A Notation System for Indexing Pesticides*

By HERMAN SKOLNIK and AVIS CLOW
Hercules Powder Company, Research Center, Wilmington, Delaware
Received December 12, 1963

Although there are relatively few toxophoric groups of importance in the literature of pesticides, the number of individual compounds in each toxophoric class is large. For example, there are more than 2000 U. S. patents describing many thousands of organic phosphorus compounds as pesticides. The January–June, 1962, *Chemical Abstracts Subject Index* required four pages (a total of 12 columns) to list the phosphorothioate and phosphorodithioate entries, most of which pertain to the preparation, composition, and uses of the compounds as pesticides. Other toxophoric groups, such as halogen, amide, urea, carbamate, etc., similarly have thousands of members which have been studied for their pesticidal activity.

Scientists involved in pesticide research and development programs require an information system that allows both retrieval (generically and specifically) and correlation, *e.g.*, chemical structure with pesticidal activity. Consequently, an index of chemicals of interest as pesticides must be designed to express the subject content of chemical structures from the following viewpoints:

1. The whole structure or essentially the whole structure
2. The toxophoric group or groups
3. Other functional groups or moieties
4. Relationships between parts of the structures within a toxophoric class
5. Relationships between toxophoric classes

Subject indexing by means of chemical nomenclature is a reasonably good mechanism for retrieval of specific compounds; it is a completely unsuitable mechanism for retrieval from the other viewpoints. This paper describes a pesticide information system which is suitable for retrieval from all five viewpoints.

With the increasing use, over the past fifteen years or so, of accounting machines and computers as documentation tools, there has been a corresponding activity in the design of machine codes for the retrieval and correlation of chemical structures. These codes consume from one-third to a full IBM card and allot a fixed position to designate a functional group or moiety or combinations of the two. They do not provide for or allow the printing of a list of compounds, and their only advantage is machine sorting, which usually relates the compounds by number or code to a structure card file.[5-10,13]

These limitations of chemical nomenclature and machine codes led us to investigate the use of notation systems as a basis for indexing chemicals of interest as pesticides. Notation systems, such as the Dyson,[4] Wiswesser,[14] Silk,[11] Duffin,[3] and others,[1,2] use a system of symbols—numerals, letters, and marks—to represent the structures of molecules. Although none matches the structural formula as a communication mechanism, most notation systems are successful in expressing structures concisely, linearly, and without ambiguity.

Because inventors of notation systems have emphasized the communications aspect, their rules and procedures have been directed towards maintaining a hard and fast consistency for writing a cipher to give one and only one correct sequence of symbols for each chemical compound. The Dyson system, for example, requires that the cipher be started with the longest chain or the most complex moiety of the molecule. The Wiswesser system requires that the cipher be started with one end of the molecule. Another constraint in the use of many existing notation systems is the absence of certain required symbols and characters in accounting machines and computers. Thus, the IBM 1401 Printer available to us does not have lower case letters, subscripts, superscripts, colon, semicolon, parentheses, equal sign, or question mark.

From the pesticide viewpoint, the most logical notation system is one which allows us the flexibility of starting the cipher with toxophoric groups or with any functional group or moiety considered pertinent. This requirement is important for ease of sorting between and within classes and for machine print-outs by desired chemical sequences or relationships. Furthermore, if machines are to be used, notation symbols and characters must be compatible with the machines available.

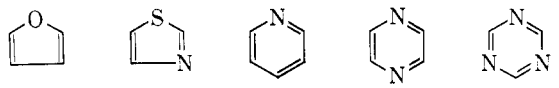
Basic Symbols.—The notation system we developed for the indexing of pesticides has eclectic features of the Dyson, Silk, and Wiswesser systems plus several new features. The basic symbols used are as follows.

A = alkyl; followed by a numeral, the number of carbons in an alkyl chain, *e.g.*, A1 = methyl, A2 = ethyl, A3 = propyl, etc. We decided to use A instead of C so that machine sorting would arrange alkyls before aromatics.

B = aromatic ring systems; followed by a numeral, the number of carbon atoms in the aromatic nucleus, *e.g.*, B6 = benzene, B66 = naphthalene, etc.

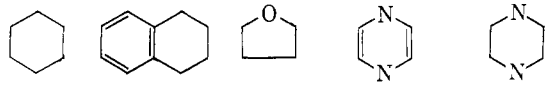
Z = heterocyclic ring systems. The number following Z indicates the ring size and the symbol following the number indicates the hetero atom. Examples are

* Presented before the Divisions of Chemical Literature and Agricultural and Food Chemistry, Pesticides Subdivision, Joint Symposium on "Problems of the Pesticide Literature and Some Solutions," 147th National Meeting of the American Chemical Society, Philadelphia, Pa., April 9, 1964.



Z5O Z5SN-3† Z6N Z6N2-4† Z6N3-3,5†

BH and ZH = partially or fully saturated aromatic and heterocyclic rings. Examples are



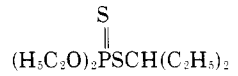
B6H6 B66H4 Z5QH4 Z6N2-4H2 Z6N2-4H6

E = unsaturation (double bond)
U = triple bond
K = carbonyl
L = halogen
Q = oxygen, hydroxyl, or oxide
N = nitrogen or -NH_2
P = phosphorus
M = metal
T = *t*-butyl
V = vinyl
X = unspecified substituent
Y = isopropyl

Basic Rules.—The cipher is started from any functional group or moiety in the chemical structure. Thus, ciphers for ethers are written as follows.

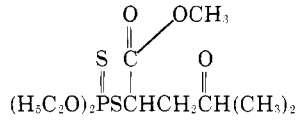
	As ethers	As alkyls	As others
CH_3OCH_3	Q.A1/2	A1QA1	
$C_2H_5OC_2H_5$	Q.A2/2	A2QA2	
$CH_3OC_2H_5$	Q.A1.A2	A1QA2	
$C_6H_5OCH_3$	Q.A1.B6	A1QB6	B6QA1

The period is used to separate parts of the cipher and to relate each part to the atom it is joined to. The stroke, on the other hand, is used to show how many units there are of a given atom or group, or to show attachments within the part, as in the following example.



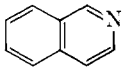
PES.SA1/A2/2.QA2/2

In structures having complex substituents, commas are used as shown below.



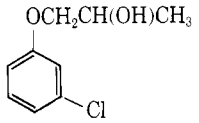
PES.SA1/KQA1/A1KN,A1,2.QA2/2 (from the phosphorus viewpoint)
KQA1.A1/SP,ES,QA2,2/A1KN,A1,2 (from the ester viewpoint)
KN/A1/2.A2/KQA1/SP,ES,QA2,2 (from the amide viewpoint)

† The numbering of the ring is in accordance with the Revised Ring Index. Locant numbers are given for all hetero atoms except the first, which is always understood to be at position 1. When the hetero atom is not in position 1, then the locant is given, *e.g.*



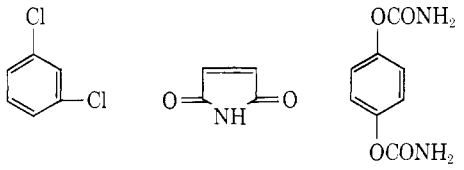
Z66N-2

Further clarification of the ciphers is obtained by separating ring substituents with hyphens, when necessary, as in the following example.



Q.A2/Q/A1.B6L (from the ether viewpoint)
B6.QA2/Q/A1.L (from the benzene viewpoint)
QA1/A1/A1QB6-L (from the alcohol viewpoint)
LB6-QA2/Q/A1 (from the halogen viewpoint)

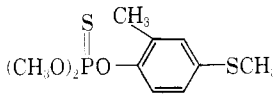
Positions of substituents on rings are not designated, as in the following.



B6L2 Z5NH2.EQ/2 B6.QKN/2
L2B6 EQ/2.Z5NH2 NKQ/2.B6
KN.QB6QKN

In an extremely large file, of course, positions of substituents would need to be designated. Within a given toxophoric class of pesticides, however, we find that position designation is an unnecessary refinement.

Ciphers are written in the order of preferred toxophoric group, followed by that portion of the molecule containing other toxophoric groups of moieties of interest. For example



PES.QB6-SA1-A1.QA1/2

The same order of citation is followed when a second index entry is made for the same compound.

S.B6-QP/ES/QA1/2-A1.A1

If it is important that a compound be considered from more than one viewpoint, as many entries as are necessary may be made in the index. Thus, the above compound may have as many as seven entries in the index.

PES.QB6-SA1-A1.QA1/2
PES.QA1/2.QB6-SA1-A1
S.B6-QP/ES/QA1/2-A1.A1
S.A1.B6-QP/ES/QA1/2-A1
B6.QP/ES/QA1/2.SA1.A1
B6.SA1.QP/ES/QA1/2.A1
B6.A1.QP/ES/QA1/2.SA1

Indexing Phosphorus Pesticides.—Phosphorus derivatives, being one of the broadest and most active in pesticide research, were chosen for our study of indexing by means of notation systems. The phosphorus compounds were selected (approximately every tenth compound) from the January–June, 1962, *Chemical Abstracts Subject Index*. Table I lists four of these compounds, in the same order as taken from *Chemical Abstracts*, with the structure, Hercules cipher, and machine-sorting order. Table II compares the same information for the Dyson and Wiswesser ciphers.

The *Chemical Abstracts* index entries group these compounds together conveniently under two major classes:

Table I

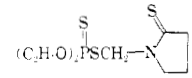
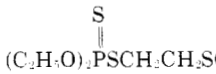
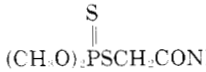
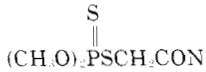
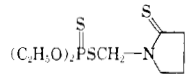
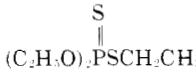
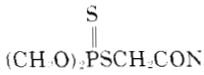
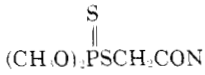
Structure (*Chem. Abstr. Index* order)	Hercules notation	Machine sorting order
	PES.SA1Z5NH4-ES.QA2/2	(3)
	PES.SA2SA1.QA2/2	(4)
	PES.SA1KN/A1V.QA1/2	(2)
	PES.SA1KN/A1/2.QA1/2	(1)

Table II

Structure (*Chem. Abstr. Index* order)	Dyson notation	Machine sorting order	Wiswesser notation	Machine sorting order
	A5ZN1ES2:CS/PES(QC$_2$)$_2$	(1)	T5NYTJ—A1SPS&O2&O2	(1)
	C$_2$:Q/$_2$PES;S/2C$_2$SC	(3)	2OPS&O2&S2S1	(4)
	C$_3$E:N/EQC$_2$:2S/PES(QC)$_2$	(4)	1U2MV1SPS&O1&O1	(3)
	C:Q/$_2$PES;S/2C$_2$EQ;N(C)$_2$	(2)	1OPS&O1&S1VN1&1	(2)

phosphorodithioic acid and phosphorothioic acid. The subclasses, however, are at the mercy of the alphabet. The Hercules notation system beginning, by definition, with toxophoric group arranges these compounds by three classes.

1. P(=O)(SR)(OR′)(OR″)
2. P(=S)(OR)(OR′)(OR″)
2. P(=S)(SR)(OR′)(OR″)

Within each class the arrangement is from simple to complex, or *vice versa* if desired, by (OR) or (SR), and (OR′) and (OR″). This is illustrated by Table III. The Dyson and Wiswesser systems, on the other hand, submerge the toxophoric group and are not amenable to a straightforward machine sort by toxophoric groups or by logical order within a toxophoric group.

Permutation or rotation of the Dyson and Wiswesser ciphers around the phosphorus atom,[12] shown in Tables IV and V, also failed to arrange the structures in a strictly toxophoric order. Permutation, moreover, consumes twice as many columns, and thus requires practically a full card for the permuted cipher. A major disadvantage is that the permuted cipher must be sorted twice: to the right of the phosphorus and to the left of the phosphorus.

Table VI compares the machine sorting orders of the Hercules ciphers with the ciphers and permuted ciphers of the Dyson and Wiswesser notations.

Cipher Detail and Depth of Indexing.—This paper purposely does not delineate a complete notation system. We do not think that a notation system for indexing can be perfectly fixed. Rather, the notation system must be responsive to the needs of the scientists who will be using the index. These needs determine the degree of detail in the cipher and the depth of indexing. A compound thus may be indexed as illustrated in the following.

PES.SA1KN/A1V.QA1/2
A1Q/2/P,ES,SA1KNA1V
A1KN/A1V.SP/ES/QA1/2
KN/A1V.A1SP/ES/QA1/2
VA1NKA1SP/ES/QA1/2
PES.QA1/2.SA1KN/A1V (an alternative of the first)

These five index entries allow the retrieval of the chemical from the phosphorus, methoxy, acetamide, amide, and vinyl viewpoints, and within each of these viewpoints from the interrelationships of the other viewpoints. A particular information file and the needs of the users may indicate a depth of indexing for the compound above from only the phosphorus viewpoint, or from the phosphorus and amide viewpoints. The degree of detail possible for a cipher is illustrated in the following compound.

Table III
Machine Sorting and Print-Out of Phosphorus Pesticides by the Hercules Notation

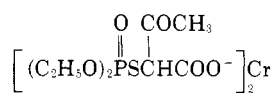
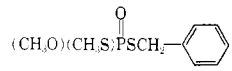
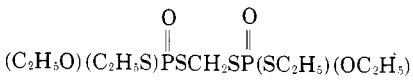
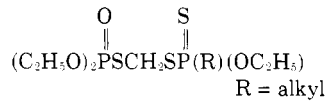
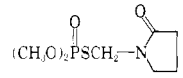
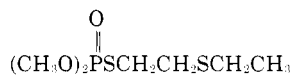
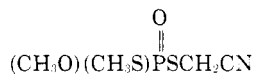
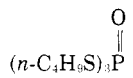
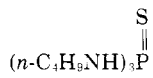
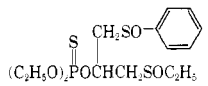
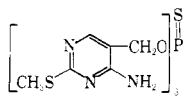
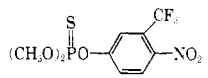
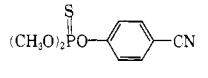
Structures	Phosphorus index entry and print-out order	Other index entries
	PEQ.SA1/KA1/KQ.QA2/2//2.MCR	K.A1KQ/SP,EQ,QA2,2.A1.2/MCR KQ.A1/KA1/SP,EQ,QA2,2.2/MCR MCR,1/2.QKA1/KA1/SP,EQ,QA2,2
	PEQ.SA1B6.SA1.QA1	
	PEQ.SA1SP/EQ/SA2/QA2.SA2.QA2	
 R = alkyl	PEQ.SA1SP/ES/A/QA2.QA2/2 also PES.A.SA1SP/EQ/QA2/2.QA2	
	PEQ.SA1Z5NH4-EQ.QA1/2	Z5NH4.EQ.A1SP/EQ/QA1/2
	PEQ.SA2SA2.QA1/2	S.A2SP/EQ/QA1/2.A2
	PEQ.SA2UN.SA1.QA1	NUA2SP/EQ/SA1/QA1
	PEQ.SA4/3	
	PES.N,A4/3	
	PES.QA1/A1SA1/2.QA2/2	S.A1/A1SA1/QP,ES,QA2,2.A1 S.A1.A1/2/A1QP,ES,QA2,2
	PES.QA1/A1S,EQ,B6/A1S,EQ,A2.QA2/2	S/EQ.A2/A1S,EQ,B6/QP,ES,QA2,2.A2 S/EQ.A2/A1S,EQ,A2/QP,ES,QA2,2.B6
	PES.QA1Z6N2-3,N,SA1/3	Z6N2-3.N.SA1.A1Q-//3,PES
	PES.QB6-NQ2-A1L3.QA1/2	Q2N/B6-A1L3-QP/ES/QA1/2
	PES.QB6-A1UN.QA1/2	NUA1/B6-QP/ES/QA1/2

Table III (*Continued*)
Machine Sorting and Print-Out of Phosphorus Pesticides by the Hercules Notation

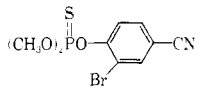
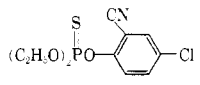
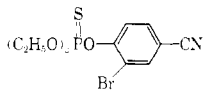
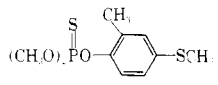
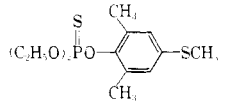
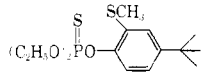
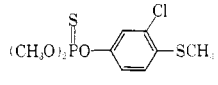
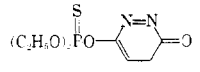
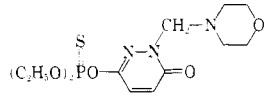
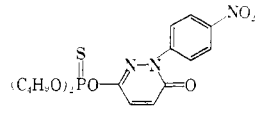
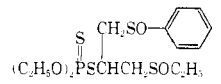
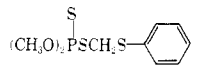
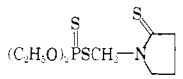
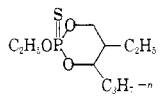
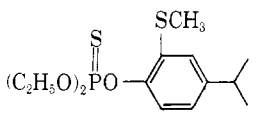
Structures	Phosphorus index entry and print-out order	Other index entries
	PES.QB6-A1UN-L.QA1/2	NUA1/B6-L-QP/ES/QA1/2 LB6-A1UN-QP/ES/QA1/2
	PES.QB6-A1UN-L.QA2/2	NUA1/B6-L-QP/ES/QA2/2 LB6-A1UN-QP/ES/QA2/2
	PES.QB6-A1UN-L.QA2/2	NUA1/B6-L-QP/ES/QA2/2 LB6-A1UN-QP/ES/QA2/2
	PES.QB6-AS1-A1.QA1/2	S.B6-A1-QP/ES/QA1/2.A1
	PES.QB6-SA1-A1/2.QA2/2	S.B6-A1/2-QP/ES/QA2/2.A1
	PES.QB6-SA1-T.QA2/2	S.B6-T-QP/ES/QA2/2.A1
	PES.QB6-SA1-L.QA1/2	S.B6-L-QP/ES/QA1/2.A1
	PES.QZ6N2-2H2-EQ.QA2/2	Z6N2-2H2.EQ.QP/ES/QA2/2
	PES.QZ6N2-2H2-EQ-A1Z6NQ-4H4.QA2/2	Z6N2-2H2.EQ.A1Z6NQ-4H4.QP/ES/QA2/2 Z6NQ-4H4.A1Z6N2-2H2-EQ-QP,ES,QA2,2
	PES.QZ6N2-2H2-EQ-B6NQ2.QA4/2	Z6N2-2H2.EQ.B6-NQ2.QP/ES/QA4/2 Q2N/B6Z6N2-2H2-EQ-QP,ES,QA4,2
	PES.SA1/A1S,EQ,B6/A1S,EQ,A2.QA2/1	S/EQ.A2/A1S,EQ,B6/SP,ES,QA2,2.A2 S/EQ.A2/SP,ES,QA2,2/AS,EQ,A2.B6

Table III (*Continued*)
Machine Sorting and Print-Out of Phosphorus Pesticides by the Hercules Notation

Structures	Phosphorus index entry and print-out order	Other index entries
$(C_2H_5O)(CH_3O)P(=S)SCH_2CONHCH_3$	PES.SA1KN/A1.QA1.QA2	KN/A1.A1SP/ES/QA1/QA2
$(CH_3O)_2P(=S)SCH_2CONHCH_3$	PES.SA1KN/A1.QA1/2	KN/A1.A1SP/ES/QA1/2
$(CH_3O)_2P(=S)SCH_2CON(CH_3)_2$	PES.SA1KN/A1/2.QA1/2	KN/A1/2.A1SP/ES/QA1/2
$(CH_3O)_2P(=S)SCH_2CONHCH_2CH=CH_2$	PES.SA1KN/A1V.QA1/2	KN/A1V.A1SP/ES/QA1/2
$(C_2H_5O)_2P(=S)SCH_2SP(=S)(OC_2H_5)_2$	PES.SA1SP/ES/QA2/2.QA2/2	VA1NKA1SP/ES/QA1/2
	PES.SA1SB6.QA1/2	S.A1SP/ES/QA1/2.B6
	PES.SA1Z5NH4-ES.QA2/2	Z5NH4.ES.A1SP/ES/QA2/2
$(C_2H_5O)_2P(=S)SCH_2CH_2SCH_3$	PES.SA2SA1.QA2/2	S.A2SP/ES/QA2/2.A1
	PZZ6Q2P-3,2H2.ES.QA2.A2.A3	Z6Q2P-3,2H2.ES.QA2.A2.A3



PES.QA/2.QBX	12 columns
PES.QA/2.QB-A-X	15 columns
PES.QA2/2.QB6-A-SA	18 columns
PES.QA2/2.QB6-Y-SA1	19 columns
PES.QA2/2.QB6-4Y-2SA1	21 columns (showing *o*- and *p*-substituents)

Degree of detail and depth of indexing are thus important factors in designing a notation system for indexing. These factors are influenced by the number of chemicals to be indexed, the complexity of the chemicals, the interrelationships that need to be correlated, and, most important, the needs of the users.

Advantages of Specific Notation System.—The notation system described in this paper was designed as an indexing tool for classes of chemical compounds of interest as pesticides. It was necessary to design this notation system because existing notation systems, which are universal in concept, are difficult to apply to pesticides. This is not a criticism of the existing notation systems, as the history of documentation has many examples of the difficulty in using universal systems, such as the Dewey, Library of Congress, and Universal Decimal Classification, for specific areas of information. Rather, our notation system owes much to the inventors of universal notation systems.

A specific notation system is no less unique in expressing structures concisely, linearly, and without ambiguity than are the universal notation systems. Approximately 35 columns of an IBM card are required by each of the notation systems for the phosphorus compounds we studied. The primary difference between our notation system and others described in the literature is the viewpoint. The difference of viewpoint is very much like that between indexing and nomenclature. Whereas there are many ways for indexing a specific compound, *e.g.*, ethyl alcohol, or a class of compounds, *e.g.*, alcohols, nomenclature must abide by convention, such as IUPAC or *Chemical Abstracts*, or by the dictates of usage. Thus ethyl alcohol, depending upon the indexer's viewpoint, may be indexed under any of the following headings.

Alcohol, ethyl
Ethyl alcohol
Ethanol
Methyl carbinol
Solvent
Antifreeze

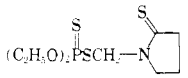
Table IV
Permuted Dyson Notations

A5ZN1ES2:CS/	P	ES(QC2)2
C2:Q/2	P	ES;S/2C2SC
C3E:N/EQC2:2S/	P	ES(QC)2
C:Q/2	P	ES;S/2C2EQ;N(C)2

Table V
Permuted Wiswesser Notations

T5NYTJ—A1S	P	S&O2&O2
2O	P	S&O2&S2S1
1U2MV1S	P	S&O1&O1
1O	P	S&O1&S1VN1&1

Table VI
Machine Sorting Order

	Hercules	*Chem. Abstr.*	Dyson	Wiswesser	Permuted Dyson		Permuted Wiswesser	
					Left sort	Right sort	Left sort	Right sort
$(CH_3O)_2PSCH_2CON(CH_3)_2$	1	4	2	2	3	3	1	2
$(CH_3O)_2PSCH_2CONHCH_2CH=CH_2$	2	3	4	3	2	1	4	1
	3	1	1	1	1	2	3	3
$(C_2H_5O)_2PSCH_2CH_2SCH_3$	4	2	3	4	4	4	2	4

The development of our notation system was prompted by the objective of indexing flexibility. Indeed, we wanted no less flexibility than we already had with our machine card file which used fixed positions in 35 columns for designating functional groups and moieties. Machine card codes based on dedicated positions require considerable machine sorting time. The greatest disadvantage, however, of such machine card codes is that they cannot be used for producing a machine print-out.

Our notation code, by design, is as flexible as any machine card code in denoting functional groups and moieties, with the additional advantages of showing relationships between functional groups and moieties and of being particularly suitable for the production of machine print-outs. The production of machine print-outs in as many ways as there are indexing viewpoints was an important accomplishment of this investigation.

BIBLIOGRAPHY

(1) Bonnett, H. T., and D. W. Calhoun, *J. Chem. Doc.*, **2**, 2 (1962).
(2) Bouman, H., *ibid.*, **3**, 92 (1963).
(3) Duffin, W. M., *ibid.*, **1** (3), 40 (1961).
(4) Dyson, G. M., "A New Notation and Enumeration System for Organic Compounds," 2nd Ed., Longmans, Green and Co., New York, N. Y., 1949.
(5) Frome, J., *J. Chem. Doc.*, **1** (1), 76 (1961).
(6) Frome, J., *ibid.*, **1** (1), 84 (1961).
(7) Frome, J., F. M. Sikora, and M. Gannon, *ibid.*, **1** (1), 88 (1961).
(8) Frome, J., *ibid.*, **2**, 15 (1962).
(9) Frome, J., and P. T. O'Day, *ibid.*, **2**, 248 (1962).
(10) Gelberg, A., W. Nelson, G. S. Yee, and E. A. Metcalf, *ibid.*, **2**, 7 (1962).
(11) Silk, J. A., *ibid.*, **3**, 189 (1963).
(12) Sorter, P. F., C. E. Granito, J. C. Gilmer, A. Gelberg, and E. A. Metcalf, *ibid.*, **4**, 56 (1964).
(13) Starker, L. N., and J. A. Cordero, *ibid.*, **2**, 12 (1962).
(14) Wiswesser, W. J., "A Line-Formula Chemical Notation," W. Y. Crowell Co., New York, N. Y., 1954.