

Report 78-15
Stanford -- KSL

Scientific DataLink

Application of the Permutation Group to
Stereoisomer Generation for Computer-
Assisted Structure Elucidation.
James G. Nourse, Jul 1978

card 1 of 1

Application of the Permutation Group to Stereoisomer Generation for
Computer Assisted Structure Elucidation

James G. Nourse
Computer Science Dept.
Stanford University
Stanford Calif. 94305 USA

The purpose of this paper is to describe a recent effort to generate, classify, and enumerate the stereoisomers of all organic chemical structures consistent with a given empirical formula. The resulting theory is then converted into an algorithm which is programmed for a computer. The resulting program is used as an aid for computer assisted structure elucidation [1]. The proper group theoretical formulation is the key to the solution of the problem and will be discussed here. Simply stated, the problem is to generate, classify, and enumerate the possible stereoisomers for an organic chemical structure of defined constitution. (Constitution is the specification of the atoms in the molecule and their connectivity by bonds).

In order to formulate the desired symmetry group it is necessary to describe the constitution of a chemical structure with a graph in which the atoms correspond to the nodes and the bonds correspond to the edges. Multiple bonds correspond to multiple edges so the desired graph is actually a multigraph. All the nodes and edges are labelled uniquely (e.g. numbered consecutively). For the purposes here the symmetry of this graph will be described by a permutation group which includes all the node symmetries plus operations which interchange double edges (but not triple edges). This group is a semidirect product of the node symmetry group and the group which interchanges double edges and is called the Graph Symmetry Group.

The desired symmetry group is called the Configuration Symmetry Group (CSG) and is defined to be the Graph Symmetry Group represented by its action on the configurations of all trivalent and tetravalent atoms in the chemical structure. The configurations at these atoms are determined by the numbers of the substituents at each atom. Thus two enantiomeric configurations can be defined for any tetravalent atom or nonplanar trivalent atom by simply giving all the substituents different numbers, even for atoms which do not exist in enantiomeric forms.

The representation of the Graph Symmetry Group on the configurations of the atoms is constructed by determining the effect of the graph permutations on these configurations. This is done by defining an ordering of the four substituents on all the atoms. This ordering is most conveniently the ordering induced by the numbering of the atoms. If a graph permutation fixes an atom it will either leave the configuration of that atom unchanged (an even permutation of the substituents) or it will invert its configuration (an odd permutation of the substituents). This information is added to the permutation. If a permutation takes one atom to another then a correspondence of the substituents must be defined. This correspondence is determined by the ordering of the substituents. Thus if atom A goes to atom B, then there is a mapping of the four substituents on atom A to the four substituents on atom B.

$$(a_1, a_2, a_3, a_4) \text{ ----> } (b_1, b_2, b_3, b_4)$$

If the permutation of the four numbered substituents defined by this mapping is even, then the configuration of A remains unchanged after it is mapped to B. If this permutation is odd, then the configuration of A inverts as it is mapped to B. This procedure is performed for all the permutations in the graph symmetry group and for all the atoms with three (nonplanar) or four substituents. The resulting group is called the Configuration Symmetry Group. Intuitively, this group is an invariance group for a stereoisomer in which the relative configurations of constitutionally identically substituted atoms is taken into account. See ref. [2] for a pictorial example.

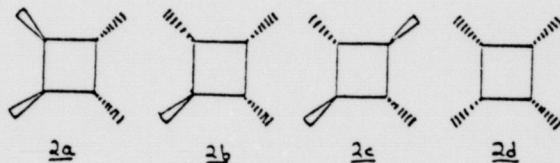
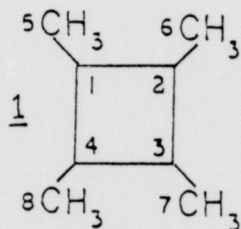
The resultant operations in the CSG are actually permutation inversions where the inversions are of individual atom configurations. The group of all such permutation inversions is the wreath product $S_n[S_2]$ for n atoms [3]. The permutation inversions in the CSG act on the 2^n possible stereoisomers to yield equivalence classes which correspond to the symmetrically distinct stereoisomers. Each possible stereoisomer can be represented by an n-tuple of the configurations of the atoms. Thus for example consider tetramethylcyclobutane 1, which has four carbon atoms which are capable of existing in two distinct configurations. For simplicity the methyl carbons are not considered.

There are 2^4 or 16 potential stereoisomers of this structure. Each of these stereoisomers can be represented as a 4-tuple such as [+--+] which means that atom 1 is in the "+" configuration (based on the atom numbering), atom 2 is in the "-" configuration etc. These 4-tuples correspond to chiral graphs [4] which are graphs augmented with "+" or "-" parity labels at the trivalent and tetravalent nodes. The configuration symmetry group for this structure is derived as described above and is given in the table. The action of these permutation inversions on the 4-tuples is (chiral graphs) as follows: A permutation inversion such as (12'34') is read: 1 goes to 2 and inverts configuration, 2 goes to 3, 3 goes to 4 and inverts configuration, and 4 goes to 1. This acts on the 4-tuple [-+++] as:

(12'34') [-+++] -----> [+++-]

Doing this for all the permutation inversions in the CSG on all the 4-tuples has the effect of collecting the 16 4-tuples into 4 equivalence classes which correspond to the 4 possible stereoisomers of tetramethylcyclobutane. These are shown in the table and the structures.

	<u>2a</u>	<u>2b</u>	<u>2c</u>	<u>2d</u>
(1)(2)(3)(4)	[++++]	[-+++]	[----]	[-+--]
(1')(24)(3')	[-+++]	[++++]	[+---]	[+---]
(12)(34)	[++++]	[++++]	[----]	[+---]
(12'34')	[+--+]	[++++]	[+---]	[-+--]
(13)(2')(4')	[+--+]	[+---]	[+---]	[+---]
(1'3')(2'4')	[----]	[----]	[----]	[-+--]
(1'43'2)	[-+++]	[-+--]	[+---]	[-+--]
(1'4')(2'3')	[----]	[----]	[----]	[+---]



This method of forming equivalence classes of the possible

stereoisomers can be used in an algorithm which generates all the distinct stereoisomers of a chemical structure of defined constitution. This algorithm has been implemented as a computer program used for structure elucidation [5].

The specification of the equivalence class of the CSG to which a given stereoisomer belongs provides a means of classification of stereoisomers. When combined with some means of uniquely (canonically) numbering the atoms of the chemical structure, a unique (canonical) name for the structure results. Each stereoisomer is represented by the numbered graph describing its constitution with parity labels at each nonplanar trivalent and tetravalent atom. For each equivalence class of the CSG a representative member is chosen which is in some sense unique. This can be the "lowest" member of the equivalence class where "+" is considered to be lower than "-" and the atoms are ordered by their node numbers [6]. This specification of the configuration of a stereoisomer is independent of geometry since only a graph (usually represented tabularly) augmented with parity labels (the chiral graph) is required to uniquely represent the stereoisomer. While these parity labels ultimately refer to geometric configurations, there is no use made of any geometric property when these parity labels are determined. Only the connectivity of the numbered graph and the parity of permutations is required. Thus the configuration (by this specification) of a stereoisomer can be separated from any geometric property. This may lead to a separation of the notions of configuration (nongeometric) and conformation (geometric) at least for typical organic structures.

This classification of stereoisomers simultaneously considers global and local properties of the structure which facilitates providing a unique name. The local property is the configuration of the atoms based on the numbering of the adjacent substituents. The global property is the overall symmetry described by the configuration symmetry group. The final specification gives just the configuration designations which were computed by using the CSG. The CSG need not be part of the designation. A specification of the configuration of a stereoisomer which relies only on local properties such as the current R/S naming system becomes very difficult when the structure has a great deal of overall symmetry [7]. A specification which relies only on a global property such as the symmetry group of the stereoisomer would not provide a unique specification since most chemical structures have little or no overall symmetry.

Using the Configuration Symmetry Group, a single counting formula can be derived which gives the number of distinct stereoisomers for a structure of defined constitution. This problem dates back to van't Hoff and no general solution has been heretofore available. The desired counting formula is one which requires only the symmetry group of the structure in analogy with the Polya enumeration theorem [8].

Combinatorial problems of this kind are generally formulated as mapping problems in which the number of equivalence classes of mappings under the action of some symmetry group is determined.

In this case the tri- and tetravalent atoms are mapped to the two possible configuration "+" or "-" based on the atom numberings. The problem is now to count the number of equivalence classes of these mappings formed by the action of the CSG. It has already been pointed out that the CSG is a subgroup of the wreath product group $S_n[S_2]$. The permutation representation of this wreath product is an example of an exponentiation group as described by Kerber [9]. For this problem, the exponentiation group is the wreath product of the group of all identical atom permutations (S_n in this case for n identical atoms) around the group which interchanges the two configurations (S_2 in this case). The exponentiation group therefore includes all the possible domain and range symmetries possible for the above mapping of atoms to configurations. The CSG is a subgroup of this exponentiation group. Kerber has given the general counting formula for the exponentiation group [9] so that counting the equivalence classes of the CSG and hence the possible stereoisomers requires the restriction of this formula to the CSG subgroup.

The counting formula will be illustrated with the tetramethylcyclobutane example, 2a-d. The computation is summarized by:

	(1)(2)(3)(4)	(12'34')	(1'3')(2'4')	(1'4')(2'3')	(1')(24)(3')
		(1'43'2)		(12)(34)	(13)(2')(4')
order	1	2	1	2	1
orbits	4	1	2	2	3
orbits with					
odd # inv.	0	0	0	0	2
contribution	16	4	4	8	0

It can be easily proved that a permutation inversion with an odd number of inversions in an orbit will contribute zero to the counting total. By Burnside's lemma the contribution of a group element to the counting total will be equal to the number of objects fixed by that group element. In this cases these objects are the 2^n possible stereoisomers. To show that an orbit with an odd number of inversions cannot fix an n-tuple, start at the beginning of the orbit. The permutation takes this first atom to another atom. If an inversion must be done here, the two atoms must have opposite configurations if this n-tuple is to be fixed by the permutation inversion. This second atom is taken to a third atom by the permutation. If an inversion is to be done here, then the second and third atoms must have opposite configurations and therefore the first and third must have the same configuration. This continues until the first atom is reached again. If there are an odd number of inversions in the orbit, the only way an n-tuple can be fixed is if the first atom has the opposite configuration of itself which is impossible. Hence, orbits with an odd number of inversions cannot fix any n-tuple.

This combinatorial result does not depend on the particular choice of numbering of the atoms in the chemical structure even though the exact form of the permutation inversions in the Configuration Symmetry Group do depend on this numbering. Renumbering the atoms corresponds to the operation of conjugation by an element of the group $S_n[S_2]$. The permutation which effects the numbering change is an element of this group. A renumbering will change the CSG into a conjugate CSG within the group $S_n[S_2]$. The key properties of the elements of this group for the counting formula are the orbit structures of the permutation inversions and the number of orbits with an odd number of inversions. Both of these properties are invariant to conjugation. Thus if a permutation inversion includes an orbit of length 4 with 3 inversions, it will go to an orbit of length 4 with either 1 or 3 inversions by the operation of conjugation (renumbering the atoms) and hence will make the same contribution to the counting total.

A limitation of this method of specifying and counting stereoisomers is that only the connectivity of the structure with parity designations is given and no account is taken of conformation (e.g. rotation around single bonds) or topology in the sense of a catenane (interlocked rings). Stated more exactly, structures which can be

interconverted by rotations around single bonds or can be interconverted by bonds passing through bonds, are given the same configuration specification. The former case includes substituted biphenyls and the later case includes catenanes (interlocked rings).

Financial support from the National Institutes of Health (grant 2R24 RR 00612-08) is gratefully acknowledged.

References

- [1] (a) L. M. Masinter, N. S. Sridharan, J. Lederberg, and D. H. Smith, *J. Amer. Chem. Soc.*, 96, 7702 (1974).
(b) R. E. Carhart, D. H. Smith, H. Brown, and C. Djerassi, *J. Amer. Chem. Soc.*, 97, 5755 (1975).
- [2] J. G. Nourse, *J. Amer. Chem. Soc.*, 97, 4594 (1975).
- [3] C. A. Mead, *Top. Curr. Chem.*, 49, 1 (1974).
- [4] R. W. Robinson, F. Harary, and A. T. Balaban, *Tetrahedron*, 32, 355 (1976).
- [5] J. G. Nourse, R. E. Carhart, D. H. Smith, and C. Djerassi, *J. Amer. Chem. Soc.*, submitted.
- [6] W. T. Wipke and T. M. Dyott, *J. Amer. Chem. Soc.*, 96, 4834 (1974).
- [7] R. S. Cahn, C. K. Ingold, and V. Prelog, *Angew. Chem. Int. Ed. Engl.*, 5, 385 (1966). See footnote on p. 397. We are merely pointing out a difficulty in designing a system of nomenclature, not proposing a new system.
- [8] (a) G. Polya, *Acta Math.*, 68, 145 (1937).
(b) B. A. Kennedy, D. A. McQuarrie, and C. H. Brubaker, Jr., *Inorg. Chem.*, 3, 265 (1964).
- [9] (a) A. Kerber, "Representations of Permutation Groups. II", Springer-Verlag, New York, 1975.
(b) A. Kerber, *Discrete Math.*, 13, 13 (1975).

Copyright © 1985 by KSL and
Comtex Scientific Corporation

FILMED FROM BEST AVAILABLE COPY