Methods for Computer Generation of Unique Configurational Descriptors for Stereoisomeric Square-Planar and Octahedral Complexes*

A. E. PETRARCA† and J. E. RUSH†

Chemical Abstracts Service, The Ohio State University, Columbus, Ohio   43210

Methods are described which would permit computer generation of unique configurational descriptors from nonunique unambiguous information supplied at input for (1) the three stereoisomeric configurations of square-planar complexes of the type Mabcd and (2) the 30 stereoisomeric configurations of octahedral complexes of the type Mabcdef. The input procedures and the general techniques are similar to those described previously for the treatment of asymmetric tetrahedral atoms and for stereoisomerism about a double bond. Inclusion of such stereochemical detail in computer structural records of inorganic compounds would increase the specificity and flexibility with which such information could be searched.

The development of techniques for translating chemical structural information into machine notation has been a subject of continuing interest in recent years. Such techniques are necessary for the manipulation of large files of structural formulas by computer methods or for conducting generic (substructure) searches by computer for all compounds with related structural characteristics.

In a recent paper[1] we described methods which would permit computer generation of unique configurational descriptions of asymmetric tetrahedral atoms, stereoisomerism about a double bond, and related systems from nonunique unambiguous information supplied at input. Details were also presented to show how the information required at input could be supplied clerically by topological coding from conventional stereochemical formulas and then manipulated for storage in computer files by extension of existing methods such as that described by Morgan.[2] The increased specificity of storage would then permit increased specificity of search by computer for information on substances having a desired set of structural characteristics.[3]

This paper describes methods for extension of the earlier techniques to the treatment of 4-coordinate square-planar and 6-coordinate octahedral stereogenic atoms. (A stereogenic atom is defined as an atom bearing several groups of such a nature that an interchange of two groups will produce a stereoisomer.[4])

TECHNIQUES FOR DESCRIPTION OF STEREOCHEMISTRY AND FOR DERIVATION OF UNIQUE CONFIGURATIONAL DESCRIPTORS

As in our earlier treatment, the basic approach is to describe the spatial arrangement of the groups attached to each stereogenic atom by means of appropriate stereodescriptors in accordance with the symmetry characteristics[5] of each particular system.[6] As illustrated for the reference structure in Figure 1, the spatial arrangement

of the peripheral atoms attached to a 4-coordinate square-planar atom can be described unambiguously in terms of a set of descriptors **A**, **B**, **C**, and **D**. The treatment for stereogenic atoms of this type is best illustrated on an atom with four different groups attached because three different stereoisomeric configurations are possible. If the peripheral atoms are ranked numerically according to a prescribed order of hierarchy, one possible description for each configuration of such a hypothetical stereogenic atom, Q, is illustrated in Figure 2. (The ordering rules we prescribe for the treatment described herein are those used
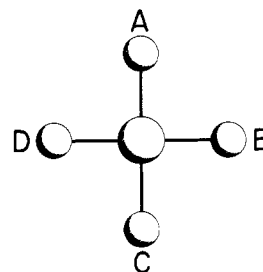

Figure 1. Reference structure for assignment of stereodescriptors to peripheral atoms attached to a stereogenic 4-coordinate square-planar atom
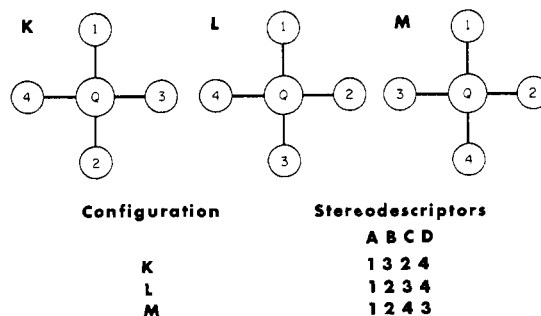

Configuration	Stereodescriptors A B C D
K	1 3 2 4
L	1 2 3 4
M	1 2 4 3

Figure 2. Descriptions of the three configurations for a hypothetical 4-coordinate square-planar stereogenic atom of the type Mabcd

for unique numbering of the atoms in a computer representation such as that described by Morgan.[52] However, other ordering rules, such as those described by Cahn et al.,[7] could also be utilized for this purpose.) As can be seen from the diagrams in Figure 2, the three configurations (arbitrarily designated here by the letters **K, L,** and **M**) are distinguishable in terms of the groups opposite each other on the two axes in the plane. The numerical stereochemical descriptions shown beneath the diagrams, however, are not unique because there are eight possible permutations for the description of each configuration in terms of the reference structure (Figure 1). The eight permutations for configuration **K** are illustrated in Figure 3. All of these permutations arise because one can start the description with any one of the four peripheral atoms while viewing the system from either side of the plane.

The technique for showing configurational identities among the various permutations and for differentiating the three configurations from each other is illustrated for two permutations of configuration **K** in Figure 4 and for one permutation each of configurations **L** and **M** in Figure 5. It involves four steps as follows:

> **Step One.** Transpose the atom numbers in the **B** and **C** descriptor columns. This is equivalent to subdividing the four coordinating atoms into two pairs of opposites which are now contained in an **AB** subset and a **CD** subset.
>
> **Step Two.** Reorder the atom numbers within each of the subsets so that the numbers in each subset are in natural order.
>
> **Step Three.** Reorder the two subsets of atom numbers so that they are in natural order with respect to the first number of each subset.
>
> **Step Four.** Configurational identities and differentation can now be established on the basis of the number of inversions contained in the entire set. (Each integer which precedes another smaller integer in the set constitutes an inversion. Computationally, the number of inversions is also equal to the minimum number of adjacent transpositions required to arrange the numbers in natural order.) Each of the three configurations is identified as follows: (a) Configuration **K** has 0 inversions; (b) Configuration **L** has 1 inversion; and (c) Configuration **M** has 2 inversions.

Note in Figure 4 that the outcome of the last step is the same for each permutation of a given configuration. Differentation between two of the three configurations is illustrated in Figure 5.

The technique for handling 6-coordinate octahedral stereochemistry is slightly more involved. As illustrated for the reference structure in Figure 6, the spatial arrangement of the coordinating atoms attached to the central atom must be described in terms of a set of six descriptors **A, B, C, D, E,** and **F.** The treatment is best illustrated on an atom with six different groups attached, for which 30 different stereoisomeric configurations (15 enantiomeric pairs) are possible. Figure 7 illustrates how two enantiomeric configurations may be described by the proposed treatment after the coordinating atoms have been ranked according to the prescribed order of hierarchy. These two configurations have been given the designations **R-2** and **S-2**, respectively, to distinguish them from the other 28 configurations described elsewhere in this treatment. The numerical part of these designations relates to the fact that both enantiomers have the same pairs

Configuration K

Permutation		Stereodescriptors
		A B C D
(a)		1 3 2 4
(b)		3 2 4 1
(c)		2 4 1 3
(d)		4 1 3 2
(e)		1 4 2 3
(f)		4 2 3 1
(g)		2 3 1 4
(h)		3 1 4 2

Figure 3. The eight permutations for description of configuration **K**
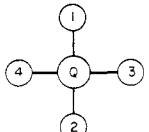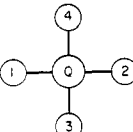
Configuration K

	A B C D	A B C D
Stereodescriptors		
Input Sequence	1 3 2 4	4 2 3 1
Step 1	12 34	43 21
Step 2	12 34	34 12
Step 3	12 34	12 34
No. of Inversions	0	0

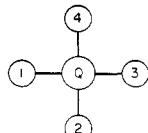Figure 4. Establishing the identity of two permutations for configuration **K**
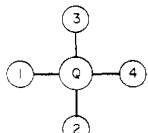
Configuration L   Configuration M

	A B C D	A B C D
Stereodescriptors		
Input Sequence	4 3 2 1	3 4 2 1
Step 1	42 31	32 41
Step 2	24 13	23 14
Step 3	13 24	14 23
No. Of Inversions	1	2

Figure 5. Differentiation of configurations **L** and **M** from each other
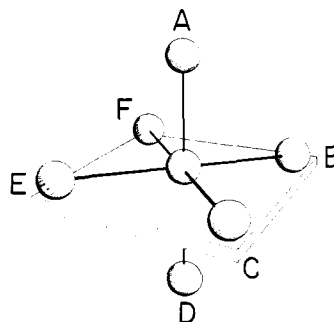
Figure 6. Reference structure for assigning stereodescriptors to peripheral atoms

R-2   **S-2**

Configuration   **Stereodescriptors**

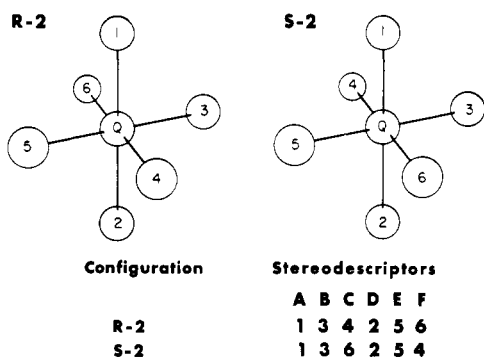|	A	B	C	D	E	F
R-2	1	3	4	2	5	6
S-2	1	3	6	2	5	4

Figure 7. Descriptions for an enantiomeric pair of hypothetical stereogenic octahedral atoms of the type Mabcdef

R-1   **R-2**   **R-3**

Configuration   **Stereodescriptors**

	A	B	C	D	E	F
R-1	1	3	5	2	4	6
R-2	1	3	4	2	5	6
R-3	1	3	4	2	6	5

Figure 8. Descriptions for the three R configurations having atoms 1-2 on the principal axis

of opposites on the three axes, while the **R** and **S** portion relates to differences in chirality in terms of the symbolism of Cahn *et al.*[7] The chirality in this instance relates to the direction of numbering of the atoms in the plane perpendicular to the principal axis i.e., the axis containing atom number one. When the atoms in this plane are viewed from atom number one, the numbers are arranged in *clockwise* order for the **R** chirality and in *anticlockwise* order for the **S** chirality.

To provide a brief glimpse of what some of the other configurations and their descriptions look like, Figure 8 shows three configurations of **R** chirality which have atoms 1 and 2 opposite each other on the principal axis, but which differ with respect to the atoms opposite each other in the plane perpendicular to the principal axis. The **R** chirality relates to the fact that when the atoms in this plane are viewed from atom number 1, the lowest-numbered atom (No. 3) and the lower-numbered atom (of the two) adjacent to it in the plane are arranged in clockwise order. The four configurations of **R** chirality shown in Figure 9 are representative of the remaining configurations in that the atom numbers opposite each other on the principal axis are 1-3, 1-4, 1-5, and 1-6, respectively.

Because of the symmetry characteristics of the octahedral system there are 6! (or 720) permutations for the numerical descriptions of all 30 configurations in terms of the reference model (Figure 6)—24 permutations for each configuration. Figure 10 shows four permutations for configuration **R-15** which are derived from orientations of the structure wherein atom 1 coincides with position **A** of the reference structure. Successive 90° rotations on the 1-6 axis account for the variations in each of the permutations shown. The remaining 20 permutations for configuration **R-15** can be derived in similar fashion from orientations in which each of the other peripheral atoms are in position **A** of the reference structure.

The technique for showing configurational identities among the various permutations and for differentiating the 30 configurations from each other is illustrated for two permutations of configuration **R-15** in Figure 11 and for one permutation each of several other configurations in Figures 12 and 13. The technique involves six steps as follows:
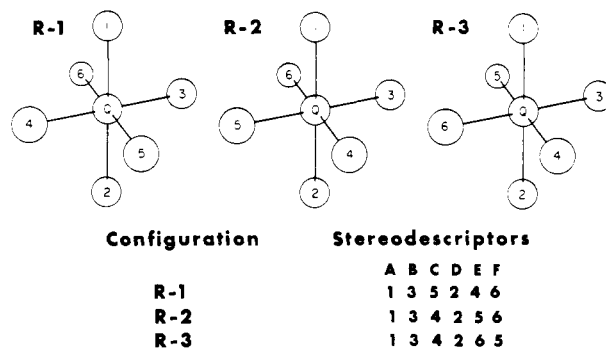
R-4   **R-7**

R-10   **R-13**

Configuration   **Stereodescriptors**

	A	B	C	D	E	F
R-4	1	2	5	3	4	6
R-7	1	2	5	4	3	6
R-10	1	2	4	5	3	6
R-13	1	2	4	6	3	5
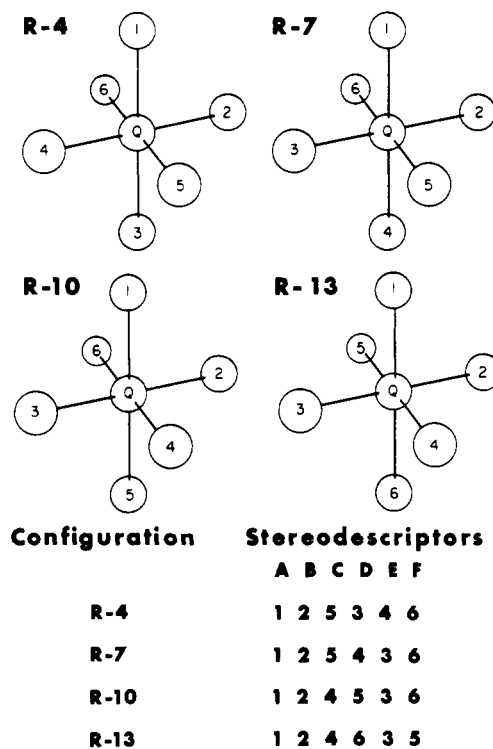
Figure 9. Descriptions for some R configurations having atoms other than 1-2 on the principal axis

Configuration R-15

(a)   (b)

(c)
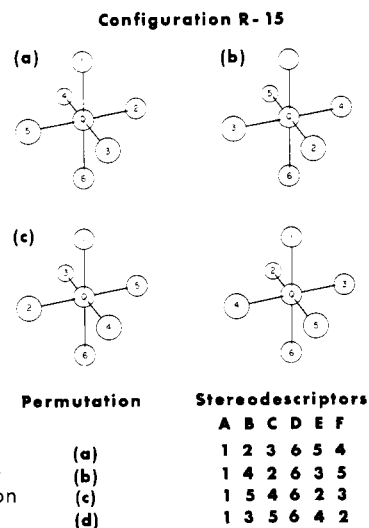
Figure 10. Four of the 24 permutations for description of configuration R-15

Permutation   **Stereodescriptors**

	A	B	C	D	E	F
(a)	1	2	3	6	5	4
(b)	1	4	2	6	3	5
(c)	1	5	4	6	2	3
(d)	1	3	5	6	4	2
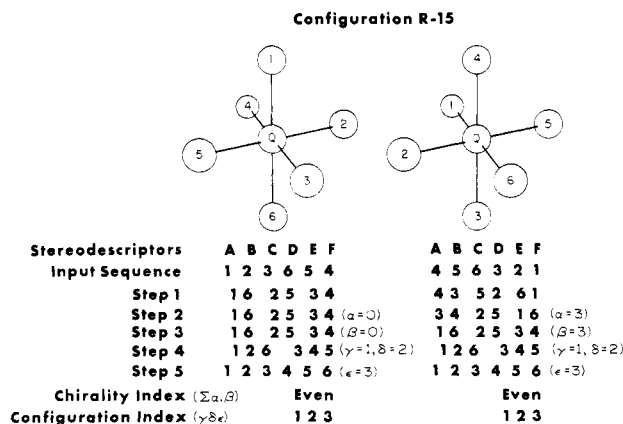
Figure 11. Establishing the identity of different permutations for a given configuration

Figure 12. Differentiation of two diastereomeric configurations from each other

Step One. Rearrange the atom numbers so that the three pairs of opposites originally in the **AD**, **BE**, and **CE** subsets are now located, respectively, in the **AB**, **CD**, and **EF** subsets.

Step Two. Reorder the atom numbers within each of the three subsets so that the numbers within each subset are in natural order. The total number of transpositions required for this operation, α, is then recorded.

Step Three. Reorder the three subsets of atom numbers so that they are in natural order with regard to the first number of each subset. The minimum number of adjacent transpositions required for this operation, β, is also recorded.

Step Four. Now subdivide the atom numbers into an **ABC** subset and a **DEF** subset and reorder the atom numbers in each subset so that they are in natural order. The minimum number of adjacent transpositions for the **ABC** subset, γ, and the minimum number for the **DEF** subset, δ, are then recorded.

Step Five. Reorder the atom numbers in the entire set so that they are in natural order, and record the minimum number of adjacent transpositions required for this operation, ϵ.

Step Six. Configurational identities and differentiation can now be established on the basis of a configuration index and a chirality index. (The configuration index is a function of the pairs of opposites on each of the three axes for a given configuration, while the chirality index relates to the chirality differences between enantiomeric configurations.) The configuration index is a three-digit number based on the value of $\gamma\delta\epsilon$; configuration indexes corresponding to the numerical parts of the designations used throughout this paper for the 15 enantiomeric paris are listed in Table I. The chirality index is the parity of the sum of α and β; an even sum indicates **R** chirality, while an odd sum indicates **S**.

Figure 13. Differentiation of enantiomeric configurations from each other

Table I. Configuration Indexes and Corresponding Numerical Designations Used for Defining Specific Pairs of Opposites in Octahedral Complexes

Config. Index	Numerical Designation	Config. Index	Numerical Designation	Config. Index	Numerical Designation
000	1	120	6	112	11
010	2	101	7	122	12
020	3	111	8	103	13
100	4	121	9	113	14
110	5	102	10	123	15

As seen in Figure 11, the outcome of the last step is the same for different permutations of a given configuration. Figure 12 illustrates the differentiation between a pair of diastereomers, while Figure 13 illustrates the differentiation between a pair of enantiomers.

A slight variation in the treatment just described for both 4-coordinate square-planar and 6-coordinate octahedral systems is necessary for stereogenic atoms having two or more identical groups attached because degeneracy reduces the possible number of configurations for each system. Details concerning this aspect of the treatment are being more fully explored.

INPUT PROCEDURES

The operations just described are to be performed within the computer after the necessary information has been
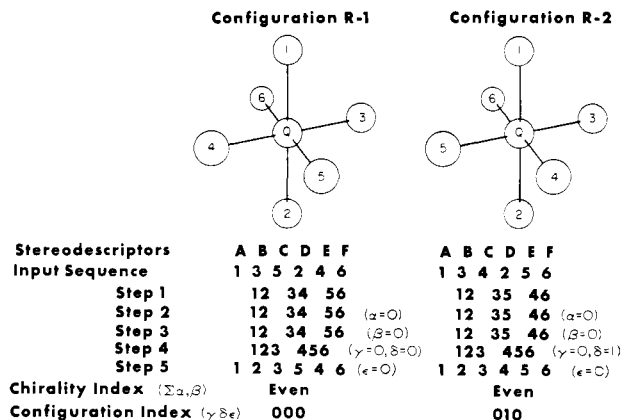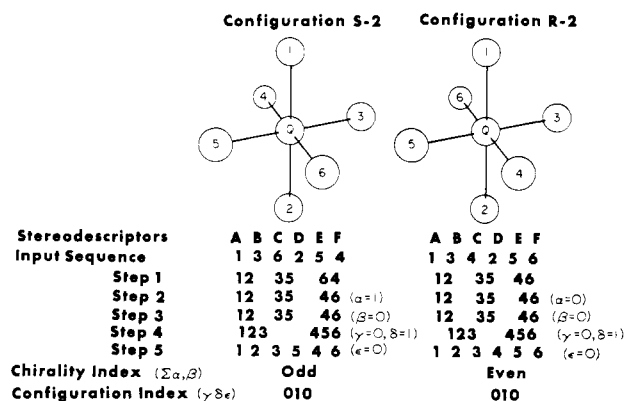
supplied by input procedures similar to those described in our earlier paper.[1] (For example, both 4-coordinate square-planar and 6-coordinate octahedral stereochemistry can be described by modified versions of the clockwise sequence rule.) Figure 14 illustrates the type of unambiguous, nonunique atom-bond connection table which would be prepared at input for a specific stereoisomer of a 4-coordinate square-planar complex.

DERIVATION OF UNIQUE DESCRIPTORS FOR SPECIFIC STEREOISOMERS

The type of unique computer representation which would be generated from the information in Figure 14, by a modification of Morgan's technique,[2] is illustrated in Figure 15. (Detailed descriptions of the component parts of the connection tables shown in Figures 14 and

**Name:** trans-Bromonitrochloro-ammine platinate ion

M.F. $BrClH_3N_2O_2$

Stereo: Absolute

Sign of Rotation: O
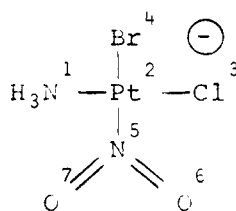
| Atom No.	Element	Group	Bond 1	Attachment 1	Bond 2	Attachment 2	Bond 3	Attachment 3	Bond 4	Attachment 4	Charge	Abnormal Valence	H Count	A	B	C	D
1	N	1	2										3				
2	Pt	1	1	1	3	1	4	1	5		-1			4	3	5	1
3	Cl	1	2														
4	Br	1	2														
5	N	1	2	2	6	2	7										
6	O	2	5														
7	O	2	5														

Figure 14. Unambiguous, nonunique atom bond connection table prepared at input for a stereoisomer of a 4-coordinate square-planar complex
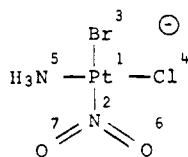

```
ATOM NO.:    1   2   3   4   5   6   7
ELEMENT:    Pt   N  Br  Cl   N   O   O
BOND FROM:       1   1   1   1   2   2
BOND TYPE:       1   1   1   1   2   2

CHARGE:     1 (-1)
STEREO:     Absolute
ROTATION:   0
CONFIG.:    1 (K)
```

Figure 15. Unique computer structural representation for the anion illustrated in Figure 14

$$Br - X - NO_2$$
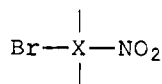
X is a 4-coordinate square-planar stereogenic atom

Figure 16. Example of a stereospecific substructure which could be searched generically by the proposed treatment

15 are provided in our earlier paper[1] and in references cited therein.) Note that the configurational descriptor for the stereogenic atom in this particular stereoisomer is **K**. The computer representations for the other two stereoisomers would differ only in the configurational descriptors for the stereogenic atom—i.e., the descriptor would be **L** or **M**—since the constitution would be the same in each case.

In similar fashion each of the various stereoisomers of an octahedral complex (both optical and nonoptical) would have its own unique computer structural representation.

GENERAL APPLICABILITY AND UTILITY OF THE METHOD

As suggested earlier, the proposed treatment would permit increased specificity of search by computer for information on substances having a desired set of structural characteristics. For example, if one were interested in searching for 4-coordinate complexes with a bromine atom and a nitro group trans to each other only those structures satisfying these stereochemical requirements—e.g., the structure in Figure 15—would be retrieved. Figure 16 illustrates how such a search request might be phrased.

In conclusion, we have described the essential details of a method which permits computer generation of unique configurational descriptors for stereogenic square planar and octahedral complexes. (A system of specification, based on similar principles, which is more readily adaptable for use in nomenclature has also been developed. This system will be described elsewhere.[5]) We have also indicated how the increased specificity of storage would permit increased specificity of search for stereochemical detail. With suitable modifications the techniques described herein can be extended to the treatment of other stereochemical systems such as the 5-coordinate trigonal bipyramid, 5-coordinate tetragonal pyramid, 6-coordinate trigonal prism, and 8-coordinate tetragonal antiprism.

ACKNOWLEDGMENT

LITERATURE CITED

(1) Petrarca, A. E., M. F. Lynch, and J. E. Rush, "A Method for Generating Unique Computer Structural Representations of Stereoisomers," J. CHEM. Doc. 7, 154 (1967).

(2) Morgan, H. L., "The Generation of a Unique Machine Description for Chemical Structures—A Technique Developed at Chemical Abstracts Service," J. CHEM. Doc. 7, 107 (1965).

(3) Cossum, W. E., M. L. Krakiwsky, and M. F. Lynch, "Advances in Automatic Chemical Substructure Searching Techniques," J. CHEM. Doc. 5, 33 (1965); Sussenguth, E. H., Jr., "A Graph Theoretic Algorithm for Matching Chemical Structures," J. CHEM. Doc. 5, 36 (1965); Hoffman, W. S., "An Integrated Chemical Structure Storage and Search System Operating at Du Pont," J. CHEM. Doc. 8, 3 (1968).

(4) See McCasland, G. E., "A New General Method for the Naming of Stereoisomers," proposal distributed by the Nomenclature Committee, Division of Organic Chemistry, American Chemical Society, 1953. (Copies are available from Chemical Abstracts Service, Columbus, Ohio.)

36

(5) Jaffe, H. H., and M. Orchin, "Symmetry in Chemistry," Wiley, New York, 1965.

(6) A similar approach was used by McDonnell, P. M., and R. F. Pasternack, "A Line-Formula Notation System for Chemical Compounds," J. CHEM. DOC. 5, 56 (1965), but our subsequent treatment differs from theirs. See also Pasternack, R. F., and P. M. McDonnell, "Designation of Ligand Positions in Coordination Complexes," Inorg. Chem. 4, 600 (1965).

(7) Cahn, R. S., C. K. Ingold, and V. Prelog, "Specification of Molecular Chirality," Angew. Chem. Intern. Ed. Engl. 5, 385 (1966).

(8) Petrarca, A. E., J. E. Rush, and M. F. Brown, "A Nomenclature-Independent Method for Specification of the Stereochemistry of Coordination Compounds," Abstracts of Papers, INOR 147, 156th Meeting, ACS, Atlantic City, September 1968. (Also to be submitted for publication in J. CHEM. DOC.).

Storage and Retrieval of Agricultural Screening Data*

J. B. HAGLIND, H. J. ACKERMANN, R. E. MAIZELL,
T. M. MANNING, and B. S. SCHLESSINGER†
Technical Information Services, Olin Mathieson Chemical Corp.,
275 Winchester Ave., New Haven, Conn. 06504

A system is described for the storage and retrieval of agricultural screening data. The system embraces a formatted means of entering laboratory results in notebooks, keypunching the data onto cards, and processing the card file onto a disk for random access of data. A FORTRAN program is used to search the disk file for specific compounds and tests. Compound structure searching is achieved by using a permuted Wiswesser Line Notation file.

For several years prior to the development of the system to be described, biological test data determined by our Agricultural Research group had been recorded in conventional laboratory notebooks and keypunched onto cards for storage and retrieval. The cards were manipulated by standard unit record equipment and information printed from them by programs written for and operated on a 1401 IBM computer. This system was effective and satisfactory as a start. However, as the files of biotest data grew and secondary screening data rose in relative importance, a new approach was clearly needed (1) to store and retrieve not only current data, but also the large volume of previous data; (2) to reduce machine processing time and costs; and (3) to give greater flexibility and capability. We proceeded to try to meet these needs in closest possible cooperation with laboratory personnel.

ANALYSIS

Availability of an IBM 1130 and 1800 at the Olin Research Center in New Haven prompted development of a system suited for these computers and also adaptable, if necessary, to an IBM 360 scheduled for later installation.

It was decided that further processing of earlier results, already available on punched cards and in quarterly print-outs, could await development of the system described here. The older files would then be integrated into the ongoing system as time and money permitted. Data could, of course, still be retrieved from older files by visual inspection of printouts or by sorting of the punched cards.

To simplify the system, it was agreed to use brief codes to indicate test results, rather than full descriptions. The input format was designed to contain the Olin Compound Registry Number, date of compound entry into the Registry, and 52 specific results for seven different test classifications (Figure 1).

DEVELOPMENT

The most important data, initially, were all primary screening information recorded in notebooks after a specified date. Since these data had been recorded in a relatively uniform manner in notebooks, selection of specific results for storage in the data retrieval system was simplified.

After deciding what information was to be stored, an input format was designed. A template was made so as to be superimposed on a notebook page to permit more rapid location of specific input for the system. Clerks manually transcribed data from notebooks to input forms for keypunch operators.

To eliminate the transcription step for current data, a unique notebook was designed to permit scientists to enter data in card column format in areas marked for rapid location by keypunch operators. Adequate space was given for additional comments or remarks (Figure 2).