonable structure -- where "reasonable" is a function of what compounds are presently known. For example, suppose that penta- and hexa-coordinated compounds[9] became common, then even the above described extensions of such a listing would have to be enlarged to include still more 'unimaginable' compounds. Actually, from a mathematical bias, the description must go beyond focusing on chemistry's traditional limitation to single, double and triple bonds;

i.e., quadruple bonds of the type found in metal clusters[10],
as well as three-center bonds (of which variations of the
boron hydrides[11] provide a prime example) and fractional bonds
(of which benzene is more accurately represented by Robinson's
picture of a circle inside a hexagon in order to emphasize the
fact that there are six equal bonds of magnitude one and one-
half, rather than **any** single or double bonds[12]) should also be
included, etc. In fact, with an active imagination, there is
no limit to the multiplicity of the bonds, including the
limiting case of an infinite number of bonds, each with
variable bond lengths.

The answer to this dilemma, is, as it should be, to
recognize that, even with a computer, a complete listing is an
impossibility. It is unrealistic to believe that any
extension of any idea will be able to include **everything** that
is mathematically possible. Regardless how well devised, a
system can only be as good as the heuristics that went into
the model being used.

REFERENCES

(1) Harary, F., "Graph Theory"; Addison-Wesley Publishing
     Company: Reading, Mass., 1969, p.9

(2) Advertisement entitled "What Is C_6H_6? Benzene?" in
     several chemistry journals for the software package JAL-
     30X produced by JEOL Ltd. of Tokyo, Japan.

(3) Scott, L. T.; Jones, M., "Rearrangements and Interconver-
     sions of Compounds of the Formula $(CH)_n$". Chem.Rev.,
     1972, 72, 183-202.

(4) Walba, D. M.; Richards, R.M.; Haltiwanger, R.C., "Total
     Synthesis of the First Molecular Moebius Strip,"
     J.Am.Chem.Soc., 1982, 104, 3219-21.

(5) Coxeter, H. M. S., "Introduction to Geometry," John
     Wiley & Sons, Inc.: New York, 1961, p. 382.
     A surface is called "orientable" if the faces of any map
     on the surface can be regarded as directed polygons in
     such a way that the two directions thus assigned to each
     edge cancel each other. If, on the other hand,
     there exists at least one map which can not have such an
     assignment, this surface is "non-orientable". An example
     of a non-orientable surface is the Moebius strip.

(6) Elk, S. B., "Topologically Different Models To Be Used
     as the Basis for Ring Compound Taxonomy," J.Chem.Inf.
     Comput.Sci. 1985 25, 17-22.

(7) Frisch, H. L.; Wasserman, E. "Chemical Topology," J.Am.
     Chem.Soc. 1961 83, 3789-3799.

(8) Schill, G. Catenanes, Rotaxanes and Knots; Academic
     Press: New York, 1971.

(9) Greenberg, A.; Liebman, J. F. Strained Organic Molecules;
     Academic Press: New York, 1978; pp. 375-385.

(10) Cotton, F. A., "Quadruple Bonds and Other Multiple
     Metal-to-Metal Bonds," Chem.Soc.Rev. 1975 4, 27-53.

(11) Cotton, F.A.; Wilkinson, G. Advanced Inorganic Chemistry;
     Interscience Publishers: New York, 2nd Ed., 1966; pp 275-
     281.

(12) Clar, E.; The Aromatic Sextet; William Clowes and Sons,
     Ltd.: London, 1972; p.12.