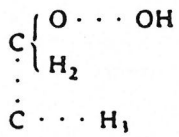
ON THE BASIC PROBLEM OF PÓLYA'S THEORY OF ENUMERATION

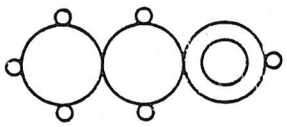
by Adalbert Kerber *

1. Historic origin

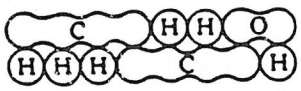
Around 1860 the isomerism problem arose: chemical substances with the same empirical formula but different properties were discovered. At about the same time different ways of drawing molecules were introduced,e.g.



Couper



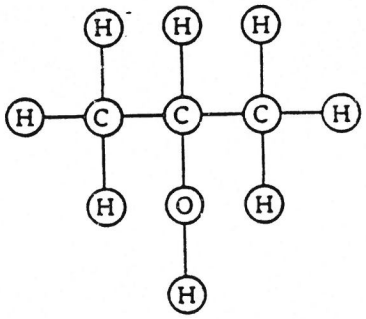
Loschmidt

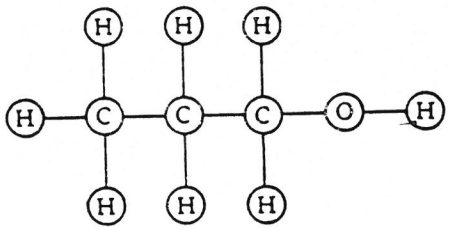


Kekulé

ethyl alcohol, C_2H_5OH

The method used by Alexander Crum Brown (Edinburgh) showed clearly that the reason for isomerism is that sometimes there are several ways to connect atoms subject to their given valency and given substructures. For example in the case of the empirical formula C_3H_7OH, which means that there are 3 atoms with valency 4, 8 atoms with valency 1 and one atom with valency 2 ,and furthermore that there must be a substructure of the form $-O-H$, he got the following two solutions:





* *Supported by the Volkswagenwerkstiftung*

(This substance is an alcohol and it seems to be this example which, according to Prof. Gillis, stimulated G. Pólya to start his lecture for the International Congress in 1936 with the sentence:"*Es gibt viele Alkohole, nicht alle sind trinkbar* ...".)

At about the same time Arthur Cayley examined *trees* and J.J. Sylvester introduced the name *graph* in a paper entitled "Chemistry and Algebra".

The resulting problem which stimulated Pólya to develop what now is called *his* theory of enumeration, and which is in fact a very interesting particular case of the theory of enumeration under finite group action reads therefore as follows: *Enumerate, classify and construct graphs with certain conditions (like degree sequence, certain subgraphs etc.).*

Reference (from which the drawings are taken):

N.L. Biggs/E.K. LLoyd/R.J. Wilson: Graph Theory 1739-1936. Clarendon Press 1976.

2. The DENDRAL Project

This project seems to be the only strong attack towards a solution of the above problem. It was started by Lederberg, Stanford, in 1964 with an algorithm for the generation of acyclic molecular graphs. The central part of system is CONGEN, a system for the CONstrained GENeration of molecular graphs. It generates graphs with side conditions like connectednes, certain radicals etc.. The reader should carefully notice that the aim is to do this for molecules with up to 100 points,and know that at present complete lists of graphs are available only up to 10 points.

DENDRAL consists of two main parts:

Heuristic DENDRAL which judges by specific chemical knowledge if a just constructed graph should be added to the set of *possible solutions.*

Meta DENDRAL which contains rule discovering methods in order to formulate hypotheses about spectroscopic properties of structures the molecular graph of which is already known

Our main interest lies in CONGEN, the basic method of which is:

Take an empirical formula, evaluate the number of cyclic substructures and distribute the atoms in all the possible and essentially different ways over the cyclic substructures and the acyclic rest.

Besides this mathematical part it is of main importance to have a good program for the *interpretation of mass spectra and other chemical information.* First steps are already done which evaluate from mass spectra the molecular weight of some fractions of the molecule in order to get possible sequences of subgraphs. This is already and successfully applied to *oligopeptides* which are sequences of amino acids.

3. The PLAN-GENERATE-TEST method

This is the basic "philosophy"of this project, and in mathematical terms it means that the conditions for the desired graphs have to be put in as early as possibly in order, i.e. to *plan* in order to keep the complexity as low as possible first of all, and that besides this there must be a method available which *tests* the generated graphs against the chemical information.

Chemistry	Mathematics
yields conditions:	
i) empirical formula C_3H_8O	≈ edge degree sequence 444211111111
ii) ...+ radical C_3H_7OH	≈ subgraph
or	
iii) mass spectrum	≈ weight of points in a connected subgraph
iv) substance is a polypeptide	≈ sequence of certain possible subgraphs (e.g. of 22 amino acids)
v) results of chemical spectroscopy	≈ amino acids can and will break at certain points
⋮	

Reference:
R.K. Lindsay/B.G. Buchanan/E.A. Feigenbaum/J. Lederberg: Applications of Artificial Intelligence for Organic Chemistry, the DENDRAL Project.
McGraw-Hill Book Company

4. Methods

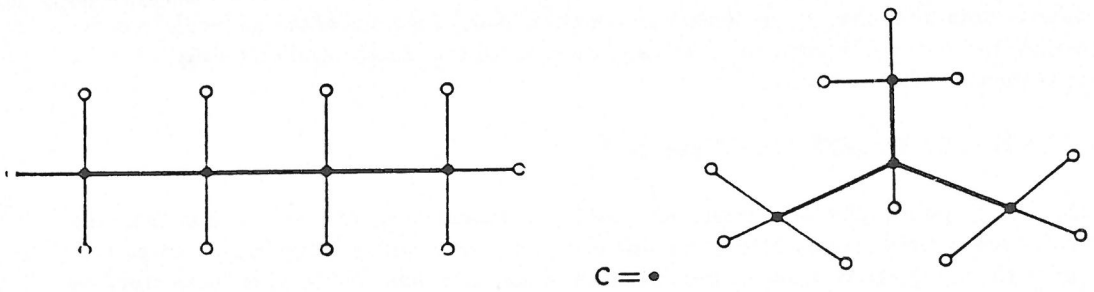
Some of the methods which we use at present are

i) Direct methods for the construction of graphs using double cosets and the lexicographic order of permutations, see e.g.

H. Brown/L. Hjelmeland/L. Masinter: Constructive graph labeling using double cosets. Discrete Math. 7(1974),1-30.

H. Brown/L. Masinter: An algorithm for the construction of the graphs of organic molecules. Discrete Math. 8(1974),227-244.

ii) Recursive procedures, e.g. the reduction of a chemical graph to its *skeleton*



iii) Generation of graphs uniformly at random (which is a very interesting method in order

to look at counterexamples, say, and which can be used in all the Pólya type cases and, more generally, in each case of enumeration under finite group action, too), see

J.D. Dixon/H.S. Wilf: The random selection of unlabeled graphs. J. Algorithms 4(1983),205-213.

It is our aim to put these things together in order to get, for mathematical purposes, a program system which redundancy-free allows to construct graphs with certain side conditions, and which, for chemical purposes, yields an expert system for molecular structure elucidation which is portable and uses advanced combinatorial and algebraic methods.

A. Kerber
Lehrstuhl II für Mathematik
Universität Bayreuth
D-8580 Bayreuth