Tetrahedron. Vol. 30, pp. 1845 to 1859. Pergamon Press 1974. Printed in Great Britain

REPRESENTATION OF THE CONSTITUTIONAL AND STEREOCHEMICAL FEATURES OF CHEMICAL SYSTEMS IN THE COMPUTER ASSISTED DESIGN OF SYNTHESES[1]

J. BLAIR, J. GASTEIGER, C. GILLESPIE, P. D. GILLESPIE and I. UGI
Laboratorium für Organische Chemie der Technischen Universität München

(*Received* 12 *August* 1973)

1. INTRODUCTION

1.1. *Synthetic planning and the involvement of computers*

Synthesis planning involves decisions between many alternatives and is carried out on the basis of large amounts of information. This suggests computer participation in the solution of the problem.

Previous applications of computers in chemistry have been primarily limited to numerical calculations and information retrieval. However, techniques and methodology of artificial intelligence[3] in general and heuristic programming in particular have now matured to the point that the problem of organic chemical synthesis planning is, in principle, soluble.

Several groups, working independently, have attacked the problem and, in some cases, have excellent programs operating which yield partially planned syntheses.[4-8] The latter have been noteworthy in a number of respects, making use of historically developed empirical chemical information. These empirical approaches to computer assisted synthesis planning have two limitations: First, the chemical information usually verbal or graphic in nature, must be transformed into a representation with which the machine can operate; secondly, the results are often limited to those presently available to a well-informed chemist. While the first of these drawbacks relates to effectiveness and becomes less important with the development of sophisticated hardware and programming, the second is an inherent property of any information retrieval system.

1.2. *Algebraic representation of chemical systems and their interconversions by BE- and R-matrices.*

As an alternative to the empirical approaches employed by the Harvard,[4] Princeton,[5] or Stony Brook[6] groups, we saw in the algebraic structures representing chemical systems[9,10] a straightforward route into a computer assisted synthesis program which bypasses some of the existing limitations. It promises, due to its open-ended nature, to yield new information concerning relations between chemical systems as well as a test of the validity of the underlying mathematical model of chemistry.[9,10]

Any synthesis involves the conversion of a set of starting materials into the target compound and its by-products by a sequence of synthetic reactions.

An equation representing a chemical reaction has as its left and right hand sides isomeric ensembles of molecules, EM. A multistep synthesis is representable as a sequence of isomeric ensembles of molecules which begins with the set of starting materials EM_A and ends with the target molecule and all by-products which are formed, EM_Z. A synthesis is embedded in that family of all isomeric ensembles of molecules (FIEM) containing the atoms of the target and its by-products.

In our context a BE-matrix[9] is a particularly convenient and effective representation of an EM. The chemical constitution of an EM is given by its covalent bonds and its free valence electrons. Thus, a BE-matrix E representing an ensemble of molecules consisting of n atoms, is an $n \times n$ matrix with integral entries where the off-diagonal entries e_{ij} represent the formal covalent bond order between the atoms A_i and A_j and the diagonal entries e_{ii} correspond to the numbers of free electrons on the atom A_i.[9c,d]* Since a bond from A_i to A_j implies a bond from A_j to A_i, the BE-matrices are symmetric. The sum over the entries in the BE-matrix is equal to the total number of valence electrons in the ensemble of molecules. The row/column sums of BE-matrices $e_i = \sum_j e_{ij} = \sum_j e_{ji}$ are the numbers of valence electrons belonging to the atom A_i, and the cross sums $c_i = 2e_i - e_{ii}$ are the overall numbers of electrons in the valence shell of the atom A_i. An FIEM of a set of atoms can be represented by a family of BE-matrices.

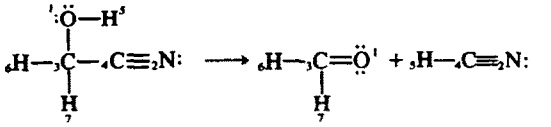
A chemical reaction corresponds to a transformation of one BE-matrix into another, $E \rightarrow E'$, which can be effected by the addition of a reaction matrix (R-matrix);[9c,d] i.e., $E + R = E'$. The entries $r_{ij} = r_{ji}$ of an R-matrix indicate which bonds are made and broken as well as the changes in the distribution of free electrons.

*Note that this definition[9c,d] is different from the one given in[9a].

The sum of the entries of an R-matrix is zero, due to the fact that electrons are conserved during chemical reactions, while the sum of the absolute values of the entries of an R-matrix, $D = \sum_{i,j} r_{ij}$ is twice the number of valence electrons participating in the reaction. Thus, an R-matrix may be considered an "electron pushing" device with double bookkeeping capability for bonds and electrons. The addition of an R-matrix to a BE-matrix, transforming $E \rightarrow E'$ occurs only if R fits E, i. e., R may have entries $r_{ij} < 0$ only where E has non-zero entries $e_{ij} \geq |r_{ij}|$. Thus the non-zero entries of E can be used to manufacture fitting R-matrices. The BE-matrix E_Y of the potential synthetic precursors of the target ensemble can be obtained from E_Z by addition of fitting R-matrices.[9d]

The reverse process, the synthesis of EM_Z from a precursor EM_Y, a synthetic step of potential interest, is represented by adding $R_Y = -R_Z$ to E_Y.

The representation of the conversion of α-hydroxy acetonitrile (1) into its components 2 + 3 is used to illustrate the action of an R-matrix on a BE-matrix.



1 2 3

SCHEME 1

Thus, one resorts to the "reverse strategy" where, beginning with the target molecule, precursors are generated until reasonable starting materials are found. Here, we are faced with the fact that when R-matrices are applied to the target BE-matrix alone, only those precursors arise that yield no synthetic by-products, i.e., only cycloadditions, ring openings, rearrangements, disproportionations etc. where no water, CO_2 etc. are formed. This requires us to augment the target BE-matrix with a small number of these by-products. That this number is small results from the observed fact that most by-products of potentially useful reactions belong to a limited list of relatively simple compounds e.g., H_2O, NaCl, CO_2, $(C_6H_5)_3PO$ etc., or their formal equivalents.

CICLOPS, a computer program we have designed based upon the foregoing mathematical representation for chemical systems, takes full advantage of the fact that BE- and R-matrices are easily adaptable to manipulations by a computer.[10]

CICLOPS is initiated by representing the target Z and its by-products Z′ as a BE-matrix E_Z. The same standard set of by-products Z′ is used for any target. Certain R-matrices which are generated from E_Z are added to the target BE-matrix producing BE-matrices E_Y of precursors. This process is repeated until precursors are found that match compounds contained in a starting material library. The R-matrices actually refer to the inverses

$$\begin{array}{c|ccccccc} & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 4 & 0 & 1 & 0 & 1 & 0 & 0 \\ 2 & 0 & 2 & 0 & 3 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 4 & 0 & 3 & 1 & 0 & 0 & 0 & 0 \\ 5 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 6 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 7 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{array} + \begin{array}{c|ccccccc} & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & & & +1 & & -1 & & \\ 2 & & & & & & & \\ 3 & +1 & & & -1 & & & \\ 4 & & & -1 & & +1 & & \\ 5 & -1 & & & +1 & & & \\ 6 & & & & & & & \\ 7 & & & & & & & \end{array} = \begin{array}{c|ccccccc} & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 4 & & 2 & & & & \\ 2 & & 2 & & 3 & & & \\ 3 & 2 & & & & & 1 & 1 \\ 4 & & 3 & & & 1 & & \\ 5 & & & & 1 & & & \\ 6 & & & 1 & & & & \\ 7 & & & 1 & & & & \end{array}$$

$E_{(1)}$ R $E_{(2)+(3)}$ Eq (1)

(The zeros in Eq (1) have been omitted from the R-matrix and product BE-matrix).

1.3. *Basic strategy of CICLOPS*

An organic synthesis planning program based upon the "forward strategy", i.e., beginning with a selection of probable starting materials and proceeding through the various intermediates to the goal molecule is, in principle, possible with these algebraic structures as discussed above. However, the problems surrounding this approach require not only solution of a target evaluation in an effort to shrink the number of reasonable starting materials but also evaluation of a large number of pathways leading to the target. While attractive, this alternative would result in a program whose machine time requirements make it presently impractical.

1.4. *The need for an encoding procedure*

Internally, any program of this nature requires a coding mechanism for rapidly identifying the constitution and stereochemical features of molecules. Such a code would allow recognition of duplicate precursors as well as those existing in a starting material library. Further, the encoding procedure should indicate constitutionally equivalent atoms whose presence might lead to the generation of equivalent precursors.

There exist many encoding procedures yielding linear representations of chemical constitutions, each with its own particular merits suited to the of synthetic reactions which have been called transforms.[4]

purposes for which it was designed.[11] However, during the development of our program it became apparent that none of the existing codes could be conveniently used to assign uniquely indexed BE-matrices to an ensemble of molecules. The row and column indices of an n×n BE-matrix can be chosen in up to n! different ways. Thus for the BE-matrix to be an efficient device suitable for the computer use which we intend for it, a procedure had to be found whereby each chemical system has an unequivocal representation in a BE-matrix or corresponding linear code.

2. SEQUENTIAL ENCODING OF CHEMICAL CONSTITUTION

2.1. *Information required for representing a chemical constitution*

The chemical constitution of a molecular system is usually represented by its constitutional formula which is a labelled graph. In this graph the lines are covalent connectivities and the vertices correspond to atoms labelled by chemical element symbols. Chemical constitution can also be represented by covalent bond lists, a list of covalently bonded pairs of indexed atoms. An unequivocal assignment of a covalent bond list to a chemical constitution requires a unique way of numbering or indexing the atoms belonging to the chemical constitution. This amounts to the solution of the problem of classifying labelled graphs.[12]

2.2. *Assignment of atomic sequence indices*

An unambiguous assignment of indices to the set of atoms belonging to a chemical constitution is accomplished as follows. This procedure stems from the nature of the representation for chemical constitution and as such lends itself to computer manipulation.[13]

1. Atoms of the same chemical element are collected into equivalence classes of elements which are arranged in descending order of atomic number.

2. Each member of an element equivalence class receives the same tentative *a*tomic *s*equence *i*ndex (ASI), V. The latter, V, is equal to one more than the total number of atoms (m) with a higher atomic number, i.e., V = m + 1.

3. When an element equivalence class contains more than one member, the tentative ASI's of the α-atoms, i.e., the first sphere neighbors of a given atom are considered. A neighbor weight vector (NWV) consisting of the neighbor ASI's is constructed. The NWV is left justified, i.e., the neighbor ASI's (tentative or final) are arranged from left to right in ascending numerical order, however, zeros come last. The total number of entries in a NWV is equal to the highest coordination number of the atom. Within an equivalence class all atoms are treated as if their coordination numbers were equal by completing their NWV with zeros if necessary.*

The NWV's of an equivalence class are compared, component for component, starting from the left. When the component with a lowest value is encountered, the corresponding central atom receives the lowest permanent ASI available in the equivalence class. Where no lowest value component is found, the components from the next position on the right are compared in the same manner. This process is continued until all atoms receive a permanent ASI or until all components have been considered.

4. When a first sphere neighbor weight vector comparison is insufficient to assign all the ASI's, the NWV of the second, third, and remaining neighbors are compared as described above until all decisions have been made or all neighbors have been considered, or, in the case of cyclic structures, backtracking to the origin has been achieved.

5. Atoms not distinguished by this procedure are considered constitutionally equivalent and can be arbitrarily assigned the remaining ASI's within the range of their equivalence class.⊖

The foregoing indexing procedure† has the following advantages:

(a) a single canonical‡ BE-matrix for any given molecule or EM is obtained by using the ASI's as row/column indices of the BE-matrix. This allows comparison with a starting material library which is coded in the same way.

(b) the ASI of atoms belonging to chemical systems which can be described by more than one VB-structure are independent of the chosen resonance formula

(c) the BE-matrix can be automatically blocked into its molecular fragments;

(d) cyclic structures are identified;

(e) constitutionally equivalent atoms and bonds are recognized;

(f) the entire process, due to our representations, is faster than other methods.

2.3. *Example for atomic sequence indexing; generation of a canonical BE-matrix*

The application of the preceding rules for coding a molecule can be illustrated with the following example.

Initially, arbitrary indices are assigned to the

*Thus, coordination numbers are taken into account, and bond orders can be neglected.

⊖The present approach permits the unique indexing of the atoms of a chemical constitution except in some rare cases of poly-homocyclic molecules, whose carbon atoms are not constitutionally equivalent but have all the same NWV. For these exceptions further rules are being introduced which are lexicographic in nature.

†Additional rules concerning oxidation state, isotopic labelling, and configurations may be added if necessary.

‡A canonical form is a standardized form particularly suited to a given purpose.

atoms of a chemical constitution (eq. Fig 1). This permits one to establish a preliminary connectivity list.

The empirical formula of the above compound, $S_1O_3N_2C_8H_{12}$, leads to the following equivalence classes and their tentative AS indices (V) (Table 1).

Equivalence class 1 has only one member, tentative and final ASI are therefore identical. For assigning final ASI's to the atoms in equivalence classes 2–5, their neighbor weight vectors (NWV) have to be considered (Tables 2 and 3).

The assignment of AS indices converts the initial labelled graph of Fig 1 into the final labelled graph of Fig 2, and results in the canonical Be-matrix of Fig. 3.

Table 1. The tentative ASI of the element equivalence classes

Equivalence classes	Tenative ASI
(1) S_1	1
(2) O_8, O_9, O_{10}	2
(3) N_4, N_{11}	5
(4) C_2, C_3, C_5, C_6, C_7, C_{12}, C_{13}, C_{14}	7
(5) H_{15}. . . . H_{26}	15

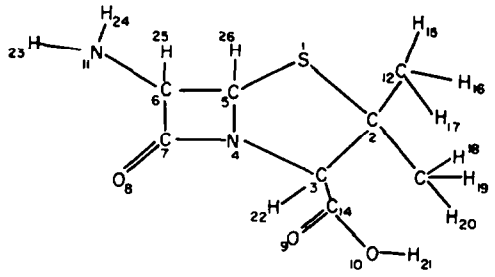


Fig 1. Formula of 6-amino-penicillanic acid (6–APA) with initial indices.

3. R-MATRIX GENERATION

For any BE-matrix E there exists a well defined set of fitting R-matrices.[9d] An R-matrix can represent a chemical reaction in which the EM participates, or a sequence of reactions that begins or ends with this EM. In other words, not every

Table 2. Assignment of the final ASI of the second element equivalence class.

Atom	Tentative ASI	1.	NWV	Intermediate ASI	2.	NWV	Final ASI
O_8	2	7	0	2		57	*3*
O_9	2	7	0	2		27	*2*
O_{10}	2	7	15	*4*			*4*

Table 3. Assignment of the final ASI to the atoms of equivalence classes 3, 4, and 5.

Atom	ASI	1. NWV	ASI	2. NWV	ASI	3. NWV	ASI		
N_4	5	7 7 7	*5*						
N_{11}	5	7 15 15	*6*						
C_2	7	1 7 7 7	*8*						
C_3	7	5 7 7 15	*11*						
C_5	7	1 5 7 15	*7*						arbitrarily
C_6	7	6 7 7 15	*12*						assigned
C_7	7	3 5 7 0	*10*						ASI
C_{12}	7	7 15 15 15	13	1 11	13	5 7 9 15	13 . . . etc.	constitutionally	*13*
C_{13}	7	7 15 15 15	13	1 11	13	5 7 9 15	13 . . .	equivalent	*14*
C_{14}	7	2 4 7 0	*9*						
H_{15}	15	13	21	8	21	1 11	21 . . .		*21*
H_{16}	15	13	21	8	21	1 11	21 . . .		*22*
H_{17}	15	13	21	8	21	1 11	21 . . .		*23*
H_{18}	15	14	24	8	21	1 11	24 . . .		*24*
H_{19}	15	14	24	8	21	1 11	24 . . .		*25*
H_{20}	15	14	24	8	21	1 11	24 . . .		*26*
H_{21}	15	4	*15*						
H_{22}	15	11	*19*						
H_{23}	15	6	16	12	16	7 10 20	16 . . .		*16*
H_{24}	15	6	16	12	16	7 10 20	16 . . .		*17*
H_{25}	15	12	*20*						
H_{26}	15	7	18						

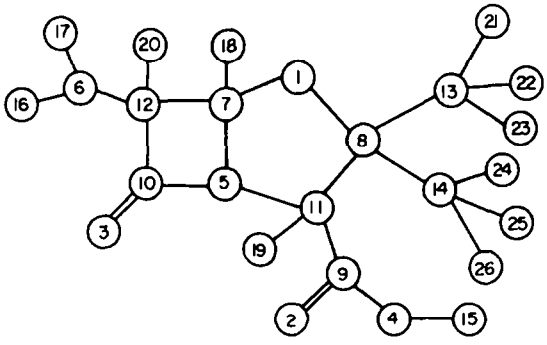


Fig 2. Final Labelled Graph of 6-APA From Fig 2 the BE-matrix of 6-amino-penicillanic acid in its canonical form (Fig 3) is obtained from the initial connectivity list by using the AS indices as row and column indices.

definitions and limited to the breaking of one, two, and three bonds shows that they may be classified into 38 general R-matrix types.[14]

The 38 R-matrix types are arranged into 3 separate groups: RGEN 1, those breaking one bond, making 0, 1 or 2 (5 types); RGEN 2, those breaking two bonds, making 0, 1, 2, (12); RGEN 3, those breaking three bonds, making 0,1,2,3 or 4, (21). R-matrices breaking 4 bonds are merely sums of

Table 4. RGEN-1 reaction types

I–J	I+J:
I–J+ Y	I+J–Y
I–J+:X	I:+J–X
I–J+:X	J–X–I
I–J+:X+Y	I–X+J–Y

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	19	20	21	22	23	24	25	26
1	4						1	1																	
2		4							2																
3			4							2															
4				4					1																
5					2		1			1	1														
6						2						1				1	1								
7	1				1							1						1							
8	1									1			1	1											
9		2	1							1															
10			2		1							1													
11					1			1	1									1							
12						1	1			1									1						
13								1												1	1	1			
14								1															1	1	1
15																									
16						1																			
17						1																			
18						1																			
19										1															
20											1														
21													1												
22													1												
23													1												
24														1											
25														1											
26														1											

Fig 3. Canonical BE-matrix, $E_{6\text{-APA}}$, of 6-amino-penicillanic acid.

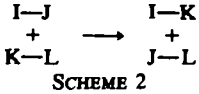
fitting R-matrix corresponds to a single step chemical reaction but can refer as well to a sequence of reactions whose R-matrix is the sum of R-matrices for the individual reactive steps. This foreshadows a representation leading to an analysis of single step mechanistic alternatives.

The closed shell R-matrices[9d] may have off-diagonal entries of O, ±1, ±2, and ±3 and diagonal entries of 0, ±2. In CICLOPS certain restrictions are placed upon them. For example, they may have only one, two or three pairs of negative off-diagonal entries ±1 and, at most, a pair of row/column sums $\Sigma_j r_{ij} = \pm 1$, all other row/column sums being zero. Analysis of these R-matrices, based upon their RGEN 2 matrix pairs.[9d] The members of RGEN 1 are given as an example for the series: (Table 4). For example the matrix (Fig 4) of RGEN 2

$$R = \begin{array}{c} \begin{matrix} i & j & k & l \end{matrix} \\ \begin{bmatrix} 0 & -1 & +1 & 0 \\ -1 & 0 & 0 & +1 \\ +1 & 0 & 0 & -1 \\ 0 & +1 & -1 & 0 \end{bmatrix} \end{array} \begin{matrix} i \\ j \\ k \\ l \end{matrix}$$

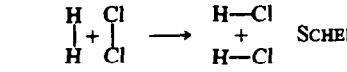
Fig 4.

corresponds to the reaction type (Scheme 2)

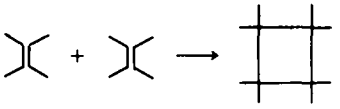


SCHEME 2

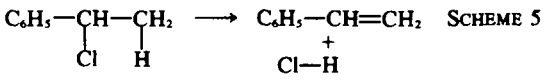
and includes among others, the chemical reactions of Schemes 3, 4 and 5



SCHEME 3



SCHEME 4



SCHEME 5

Each reaction type is generated with the assignment of non-zero entries of an R-matrix. All possibilities for each R-matrix type acting upon a single BE-matrix are achieved by permutations on the indices. In order to avoid meaningless, redundant or uninteresting reactions a sub-set of R-matrices is chosen on the basis of some selection rules. First, in the BE-matrix, certain bonds including all multiple, carbon or hydrogen to heteroatom and those one and two neighbors distant are flagged as breakable. Before fitting the R-matrix, atoms of the target molecule are scanned and evaluated against the chemical limits table. The latter is comprised of electronic data for atoms, e.g., maximum permissible charge, valencies, electron population etc. This valence analysis eliminates many R-matrices which, although fitting, would generate electronically unacceptable BE-matrices. For example, if atom A_i has its maximum permissible number of bonds and free electrons, R-matrices increasing the number of surrounding electrons are ignored. R-matrices exchanging atoms of the same element are similarly ignored. Other, more specific rules, avoiding certain multiple bonds, etc. are included.

4. REPRESENTATION OF STEREOCHEMISTRY

Those stereochemical features of a synthesis which must be taken into account may be represented in terms of monocentric configurations and their combinations. With very few exceptions, the configurational aspect of organic syntheses can be confined to tri- and tetra-coordinate skeletons.[15a]

4.1. *Configurations and mappings; Stereochemistry of molecules*

A configuration is determined by a set of m ligands $L = \{L_1, \ldots L_m\}$ and their placement on the ligand sites of a skeleton $S = \{S_1, \ldots S_m\}$ with a coordination number m and a characteristic rigid model symmetry.[15] The description of a configuration invokes the assignment of indexed ligands to independently indexed skeletal sites.[15] Although the Cahn–Ingold–Prelog Sequence Rules[16] are widely accepted and used to describe the stereochemistry of carbon compounds, their incorporation into a computer program is an awkward procedure. An indexing algorithm originating from a mathematical model makes better use of the advantages a computer has to offer.

We represent a configuration by mapping the relative atomic sequence indices (RASI) of the coordination center's α-atoms onto indexed skeletal positions. The α-atoms are those directly bonded to the central atom of the configuration. RASI = 1, . . .m(m = 3 or 4) are the indices of the α-atoms assigned from the numerical order of their ASI's.

Skeletal indices may be arbitrarily assigned but then must be retained. The indexing of tri- and tetra-coordinate skeletons is described in Fig 5.

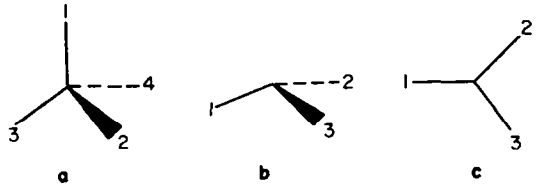


Fig 5.

For the skeleton **a** and **b** there are two distinguishable ways of indexing. For the planar skeleton **c**, however, two distinguishable ways of indexing result only if one specifies from which side of the skeletal plane it is being viewed. Pairs of indexed type **c** skeletons are used to specify the stereochemical features of idealized planar double bond systems.

4.1.1. *Stereochemical descriptors.* That configuration in which the α-atom with the lowest RASI occupies the skeletal site 1, the α-atom with the second lowest RASI occupies the skeletal site 2, etc, is the reference isomer.[15b] A permutation of the α-atoms which transforms the reference isomer into the configuration in question can be used as the descriptor of the configuration.

The reference isomer in the case of a 4-coordinate central atom with T_d skeletal symmetry can be represented by Eq (2) where l stands for the RASI and s for the skeletal indices as shown in Fig 6.

$$\binom{l}{s}_{ref} = \binom{1\,2\,3\,4}{1\,2\,3\,4}_{ref} \qquad \text{Eq (2)}$$

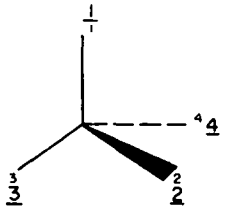


Fig 6.

$$P\begin{pmatrix}1\\ s\end{pmatrix}_{ref} = (12)\begin{pmatrix}1&2&3&4\\ \bar{1}&\bar{2}&\bar{3}&\bar{4}\end{pmatrix}_{ref} = \begin{pmatrix}2&1&3&4\\ \bar{1}&\bar{2}&\bar{3}&\bar{4}\end{pmatrix} \quad \text{Eq (3)}$$

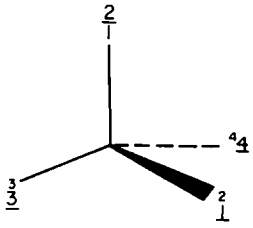


Fig 7 is derived from figure 6 by application of a RASI permutation P = (12) as shown in Eq (3).

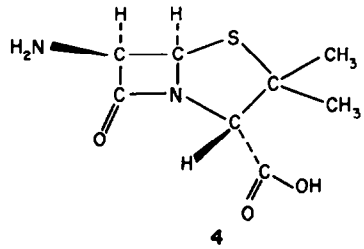
There exist pairs of configurations corresponding either to the reference configuration, or obtainable from the latter by odd* permutations of α-atoms. Accordingly, (+1) and (−1), indicating the parity of the permutation, are suitable descriptors for such configurations.†

Thus the algebraic sign group {(+1), (−1)} and its direct products e.g. {(+1) (+1) . . .} can be used for the treatment of problems involving molecules with one or more monocentric configurations.[17]

4.1.1.1. *Tetrahedral configurations.* For configurations with tetrahedral skeletons the reference isomer is chosen as in Fig 6 and is assigned the parity descriptor (+1).

The assignment of a parity descriptor is illustrated with **4** as an example.

The molecule has three chiral monocentric configurations. The ASI's are contained in Fig 2, comparison of the mapping of the RASI's onto the



4

three monocentric reference skeletons lead to the parity descriptors (Table 5).

Table 5. The assignment of parity descriptors to 6-aminopenicillanic acid (4)

Center	α-atom ASI	α-atom RASI	Parity descriptor
7	1	1	(+1)
	5	2	
	12	3	
	18	4	
11	5	1	(−1)
	8	2	
	9	3	
	19	4	
12	6	1	(+1)
	7	2	
	10	3	
	20	4	

The parity vector is $P_{6\text{-APA}} = (0,0,0,0,0,0,+1,0,0,0,-1,+1,0,0,0,0,0,0,0,0,0,0,0,0,0,0)$, and the pair ($P_{6\text{-APA}}$, $E_{6\text{-APA}}$), (see Fig 3) represents the chemical constitution and the essential stereochemical features of 6-APA (4).

If a chemical system contains two centers of chirality A and B, both capable of existence in a (+1) and a (−1) configuration, then there are four diastereomers $A_{(+)}-B_{(+)}$, $A_{(-)}-B_{(-)}$, $A_{(+)}-B_{(-)}$, $A_{(-)}-B_{(+)}$, of which two each are pairs of chiral antipodes. These are equivalent with respect to scalar properties. For many purposes, it is advantageous to use the product of the algebraic signs as a common descriptor for pairs of antipodes, i.e., (+1) for the (+1) (+1) and the (−1) (−1) diastereomer, and (−1) for the (+1) (−1) and the (−1) (+1) case.[17]‡

4.1.1.2. *Double bond systems.* The reference configuration for the monocentric tri-coordinate center is chosen as in Fig 8 and is assigned the stereochemical descriptor (+1).

This descriptor is the parity resulting from a permutation similar to that described for the tetracoordinate system.

A pair of such tricoordinate monocentric reference configurations can be used to represent the idealized planar reference isomer for double bond systems. All triogonal centers of π-bond systems must be viewed from the same side of the molecular plane. Although the descriptors of the

*The exchange of two adjacent symbols, as represented by the permutation (12) in Eq (3) is called a transposition. A permutation which contains an odd number of transpositions is an odd permutation and has a parity of −1. An even permutation contains an even number of transpositions. A cyclic permutation of n symbols can be achieved by n − 1 transpositions, i.e., a cycle of even numbers of symbols is an odd permutation and vice-versa.

†To regenerate the R,S-nomenclature[16] commonly used in chemistry today a subroutine re-indexing the ligands around a chiral center is implemented at the end of a synthesis search. In most cases, however, the parity descriptor (+1) corresponds to (R), and (−1) to (S). A direct extension of the Cahn-Ingold-Prelog Sequence Rules into an indexing system of the atoms of a chemical constitution would not serve the present purpose. With these rules the different resonance structures of delocalized bond systems can lead to different indexing of atoms, and constitutionally equivalent atoms, such as the two ortho-carbons of chlorobenzene would be indexed, as if they were not equivalent.

‡Sometimes the necessity arises for one to represent configurations and configurational relations in racemic compounds. There (±1) and (±1) are used as the respective components of the parity vector, and multiplication according to (±1)(±1) = (+1) and (±1)(∓1) = (−1) represents the pair-wise configurational relation.

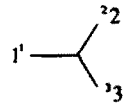
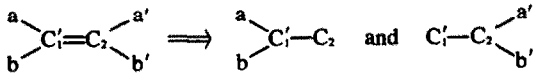


Fig 8.

monocentric configurations change their sign when viewed from the opposite side of the plane the signs of their pairwise products remain invariant and can be used similarly to *cis trans* - or E-Z[18] descriptors of double bond systems.

To generate the tri-coordinate centers one dissects a simple olefin (5) as follows:



5 6a 6b

SCHEME 6

In other words the central atom of one configuration is the α-atom of the other one.

Translation of the pairty descriptor $P_{OL} = P_1 \cdot P_2 (= \pm 1)$ for olefinic systems into an ASI-based *cis-trans* or E-Z nomenclature P_{CT} makes use of the fact that the sum of RASI's for the two doubly bonded sp^2 carbon centers, i.e., Σ RASI = $RASI_C$ (C) + $RASI_C$ (C'), correlates the vector relationship for the remaining pairs of α-atoms. When this sum is even, the + 1 parity descriptor defines a *trans*- or E-arrangement of the two ligands with the higher priority indices. Conversely, the odd sum + 1 parity descriptor defines a *cis* relation for the priority indexed pair. The following equation and Table 6 summarize this correlation.

$$P_{CT} = P_{OL} \cdot (-1)^{\Sigma \text{ RASI}} \qquad \text{Eq (4)}$$

4.2. *Representation of the stereochemical features of reactions*

In chemical reactions there is a correlation between the stereochemical features of starting materials and products which is characteristic of the reaction involved. Accordingly, the stereochemistry of synthetic precursors can be derived from the target and the reactions leading to it. Formally, reactions can be represented by Eq (5)

$$(E, P) \xrightarrow{(Q, S)} (E', P') \qquad \text{Eq (5)}$$

Here E and E' stand for the chemical constitution of the target and its precursor; P and P' represent their stereochemical features. Q is an operator representing a reaction and acting upon the chemical constitution of the initial system transforming it into the chemical constitution of the final system. S is an operator accomplishing the associated changes in the stereochemistry.

In our representation of the constitution, E and E' are canonical BE-matrices. Then the transformation $E \xrightarrow{Q} E'$ is effected by adding a reaction matrix R to E and re-indexing the resulting BE-matrix. $P = (P_1, P_2, \ldots, P_n)$ and $P'_{i'} = (P'_1, P'_2, \ldots, P'_n)$ are parity vectors whose components P_i and P'_i ($= 0, \pm 1$) are the parity descriptors of the monocentric configurations about the original atom A_i whose index after transformation Q is i'.

The operator $S = (S_1, S_2, \ldots, S_n)$ transforms P into P' such that

$$S_i \cdot P_i = P'_{i'} \qquad \text{Eq (6)}$$

The combination of S_i with P_i is defined as a multiplication if $S_i \neq 0$ and $P_i \neq 0$ and as an addition if $P_i = 0$, i.e.,

$$P'_{i'} = \begin{cases} S_i \times P_i, & \text{if } S_i \neq 0, P_i \neq 0 \\ S_i + P_i, & \text{if } P_i = 0 \end{cases} \qquad \text{Eq (6a)}$$

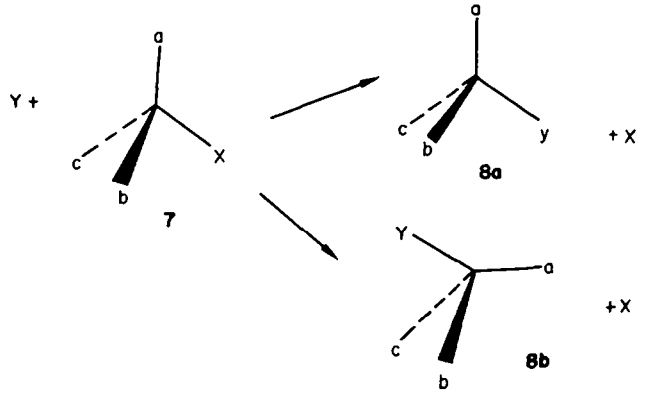
The components S_i ($= 0, \pm 1$) indicate the formal configurational changes at the atom A_i and take into account substitution, rearrangement, changes in coordination numbers, as well as changes in the relative indices of the α-atoms.

4.2.1. *Interconversions of tetracoordinate systems.* If a chemical reaction replaces a ligand of a tetrahedral configuration with another ligand, (Scheme 7), the entering moiety can occupy either the same skeletal site (considering rotations of the skeletal symmetry group as equivalent) as the

Table 6. Correlation of parity descriptors P_{OL} with *cis trans* nomenclature P_{CT}

RASI of sp^2 centers		$\sum$ RASI	$\sum^{\text{RASI}}$	parity descriptor P_{OL}	
				= +1	= −1
C'_1	C_2		(-1)	corresponds to P_{CT}	
1	1	even	+1	*trans*	*cis*
1	2	odd	−1	*cis*	*trans*
1	3	even	+1	*trans*	*cis*
2	2	even	+1	*cis*	*trans*
2	3	odd	−1	*trans*	*cis*
3	3	even	+1		

leaving group or one obtainable from the original site by a skeletal reflection operation.



SCHEME 7

In order to recognize retention or inversion in the product one conceptually superimposes initial and final tetracoordinate centers upon each other such that entering and leaving groups occupy the same site.

Then, in the product of a reaction occuring with inversion, the arrangement of the other ligands corresponds to the mirror image of the starting material (heterochiral relationship)[19] and the product of a substitution with retention resembles the configuration of the starting material itself (homochiral).[19] Correlation of the configurational descriptors of starting materials and products must take into consideration the stereochemical course of the reaction (inversion or retention) as well as the fact that the RASI of the entering and leaving groups may differ.

Throughout such reactions the RASI's of the remaining three α-atoms stay in the same order with respect to each other. Chemical reactions by which the RASI of an α-atom of a coordination center is changed, by a reaction occuring in the outer part of a ligand, without breaking or making bonds to their central atom, are formally treated as retentive substitution reactions.

Different RASI for entering and leaving groups induce changes in the ordering of the α-atom set. These changes are due only to the choice of indexing system used to represent stereochemistry, not to the nature of the phenomenon itself. This can result in an identical stereochemical descriptor for the starting material and product of a reaction which proceeds with inversion, or the corollary, different descriptors for a retention reaction.*

To eliminate this problem, a nomenclature factor F is introduced by mapping the product onto the starting material, both in their reference isomer form (see 4.1.), such that the entering group occupies the position of the leaving group while the remaining ligands retain their previous skeletal positions. This can easily be accomplished by the proper ligand permutation. The number K' of transpositions contained in this is equal to the absolute value of the difference in the RASI of entering and leaving α-atoms.

$$K' = |RASI_{(7)}(X) - RASI_{(8)}(Y)| \qquad \text{Eq (7)}$$

This value K' can be used to determine the parity F of the permutation effected on the α-atom set by the choice of the nomenclature system.

$$F = (-1)^{K'} \qquad \text{Eq (8)}$$

It can be seen that $F = +1$ if the RASI of both X and Y are either even or odd and $F = -1$ if one is odd and the other is even. Therefore also Eq (9) can be used to determine the parity of the permutation

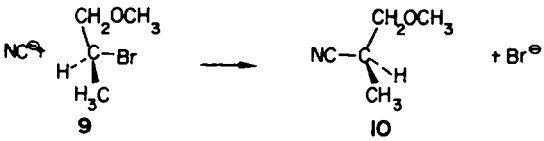
$$F = (-1)^{K}, \text{ with } K = RASI_{(7)}(X) + RASI_{(8)}(Y) \qquad \text{Eq (9)}$$

Table 7 expresses this consequence in detail. The component S_i of the stereochemical operator associated with a ligand replacement reaction at center A_i is then expressed as

$$S_i = M \cdot F_i \qquad \text{Eq (10)}$$

where F_i is the nomenclature factor and M is the descriptor for the stereochemical type of the reaction, $M = +1$ for retentions (homochiral relationship of starting material and product); $M = -1$ for inversion (heterochiral). For instance, the S_N2 reaction of cyanide with (S) – 1-methoxy-2-bromopropane ($P_i = -1$) (**9**) leads to (S) – 1-methoxy-2-cyano-propane ($P'_i = -1$) (**10**). Although this reaction (Scheme 8) proceeds with inversion, $M = -1$, both starting material and product have the same

*This problem exists in other nomenclature systems (e.g. R, S) as well.



SCHEME 8

stereochemical descriptor, because the RASI (Br) = 1 (odd), whereas the C-atom of the entering cyanide ion has RASI (C) = 2 (even).

The ASI sequence of the α-atoms of the ligands in the starting material is (Xabc), in the product (aYbc), with the ligand $a = CH_2OCH_3$ having changed its RASI from 2 to 1.

To restore the original order of the ligand set (X and Y occupying the same skeletal site), one (K' = 1) transposition (aY) is necessary, as given by Eq (7).

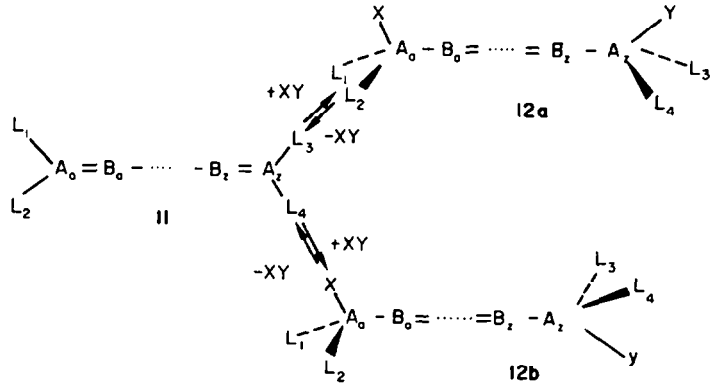
The parity change due to the nomenclature system (nomenclature factor) is then $F = (-1)^1 = -1$ (Eq (8)).

The component of the stereochemical operator S of the reaction type is (Eq (10))

$$S_i = M \cdot F_i = (-1) \cdot (-1) = (+1) \qquad (i = 3) \qquad \text{Eq (10)}$$

Thus, from Eq.(8) one obtains $P'_{i'} = S_i \cdot P_i = (+1) \cdot (-1) = -1$.

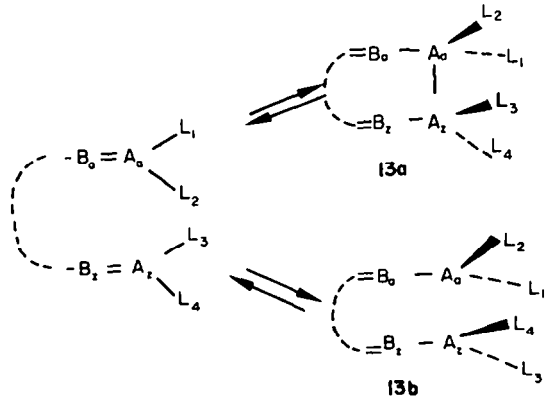
4.2.2. *Interconversions of tri- and tetracoordinate systems.* There is a great variety of reactions involving interconversions of systems containing pairs of tri- and tetra-coordinate centers, e.g., additions, eliminations, electrocyclic reactions, and cycloadditions. These reactions can be represented by Schemes 9 and 10.



SCHEME 9

Table 7. The nomenclature factor of substitution reactions as a function of the RASI of the α-atom of leaving and entering group

$RASI_{(7)}$ (X)	$RASI_{(\alpha)}$ (Y)	K'	F
1	1	0	+1
1	2	1	−1
1	3	2	+1
1	4	3	−1
2	1	1	−1
2	2	0	+1
2	3	1	−1
2	4	2	+1
3	1	2	+1
3	2	1	−1
3	3	0	+1
3	4	1	−1
4	1	3	−1
4	2	2	+1
4	3	1	−1
4	4	0	+1



SCHEME 10

The tricoordinate reactive centers and their α-atoms (A_a, B_a, L_1, L_2 and A_Z, B_Z, L_3, L_4, respectively), are coplanar in the reactant (11). The general formula 11 also includes simple olefins; here B_a is identical with A_Z and B_Z with A_a.

The reaction **11** → **12a** is a *cis*-addition reaction where the addends X and Y enter from the same face of **11**. This may also be called a *homofacial* addition.* The reverse process is a *cis*- or *homofacial* elimination. The conversion of **11** → **12b** involves a *trans*-addition where the reactants X and Y are attached to opposite faces of **11**, which may be called a *heterofacial* addition.* In Scheme 10 the conversion **11** → **13a** is a disrotatory electrocyclic process which can be described as a homofacial addition with A_a being added to A_z, and A_z to A_a. The reverse process corresponds to a disrotatory ring opening or homofacial elimination.

The interconversion **11** ⇄ **13b** comprises conrotatory[21] or heterofacial processes. In cycloadditions, the participants are formally treated as if they undergo two separate addition reactions according to Scheme 9.

In reactions of this type where tricoordinate centers are converted into tetracoordinate systems or the reverse, the correlation of the parity descriptors for the two different skeletal geometries is the main problem. Stereochemical features for both systems can be represented with ±1's (Section 4.1.1.). Analogous to the case involving reactions affecting tetrahedral centers, components S_a and S_z of the stereochemical operator S account for the faciality of the reaction and the associated RASI changes. (Eqs 11 and 12)

We have

$$S_a \cdot P_{a,tri} = P_{a,tetra}$$
$$S_z \cdot P_{z,tri} = P_{z,tetra} \qquad \text{Eq (11)}$$

and for the reverse process

$$S_a \cdot P_{a,tetra} = P_{a,tri}$$
$$S_z P_{z,tetra} = P_{z,tri} \qquad \text{Eq (12)}$$

Since the faciality M of a reaction relates to pairs of configurations, the components S_a and S_z must be treated as pairwise dependent.

$$S_a \cdot S_z = M \cdot F_a \cdot F_z \qquad \text{Eq (13)}$$

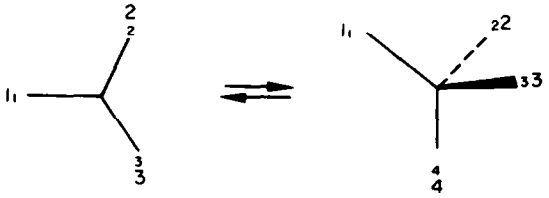
There F_a and F_z are the nomenclature factors accounting for RASI changes at the centers A_a and A_z.

M, the reaction faciality operator, where +1 denotes homofacial and −1 heterofacial attack (or elimination) is chosen as a function of the reaction type. The latter is obtained from a combination of the target BE-matrix and the R-matrix. One fourth of the sums over the absolute values of the entries of an R-matrix ($D = \sum_{ij} r_{ij}$) is the number of electron pairs participating in a reaction.[9c,d] Accordingly, the faciality of concerted electrocyclic, sigmatropic and related reactions[21c] can be determined.

Eq (13) indicates that only the product of S_a and S_z is uniquely determined but not the components themselves. $S_a \cdot S_z = +1$ implies that both components have the same sign while the negative product of the two results from opposite parities.† This reflects the fact that the double bond system has a plane of symmetry and can be viewed or attacked from both sides. Thus, pairs of enantiomeric systems with two centers of chiralty each are involved simultaneously and treated as equivalent.

The change in coordination number prevents the use of starting material configuration as a reference for the product. To obtain the nomenclature factors for these reactions one must correlate the reference isomers of the tri- and tetracoordinate configurations.

In the tricoordinate reference isomer the skeleton is indexed as in section 4.1.1., and the α-atoms RASI's 1, 2, 3 are mapped onto the numerically equal skeletal indices. The tetracoordinate reference isomer is represented by a mapping of the RASI's 1, 2, 3 and 4 of the α-atoms onto the skeletal indices 1, 2, 3, 4 (Fig 6, section 4.1.1.). The α-atom RASI's 1, 2, 3 refer to the same atoms in the tricoordinate and tetracoordinate species. The α-atom of the additional ligand X or Y with RASI = 4 occupies skeletal position 4. The transition of the tricoordinate reference configuration into the tetracoordinate reference configuration is illustrated by Scheme 11. For the interconversion of the two reference isomers the nomenclature factor F = +1.



SCHEME 11

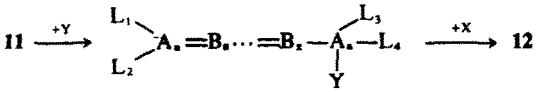
The participants of a reaction must be compared with their respective reference isomers in order to determine the nomenclature factor. However, possible RASI changes resulting from reactions converting tri- into tetracoordinate centers may prevent the placement of α-atoms on their proper reference sites. Thus, the α-atom priority sequence defined in the reference isomer must be restored. The parity of the permutation accomplishing this reordering is used to generate the nomenclature factor, F.

In the reaction **11** → **12** of Scheme 9 the attachment of Y to A_z and X to A_a illustrates the two different ways in which the RASI's may be affected. Scheme 12 depicts the stepwise addition

*The terms "homofacial" and "heterofacial" which were proposed by Ruch to Prelog[29] seem to be more appropriate in this context than suprafacial and antarafacial.

†$S_a \cdot S_z = +1 \Rightarrow S_a = S_z = \begin{cases} +1 \\ -1 \end{cases}$

$S_a \cdot S_z = -1 \Rightarrow S_a = -S_z = \begin{cases} +1 \\ -1 \end{cases}$



14

SCHEME 12

of XY to **11** used to explain the two step procedure employed in finding F. Here, the tricoordinate moiety in question has atom A_a at its center.

Step 1: Addition of Y to A_z may change the RASI of B_a in the subset L_1, L_2, B_a, i.e., $RASI_{(11)}$ $(B_a \in \{L_1, L_2, B_a\})$ may differ from $RASI_{(14)}$ (B_a) $(B_a) \in \{L_1, L_2, B_a\})$. Since the sequence of the remaining two α-atoms of L_1 and L_2 is maintained, the reordering required for generating the reference isomer can be effected with a permutation where the number of transpositions, K_1' is given by Eq (14)

$$K_1' = |RASI_{(11)}(B_a) - RASI_{(14)}(B_a)| \quad \text{Eq (14)}$$

and the parity of the permutation is $(-1)^{K_1} = (-1)^{K_1'}$ where

$$K_1 = RASI_{(11)}(B_a) + RASI_{(14)}(B_a) \quad \text{Eq (15)}$$

In the actual case, it is unnecessary to consider the hypothetical intermediate **14**. The $RASI_{(12)}$ (B_a), when determined without considering X is equal to $RASI_{(14)}$ (B_a). In other words one calculates the $RASI_{(12)}$ (B_a) *only* in the α-atoms subset of L_1, L_2, and B_a.

Step 2: The second part of the addition process is the attachment of X to A_a, reaction **14**→**12** in Scheme 12. Since the α-atom of X has been defined by the reference isomer to have RASI 4 and be mapped onto position 4, a permutation must be applied to the α-atom set when, in fact, X has a RASI other than 4. This is accomplished in a manner similar to that in step 1 except that in this case.

$$K_2 = 4 - RASI_{(12)}(X) \quad \text{Eq (16)}$$

Further, as 4 is even, the parity of the required permutation is $_{(-1)}RASI_{(12)}(X) = (-1)^{K_2}$

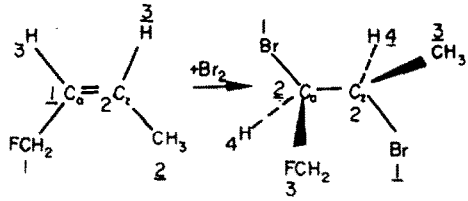
Combining steps 1 and 2, we can calculate K_a:

$$K_a = K_1 + K_2 = RASI_{(11)}(B_a) + RASI_{(12)}(B_a) \quad (B_a \in \{L_1, L_2, B_a\}) + RASI_{(12)}(X) \quad \text{Eq (17)}$$

Then, the nomenclature factor F is formed from Eq (18):

$$F_a = (-1)^{K_a} \quad \text{Eq (18)}$$

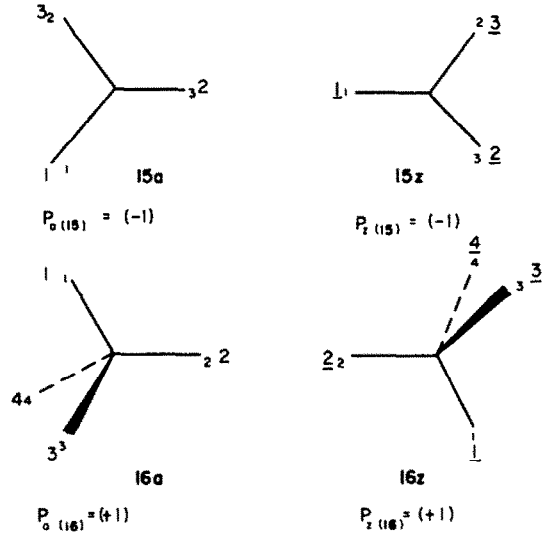
An example illustrates the above procedure: The heterofacial addition (M = −1) of bromine to **15** yields **16** and its enantiomer $\overline{16}$.



15 16

SCHEME 13

The indices in Scheme 13 are the RASI's of the ligands' α-atoms of the four centers in question. Parity descriptors are obtained according to the formal dissection:



SCHEME 14

The nomenclature factor for the reaction **15**→**16**+$\overline{16}$ is obtained by substituting in Eq (18) as follows:*

$$K_a = RASI_{(15)}(C_z) + RASI_{(16)}(C_z)(C_z \in \{H, CH_2F, C_z\}) + RASI_{(16)}(Br) = 2 + 1 + 1 = 4 \quad \text{Eq (19)}$$

$$K_z = RASI_{(15)}(C_a) + RASI_{(16)}(C_a)(C_a \in \{H, CH_3, C_a\}) + RASI_{(16)}(Br) = 1 + 1 + 1 = 3 \quad \text{Eq (20)}$$

Now, the nomenclature factors are:

$$F_a = (-1)^{K_a} = (-1)^4 = +1$$
$$F_z = (-1)^{K_z} = (-1)^3 = -1 \quad \text{Eq (21)}$$

*Note that $RASI_{(16)}(C_z) \equiv RASI_{(16)}(C_z)(C_z \in \{H, CH_2F, C_z, Br\}) = 2$, but $RASI_{(15)}(C_z)(C_z \in \{H, CH_2F, C_z\}) = 1$, and that there are analogous discrepancies at the other center of **16**.

Thus we have

$$S_s \cdot S_z = M \cdot F_s \cdot F_z = (-1)(+1)(-1) = (+1) \qquad \text{Eq (22)}$$

and we see that, according to Eq (11)

$$S_s \cdot S_z \cdot P_{s(15)} \cdot P_{z(15)} = P_{z(16)} \cdot P_{z(16)} = +1 \qquad \text{Eq (23)}$$

From this follows

$$P_{s(16)} = P_{z(16)} = \begin{cases} +1 \\ -1 \end{cases}$$

corresponding to **16** or $\overline{\mathbf{16}}$.

5. CICLOPS

5.1. *General features*

Transformation of a BE-matrix E_Z of the augmented target by all fitting R-matrices generates the complete family of BE-matrices of the FIEM containing EM_Z. This could be used to synthetic design by first analyzing the FIEM for those EM_A that contain only available starting materials, and then establishing pathways from these EM_A to EM_Z via sequences of intermediates. Instead, a tree of pathways through a subset of the FIEM is created.

CICLOPS has been tailored such that it corresponds to a practical approximation for such an idealized synthetic program, promising effectiveness without serious neglect of interesting intermediates or pathways. This is accomplished with the introduction of selection rules (Section 3.0.) as well as evaluation procedures for precursors which limit their number. Further, CICLOPS is confined to closed shell chemistry, i.e., to chemical systems whose valence orbitals contain pairs of electrons with opposite spin. The diagonal entries of closed shell BE-matrices are zero or even integers. While this eliminates radical chemistry no pathways of synthetic interest are lost, because for any radical pathway there exists a nonradical alternative.[94]

CICLOPS has been developed on an IBM 360/91 computer using PL/1, a computer language possessing the flexibility needed for list handling, algebraic manipulation and calculations.[22] Initial design and testing stage have been accomplished in the batch mode, i.e., the program runs to completion without interaction with the chemist, such that the algebraic models could be tested with as little human prejudice as possible. Presently, molecules are output as constitutional formulas.* For the duration of the initial testing period, whose primary interest rests with R-matrix generation and evaluation refinements, stereochemistry has been temporarily set aside for the sake of simplicity and speed.

*The conversion of BE-matrices into constitutional formulas is done with a program developed by R. J. Feldmann[95] at NIH.

5.2. *Data structures*

5.2.1. *BE-matrices.* Storage of the entire BE-matrix and corresponding atomic numbers would require $n(n+1)$ entries consisting mainly of zeros. Therefore, internally the BE-matrix is represented by an upper triangular, packed bond matrix (a weighted connectivity matrix), a vector of unshared valence electron pairs (diagonal entries of the BE-matrix), and a vector of atomic numbers. This is illustrated using 6-amino-pencillanic acid (Section 2.3) as an example.

Row	1	1	2	3	4	4	5	5	5	6	6	6	
Column	7	8	9	10	9	15	7	10	11	12	16	17	
Bond	1	1	2	2	1	1	1	1	1	1	1	1	

Row	7	7	8	8	9	10	11	12	13	13	13	14	14	14
Column	12	18	13	14	11	12	19	20	21	22	23	24	25	26
Bond	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Atom	1	2	3	4	5	6	7	8	9	10	11	12
Free Electrons	4	4	4	4	2	2	0	0	0	0	0	0

Atom	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Free Electrons	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Atom	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Atomic Symbol	S	O	O	O	N	N	C	C	C	C	C	C	C	C

Atom	15	16	17	18	19	20	21	22	23	24	25	26		
Atomic Symbol	H	H	H	H	H	H	H	H	H	H	H	H		

When used by the RGEN program, the upper triangular BE-matrix is expanded to a complete packed matrix enabling fast access to row sums and connectivities.

5.2.2. *R-matrices.* R-matrices are not represented internally in matrix form. Rather, the negative off-diagonal entries are added to the packed BE-matrix, the diagonal entries are added to the BE-diagonal vector and the positive off-diagonal entries are contained in an 8×3 matrix of made bonds.

The resulting Be-matrix will be evaluated by EVAL 1, a procedure that eliminates trivial or unreasonable structures based upon thermochemical, steric and similar considerations. If not rejected, the new BE-matrix is separated into molecular fragments which are sequentially ordered and further evaluated.

5.3 *Basic design*

CICLOPS takes a target molecule as input, sequentially orders it, augments it with by-products and applies pre-selected R-matrices generating first level precursors. Based upon several evaluation procedures, a selected group of precursors is saved on a file. This file is then used as input for generation of the next level's precursors which are

similarly evaluated and saved on the corresponding level file. This process is repeated using the most recently generated file as input creating a tree structure of synthesis. A synthesis is now considered complete when each component is found in the starting material library. The foregoing is illustrated in the design flow chart of Fig 9. At the time of submission of this manuscript CICLOPS had been implemented to the point where the First level of precursors is written on the save File.

6. CONCLUSION

CICLOPS originates with an algebraic model for the representation of chemical constitution and interconversions thereof. In the process of program development it became necessary not only to define unique BE-matrices but also to establish a formalism for including the synthetically relevant stereochemical features of chemical systems and reactions.

An indexing algorithm has thus been implemented providing BE-matrices in canonical form as well as a linear code for describing chemical constitution.* Further, constitutionally equivalent atoms are identified. This algorithm also leads to the assignment of permutational parities for tri- and tetracoordinate configurations as their descriptors. An operator, acting upon this descriptor serves to affix any stereochemical changes resulting from a reaction. Combining the stereochemical descriptor with the linear code yields the key with which a starting material compound library can be quickly searched and accessed.

The code, as presented here, or its compact form* treats any aspect of chemical constitution which, independent of CICLOPS, could find applications in many areas of chemical documentation and information retrieal. This coding of chemical constitution can be easily understood and accomplished by non-specialists. Further, since the linear code and stereochemical descriptors are all numbers, they can be handled quickly and stored easily by the computer. This connectivity code is easily translated into graphic output by the computer as well. It should be noted that R-matrices have the potential for forming the basis of a system for documentation of reaction types and mechanisms.

ABBREVIATIONS

6-APA	6-Amino Penicillanic Acid
ASI	Atomic Sequence Index
BE-MATRIX	Bond and Electron Matrix
CICLOPS	Computers in Chemistry, Logic Oriented Planning of Syntheses
EM	Ensemble of Molecules
FIEM	Family of Isomeric Ensemble of Molecules
NWV	Neighbor Weight Vector
RASI	Relative Atomic Sequence Index
RGEN	Reaction Matrix Generator
R-MATRIX	Reaction Matrix

*Further abbreviation of packed upper-triangular BE-matrices (Fig. 8) is possible by neglecting bonds involving H-atoms and indicating only coordination numbers differing from the respective standard ones. Beyond that, classes of constitutionally equivalent bonds can be represented by sets of their numbers. The ASI also provide a partitioning of empirical formulas into classes of constitutionally equivalent atoms yielding more detailed empirical formulas.

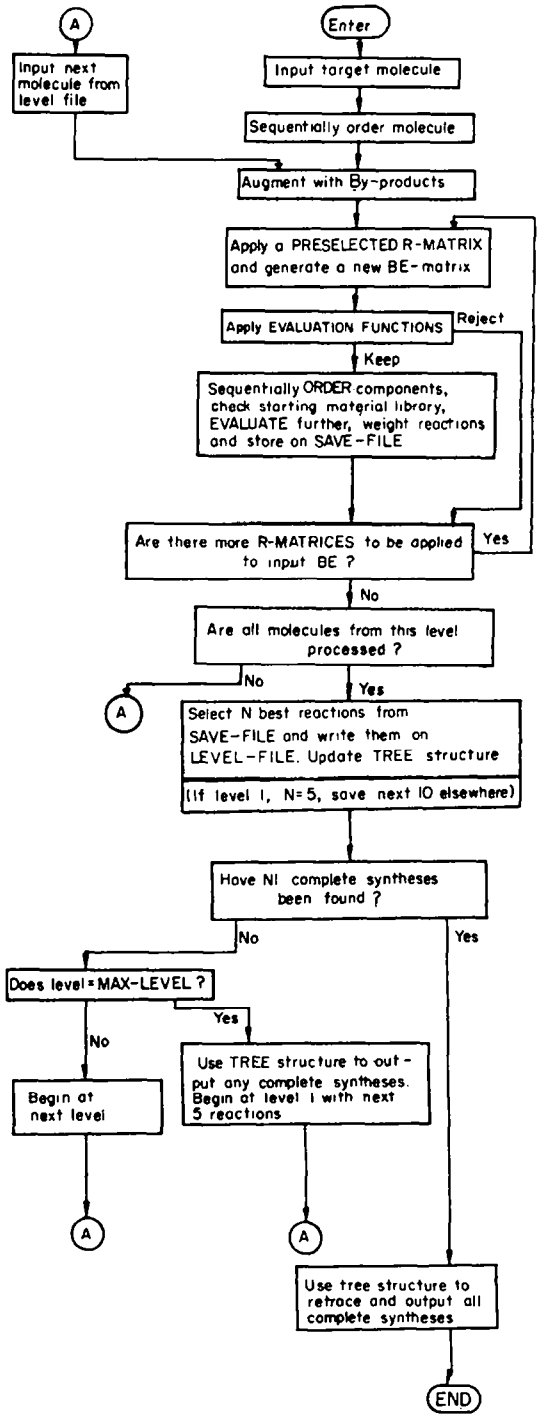


Fig 9. Function flow.

Acknowledgements—Hereby we gratefully acknowledge the partial support of this work by the National Science Foundation (USA), (Grant No. G. P. 28927X) Deutsche Forschungsgemeinschaft and Volkswagen-Stiftung e. V.

We also thank Ms. Sharon Cutri and Ms. Helga Hofmann for their aid in preparing the manuscript.

We appreciate helpful discussions with Prof. J. Dugundji and Dr. D. Marquarding. Further, the authors thank Dr. Karl Kompa for his help and advice and Richard Feldmann for providing his graphic input/output program for our use.

REFERENCES

[1] *Chemistry and Logical Structures*, Part 6.-Part 5: see Ref 2

[2] P. Gillespie, F. Ramirez, I. Ugi and D. Marquarding, *Angew. Chem.* **85**, 99 (1973); *Ibid.* Intern. Edit. **12**, 91 (1973)

[3] N. J. Nilsson, *Problem-Solving Methods in Artificial Intelligence.* McGraw-Hill., New York, N.Y. (1971) and refs cited

[4] E. J. Corey and W. T. Wipke, *Science* **166**, 178 (1969); E. J. Corey, *Quart. Rev.* **25**, 455 (1971); E. J. Corey, W. T. Wipke, R. D. Cramer and W. J. Howe, *J. Am. Chem. Soc.* **94**, 421, 431 (1972); E. J. Corey, R. D. Cramer and W. J. Howe, *Ibid.* **94**, 440 (1972); E. J. Corey and G. E. Petersson, *Ibid.* **94**, 460 (1972)

[5] W. T. Wipke, *Proceedings of the NATO/CNA ASI on Computer Representation and Manipulation of Chemical Information* (Edited by W. T. Wipke, S. Heller R. Feldmann, and E. Hyde) John Wiley, New York, N.Y. (1974)

[6] H. Gelernter, N. S. Sridharan, A. J. Hart, S. C. Yen, F. W. Fowler and H. J. Shue, *Topics Curr. Chem.* **41**, 113 (1973); H. Gelernter, *Angew. Chem.* in preparation

[7] J. B. Hendrickson, *J. Am Chem. Soc.* **93**, 6847, 6854 (1971)

[8] R. V. Stevens, personal communication

[9a] I. Ugi and P. Gillespie, *Angew. Chem.* **83**, 980 (1971); *Ibid.* Intern. Edit. **10**, 914 (1971); [b] I. Ugi and P. Gillespie, *Ibid.* **83**, 982 (1971); *Ibid* Intern. Edit. **10**, 915 (1971); [c] I. Ugi, P. D. Gillespie and C. Gillespie, *Trans. N.Y. Acad. Sci.* **34**, 416 (1972); [d] J. Dugundji and I. Ugi, *Topics Curr. Chem.* **39**, 19 (1973)

[10] J. Blair, J. Gasteiger, C. Gillespie, P. Gillespie and I. Ugi, *Proceedings of the NATO/CNA ASI on Computer Representation and Manipulation of Chemical Information* (Edited by W. T. Wipke, S. Heller, R. Feldmann and E. Hyde) Wiley, New York, N.Y. p. 129 (1974).

[11a] M. Gordon, C. E. Kendall and W. H. T. Davison, *Proceedings 11th Congress of Pure and Applied Chemistry* Vol. II/3, 115 (1947); [b] W. J. Wiswesser, *A Line-Formula Chemical Notation.* Crowell., New York, N.Y. (1954); [c] E. G. Smith, *The Wiswesser Line-Formula Chemical Notation.* McGraw-Hill, New York, N.Y. (1968); [d] W. J. Wiswesser, *J. Chem. Doc.* **8**, 146 (1968) and literature cited; [e] R. Fugmann, W. Braun and W. Vampel, *Angew. Chem.* **73**, 745 (1961); [f] H. Hilz, *J. Chem. Doc.* **4**, 173 (1964); [g] L. Spialter, *Ibid.* **4**, 261 (1964); [h] H. L. Morgan, *Ibid.* **5**, 107 (1965); [i] R. H. Penny, *Ibid.* **5**, 113 (1965); [j] E. Meyer, *Angew. Chem.* **82**, 605 (1970); [k] J. E. Dubois, *Proceedings of the NATO/CNA ASI on Computer Representation and Manipulation of Chemical Information* (Edited by W. T. Wipke, S. Heller, R. Feldmann, and E. Hyde) Wiley, New York, N.Y. (1973)

[12] F. Harary, *Graph Theory*, Addison-Wesley Reading, Mass. (1972)

[13] J. Blair, in preparation

[14] C. Gillespie, in preparation

[15a] T. Gasteiger, P. Gillespie, D. Marguarding and I. Ugi, *Topics Curr. Chem.* **48**, 1 (1974)

[15b] I. Ugi, D. Marquarding, H. Klusacek, G. Gokel and P. Gillespie, *Angew. Chem.* **82**, 741 (1970; *Ibid.* Intern. Edit. **9**, 703 (1970)

[16a] R. S. Cahn, C. K. Ingold and V. Prelog, *Experientia* **12**, 81 (1956); [b] R. S. Cahn, C. K. Ingold and V. Prelog, *Angew. Chem.* **78**, 413 (1966)

[17a] I. Ugi, *Z. Naturforsch.* **20B**, 405 (1965); [b] E. Ruch and I. Ugi, *Theoret. Chim. Acta Berlin*, **4**, 287 (1966); *Top. Stereochem.* **4**, 99, (1969); [c] I. Ugi, *Intra-Sceince Chem. Rep.* **5**, 229 (1971)

[18] J. E. Blackwood, C. L. Gladys, K. L. Loening, A. E. Petrarca and J. E. Rush, *J. Am. Chem. Soc.* **90**, 509 (1968)

[19] E. Ruch, *Accounts Chem. Res.* **5**, 49 (1972)

[20] V. Prelog, G. Helmchen, *Helv. Chim. Acta* **55**, 258 (1972)

[21a] M. G. Evans and M. Polanyi, *Trans. Farad. Soc.* **34**, 11 (1938); M. G. Evans and E. Warhurst, *Ibid.* **34**, 614 (1938); M. G. Evans, *Ibid.* **35**, 824 (1939); [b] K. E. Schuler, *J. Chem. Phys.* **21**, 624 (1953); [c] M. J. S. Dewar, *Tetrahedron* Suppl. to **8**, 75 (1966); M. J. S. Dewar, *Angew. Chem.* **83**, 859, *Ibid.* Intern. Edit., **10**, 761 (1971); [d] R. B. Woodward and R. Hoffmann, *Ibid.*, **81**, 797 (1969); *Ibid.* Intern. Edit. **8**, 781 (1969); [e] W. J. van der Hart, J. J. C. Mulder and L. J. Oosterhoff, *J. Am. Chem. Soc.* **94**, 5724 (1972); J. J. C. Mulder and L. J. Oosterhoff, *Chem. Commun.* **305**, 307 (1970); [f] K. Fukui, *Acc. Chem. Res.* **4**, 57 (1971); [g] N. T. Anh *Die Woodward-Hoffmann-Regeln und ihre Anwendung.* Verlag Chemie, Weinheim (1972); [h] T. L. Gilchrist and R. C. Storr *Organic Reactions and Orbital Symmetry.* Cambridge University Press, London (1972)

[22] I. Weinberg. *PL/1 Programming: A Manual of Style*, McGraw-Hill, New York, N.Y. (1971)

[23] R. J. Feldmann: *Interactive Graphic Chemical Structure Searching* in *Proceedings of the NATO/CNA ASI on Computer Representation and Manipulation of Chemical Information* (Edited by W. T. Wipke, S. Heller, R. Feldmann and E. Hyde, Wiley, New York, N.Y. (1974)