

- McGraw-Hill, New York, N.Y., 1968.
- (5) More than 20 papers under the main title "Application of Artificial Intelligence for Chemical Inference" by the Artificial Intelligence Group of Stanford University have appeared in *J. Am. Chem. Soc.* since 1969. Pattern recognition is a branch of artificial intelligence familiar to chemists. Papers of general interest can be found in Computer and Control Abstracts, Section 6.440.
 - (6) E. J. Corey, W. T. Wipke, R. D. Cramer III, and S. J. Howe, "Techniques for Perception by a Computer of Synthetically Significant Structural Features in Complex Molecule", *J. Am. Chem. Soc.*, **94**, 431 (1972).
 - (7) J. B. Hendrickson, "Systematic Synthesis Design. IV. Numerical Codification of Construction Reactions", *J. Am. Chem. Soc.*, **97**, 5784 (1975).
 - (8) A. J. Stuper and P. C. Jurs, "ADAPT: A Computer System for Automated Data Analysis Using Pattern Recognition Technique", *J. Chem. Inf. Comput. Sci.*, **16**, 99 (1976).
 - (9) W. E. Brugger, A. J. Stuper, and P. C. Jurs, "Generation of Descriptors from Molecular Structures", *J. Chem. Inf. Comput. Sci.*, **16**, 105 (1976).
 - (10) D. H. Smith, L. M. Masint, and N. S. Sridharan, "Heuristic DENDRAL: Analysis of Molecular Structure", in "Computer Representation and Manipulation of Chemical Information", W. T. Wipke, S. R. Heller, R. J. Feldmann, and E. Hyde, Ed., Wiley, New York, N.Y., 1974.
 - (11) M. Osinga and A. A. V. Stuart, "Documentation of Chemical Reactions. III. Encoding of the Facets", *J. Chem. Inf. Comput. Sci.*, **16**, 165 (1976).
 - (12) G. W. Adamson and D. Bawden, "An Empirical Method of Structure-Activity Correlation for Polysubstituted Cyclic Compounds Using Wiswesser Line Notation", *J. Chem. Inf. Comput. Sci.*, **16**, 161 (1976).
 - (13) E. V. Krishnamurthy, P. V. Sankar, and S. Krishnan, "ALWIN-Algorithmic Wiswesser Notation System for Organic Compounds", *J. Chem. Doc.*, **14**, 130 (1974).
 - (14) C-H. Lin, "Computer on the Lab Bench—A Primary Step toward Computer Chemistry", *Kagaku no Ryoiki*, **28**, 771 (1974).
 - (15) S. Rossler and A. Kolb, "The GREMAS System, an Integral Part of the IDC System for Chemical Documentation", *J. Chem. Doc.*, **10**, 128 (1970).
 - (16) M. Z. Balent and J. M. Emberger, "A Unique Chemical Fragmentation System for Indexing Patent Literature", *J. Chem. Inf. Comput. Sci.*, **15**, 100 (1975).
 - (17) C-H. Lin, "Chemical Inference Based on SEFLIN. I. Basic Cognizance of Molecular Shape, Fragments, and Atomic Environment of Organic Compounds", *J. Chem. Inf. Comput. Sci.*, following paper in this issue.
 - (18) H. Skolnik, "A New Linear Notation System Based on Combinations of Carbon and Hydrogen", *J. Heterocycl. Chem.*, **6**, 689 (1969).
 - (19) H. Skolnik, "A Computerized System for Storing, Retrieving, and Correlating NMR Data", *Appl. Spectrosc.*, **26**, 173 (1972).
 - (20) The set of square brackets is an identifier and is not included in the computer memory as part of SEFLIN.
 - (21) Hydrogen atoms in the molecule become apparent as the SEFLIN is transformed into the atom-environment table; see ref 17.

Chemical Inference Based on SEFLIN. I. Basic Cognizance of Molecular Shape, Fragments, and Atomic Environment of Organic Compounds

CHI-HSIUNG LIN

Department of Chemistry, National Cheng Kung University, Tainan, Taiwan, Republic of China

Received September 6, 1977

Algorithms are presented to obtain basic information in the computer memory from input SEFLIN of organic structures. The group connectivity table and the atom environment table with a ring locant table are used as bases for treating physical and chemical behavior of organic substances. A scheme of automatic generation of descriptor sets from SEFLIN for pattern recognition is shown. A subroutine for rearranging the atom environment table after the chemical transformation of the compound in reaction is presented.

The Separate Feature Linear Notation (SEFLIN) is a new structural formula system for chemical study under the man-machine interactive environment.¹ It is chemist-oriented in that the encoding is rather straightforward from the conventional structural formula. It is also machine-oriented because it is easily generated in the computer if some chemical characteristics are assigned. To a computer, SEFLIN is a communication device with the chemist as well as a database with which chemical behavior of the corresponding substance is to be studied.

The chemist conceives the chemical nature of a substance from the structural formula, and its fragments and their combinational characteristics. These are equally evident from the SEFLIN, and therefore the chemist may learn to look into the SEFLIN through the chemical common sense he already possesses. This common sense is difficult for a computer to acquire. In this series of papers, programming strategies which endow the computer with such chemical common sense will be demonstrated.

The term "chemical inference" is quite vague. It may mean any chemically significant judgment based on molecular structure which is thought to be approved by most chemists. It may also include schemes of deduction most chemists prefer in reaching an assumption or conclusion. It may be strictly logical or somewhat intuitive. But never is it absolutely accurate. This is because the object of chemistry is a huge

number of substances under infinitely diverse situations. A valid chemical inference has a statistical nature. Thus, a better inference may be what leads to a prediction closer to the statistical means of those tested substances. Since the number of tested substances is continuously increasing, any approved inference is always subject to improvement. What is really important, therefore, is to settle down to exploit the algorithm with which a certain chemical inference can be made by the average chemist or, as we hope, the computer.

There are several levels of structural problems on which an inference is made. We may look at the size and shape of the molecule. We may say there does exist a certain functional group. We may find that a certain atom or group is situated at a certain position in reference to another atom or group. We ask for the electron density distribution. We are even interested in such matters as the through-the-space interaction of electronic or nuclear spins. All of these are studied by the chemist on the same structural formula of the substance. This is to say that the conventional structural formula is successful in providing levels of information. But, we should not forget the hidden factor which is the amazing human intelligence. We look into the structural formula with a proper depth of focus for the occasion and deduce therefrom different concepts for the specific purpose. This is the mysterious power of cognizance. The artificial intelligence we need to build for the study of chemistry should be endowed with the capability

Table IV. Group KTBL of Quinine, Riboflavin Monoacetate, and Riboflavin Monobenzoate

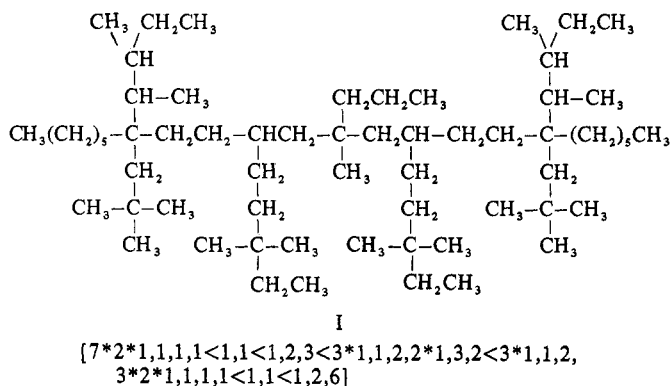
Quinine	Riboflavin monoacetate	Riboflavin monobenzoate
KSET {t x y z} {o}	KSET {t x y z} {o}	KSET {t x y z} {o}
-1 R) == -2 0	-1 R) == -2 0	-1 R) == -2 0
-2 I) == -3 -1	-2 I) == -3 -1	-2 I) == -4 -3 -1
-3 1 0 0 -4 -2	-3 J) == 0 -2	-3 J) == 0 -2
-4 R) == -5 -3		-4 5 0 0 -5 -2
-5 I) == 0 -4		-5 O . . -6 -4
		-6 V = . -7 -5
		-7 R) == 0 -6

Table V. Generation of the Group-KTBL

[4*OV2,1<1,1,3<1N <1,1,2*1,1,2]		$ \begin{array}{c} \text{CH}_3\text{CH}_3 \\ \\ \text{CH} \\ \\ \text{CH}_3\text{CH}_2\text{CH}_2\text{CCH}_2\text{CH}_2\text{CHCH}_2\text{CCH}_2\text{CH}_3 \\ \quad \quad \\ \text{O-COCH}_2\text{CH}_3 \quad \text{CH}_3 \end{array} $	
KSET {t x y z}		KSET {t x y z} {o}	
-1 4 x y z		-1 4 -5 -4 -2 0	
-2 O . . z		-2 O . . -3 -1	
-3 V = . 2		-3 V = . 2 -2	
-4 1 0 1 1		-4 1 0 1 1 -1	
-5 3 0 y z	⇒	-5 3 0 -8 -6 -1	
-6 1 0 0 z		-6 1 0 0 -7 -5	
-7 N . 1 1		-7 N . 1 1 -6	
-8 2 1 1 z		-8 2 1 1 -9 -5	
-9 2 0 0 0		-9 2 0 0 0 -8	

subroutine, all symbols except numerals followed by commas are picked up to compose KSET referring to the {t x y z} format in Table III. One, two, or three 0 symbols are put into the carbon-chain KSET which is followed by a two-way branch [<], a ring or heteroatom symbol, or nothing at all, respectively. The remaining numerals in the SEFLIN are then properly put into these KSET. Connection numbers are finally put into undefined spaces orderly as demonstrated in the example.

The Group-KTBL is a device for the computer to make judgment on bulk, molecular properties. The main chain length of compound I, for instance, is found by tracing the t



data from KSET(-1) along the route led by first connection numbers. The total is 23 carbon atoms in this example. Reducing 1 from the first t datum, a series {6,3,2,2,3,6}, called the main-chain series, is obtained. The remaining data in their xyz triads give another series {(-3,-2)(0,-6)(1,3)(0,-9)(-12,-11)}. Symmetry of the molecule is apparent in this study (see Table VI).

The reverse transformation from Group-KTBL to SEFLIN is similarly simple. The generation of all possible isomeric skeletons can be conducted by a systematic generation of Group-KTBL, on which the uniqueness of the isomer is more easily tested.

Table VI. Group-KTBL of Compound I

KSET {t x y z} {o}
-1 7 -5 -3 -2 0
-2 2 1 1 1 -1
-3 1 0 1 1 -4 -1
-4 1 0 1 2 -3
-5 3 0 -7 -6 -1
-6 3 1 1 2 -5
-7 2 1 3 -8 -5
-8 2 0 -10 -9 -7
-9 3 1 1 2 -8
-10 3 -14 -12 -11 -8
-11 2 1 1 1 -10
-12 1 0 1 -13 -10
-13 1 0 1 2 -12
-14 6 0 0 0 -10

Table VII. AETBL of Compound II

$ \begin{array}{c} \text{F} \quad \text{CH}_2\text{CH}_2\text{CHO} \\ \quad \\ \text{CH}_3-\text{C}-\text{CH}=\text{C} \\ \quad \\ \text{F} \quad \text{CH}=\text{CH}-\text{CN} \end{array} $	
II	
[4<2VH,2Y/.3-4.9-10/2F2F]	
AESET {t x y z} {o}	
-1 1 0 0 -2 0	
-2 1 F F -3 -1	
-3 1 0 . -4 -2	
-4 1 . -9 -5 -3	
-5 1 0 0 -6 -4	
-6 1 0 0 -7 -5	
-7 V = . -8 -6	
-8 H = = -7	
-9 1 0 . -10 -4	
-10 1 0 . -11 -9	
-11 Y = . -10	

ATOM ENVIRONMENT TABLE

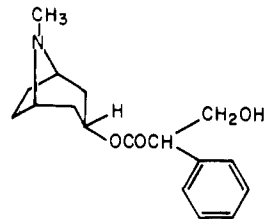
In the study of reactivity of substances, interest is focused on the individual atom rather than on the bulk structure of the molecule. Modes of linkages on the given atom, effects exerted by nearby atoms and groups, and the charge or the electron density distribution are generally the factors taken for considering the characteristics of atoms in reaction. In the automatic design of synthesis,⁵ some kind of activity index for each atom in the molecule is hoped to be generated swiftly in the computer memory. The atom environment table (AETBL) is devised for this purpose. The transformation to this from SEFLIN can be processed easily as in the case of the Group-KTBL (Table VII).

The AETBL is orderly assemblage of AESET of every atom in the skeleton. Information in the second and the third features of SEFLIN is included in AESET. The atomic orbital donated for π -bond formation is represented with a dot symbol (.) in AESET. Each substitution group in the third feature occupies a space in the xyz triad of the proper AESET.

A ring-locant table (Ring-LTBL) is used in generating the AETBL of a compound with ring structures (see Table VIII). The symbol [R5], for instance, triggers the assignment of five locants A, B, C, D, and E in the Ring-LTBL and five consecutive AESET in the AETBL. The arc symbol [A3C] assigns locants IA and IE to the AESET with locants A and C, and opens three AESETs with locants IB, IC, and ID.

In AETBL, the symbol [0] corresponds to a hydrogen atom attached to the skeleton. The environmental situation of the hydrogen atom, important in some chemical inference such as the analysis of the ¹H NMR spectrum,⁶ is clear in the table. The connectivity of dot symbols is used in estimating π -electron density distribution. This and other subjects for a proper chemical inference will be presented in later papers of this

Table VIII. AETBL of Atropine and the Corresponding Ring-LTBL



[R5.BN)B1,A3C)ICOV1<1,R6)/.14-19/13Q]

{R I}	AESET	{t x y z}	{o}
A A	-1	1 0 -7 -2	-5
B -	-2	N . -6 -3	-1
C E	-3	1 0 -9 -4	-2
D -	-4	1 0 0 -5	-3
E -	-5	1 0 0 -1	-4
-	-6	1 0 0 0	-2
- B	-7	1 0 0 -8	-1
- C	-8	1 0 -10 -9	-7
- D	-9	1 0 0 -3	-8
-	-10	O . . -11	-8
-	-11	V = . -12	-10
-	-12	1 0 -14 -13	-11
-	-13	1 0 0 Q	-12
A -	-14	1 . -19 -15	-12
B -	-15	1 0 . -16	-14
C -	-16	1 0 . -17	-15
D -	-17	1 0 . -18	-16
E -	-18	1 0 . -19	-17
F -	-19	1 0 . -14	-18

series. In the AETBL of atropine, it is quite apparent that this substance is a tertiary amine (AESET(-2)) and primary alcohol (AESET(-13)); the only methyl group (AESET(-6)) is attached to a nitrogen atom (AESET(-2)). The *o* datum of AESET(-1) is set 0 if the substance is acyclic.

The selection of the most possible change of the substance under a given reaction condition is mostly made on AETBL. The heterolysis of the C₁-N bond of atropine after it is changed to a quaternary ammonium salt corresponds to the replacement of -2 in AESET(-1) and -1 in AESET(-2) with symbols [+] and [-], respectively. In the AETBL of the quaternary ammonium salt, the *x* datum of AESET(-2) is a connection number.

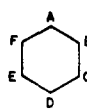
AESET COLLOCATION

The heterolysis in the above example destroys the original base ring and the resultant compound is encoded into a quite different SEFLIN. The numbering scheme and the locant assignment change drastically. A similar situation happens in case of the ring closure. Thus, in the machine manipulation of a reaction process, the AETBL should be checked at every stage of transformation, and a proper rearrangement is made to bring it into the form from which a good SEFLIN can be generated. This process is called the "AESET collocation" and is handled by a specific subroutine.

The collocation subroutine consists of four operations, \hat{W} , \hat{L} , \hat{R} , and \hat{H} . The operation \hat{W} corrects the arrangement of connection numbers in the *xyz* triad. The operation \hat{L} generates a general locant table (LTBL) based on the temporary AETBL. This LTBL contains ordinary ring locants and other symbols, including connection symbols Q, S, and T and chain locants I, X, Y, W, etc. Most rearrangements of the temporary AETBL are made in this operation. A simple example is shown in Table IX. The operation \hat{R} properly revolves, rotates, and turns rings so as to let the result conform to the ring numbering scheme of the SEFLIN system. The operation \hat{H} adjusts the order of chains with the embedded hierarchy rules.

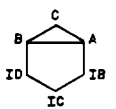
The subroutine as a whole is considerably complex, but in

Table IX. Change of Atom Numbering and Locant Specification during a Structural Transformation



[R6]

→ B-F linkage → new R-ring →



[R3]A3B))

{R}			{R I}			{R I}			{R I}		
	#			(R#) ^o		(R#) ^o		(R#) ^o		#	(R#) ^o
A	-1	A -	(A-1)	A -	(A-1)	A A	-1	(B-2)			
B	-2	B A	(B-2)	B A	(B-2)	B E	-2	(F-6)			
C	-3	- B	(C-3)	C E	(F-6)	C -	-3	(A-1)			
D	-4	- C	(D-4)	- B	(C-3)	- B	-4	(C-3)			
E	-5	- D	(E-5)	- C	(D-4)	- C	-5	(D-4)			
F	-6	C E	(F-6)	- D	(E-5)	- D	-6	(E-5)			

Table X. Four Kinds of Temporary Locant Combination Suggestive of a Ring or an Arc

Combination	Example	Ring locants ^a	Condition
R-V	{R I I I I V - - -}	{A B C D E F - - -}	Unconditional
T-W	{- - T - - - W - -}	{- - A - - - B* - -}	Conditional
T-V	{- - T - - Y I V -}	{- - A - - B C D* -}	Conditional
R-U	{- - R I I - - U -}	{- - B C D - - A -}	Conditional

^a An asterisk designates: turn upward to another column and trace down.

practice the collocation is done in a short time, almost instantaneously, under the interactive operation with a personal computer. After every AESET is checked for the order of existing connection numbers in the sequence $|N(x)| > |N(y)| > |N(z)|$, S, R, and T columns are set out in the LTBL through a survey from AESET(-1) and downward. AESETs are assigned with locants R or S in respect to whether the *o* datum is a connection number pointing to a posterior AESET or a symbol, respectively. If the *y* datum is a connection number, the AESET is assigned with a locant T.

To each T column, locants Y and X are assigned to the AESET pointed by the *y* and the *x* data of the T-AESET. If the *z* datum of this Y- or X-AESET is a connection number pointing back to the T-AESET, the locant assignment is canceled. If it points to any other preceding AESET, then the locant is replaced with W suggesting a "turning point". If a T-AESET is itself a turning point, the locant is replaced with U, and the pointed AESET is given the locant R. Locant S is changed to locant \$ if the AESET is a turning point.

Starting from locant S, R, Y, and X, connection routes are traced down referring to the *z* data, and every AESET on the routes is assigned with a locant I. A turning point is assigned with a locant V instead, and if this V-AESET has a locant W in another LTBL column, this latter locant is wiped off. A column with a solitary locant T is cancelled any time this happens.

Another survey is conducted from the bottom for AESET with locant R or T. If the AESET also has a locant I which is succeeded by one or more locants I in the column not ending up in a locant V, then these successive locants I are transferred to the R or the T column. If this is a T column, the leading locant I transferred is replaced with a locant Z.

In the next process, ring structures are looked for by referring to Table X. The condition that T-W, T-V, and R-U combinations are arcs or rings is that the tracing terminates on the starting AESET, on the X-AESET that follows the starting T-AESET, or on an AESET to which a ring locant is already assigned.

After the assignment of ring locants, the subroutine conducts a rearrangement in which a portion of the temporary AETBL

Table XI. Elements of Operation in the Operation \hat{R}

Element of operation	Collocation	(xyz \rightleftharpoons o)	Example (R-ring)		Original locants ^a
			From	To	
Revolution	Circular collocation	No	$\begin{Bmatrix} R \\ I \end{Bmatrix} \begin{Bmatrix} A & B & C & D & E & - & - \\ - & - & A & D & - & B & C \end{Bmatrix}$	$\begin{Bmatrix} R \\ I \end{Bmatrix} \begin{Bmatrix} A & B & C & D & E & - & - \\ A & D & - & - & - & B & C \end{Bmatrix}$	{CDEAB}
Rotation	Reversing collocation	Yes	$\begin{Bmatrix} R \\ I \end{Bmatrix} \begin{Bmatrix} A & B & C & D & E & - & - \\ A & - & - & - & D & B & C \end{Bmatrix}$	$\begin{Bmatrix} R \\ I \end{Bmatrix} \begin{Bmatrix} A & B & C & D & E & - & - \\ A & D & - & - & - & B & C \end{Bmatrix}$	{A'E'D'C'B'}
Exchange	Reversing collocation	Yes	$\begin{Bmatrix} R \\ I \end{Bmatrix} \begin{Bmatrix} A & B & C & D & E & - & - \\ D & A & - & - & - & B & C \end{Bmatrix}$	$\begin{Bmatrix} R \\ I \end{Bmatrix} \begin{Bmatrix} A & B & C & D & E & - & - \\ A & D & - & - & - & B & C \end{Bmatrix}$	{B'A'E'D'C'}
Alphabetization	Rearrangement	No	{A H F G D E B C}	{A H B C D E F G}	{AHBCDEFG}

^a A', etc., represent AESET with data exchange.

is replaced and inserted in another place. The portion, if ending up in a locant V, is inverted prior to the insertion.

In the LTBL, there may be columns which have locants I for the AESETs that have ring locants. Such doubled locants I are changed to locants Q expressing the joint positions for side chains. If a column has two or more locants Q, these parts are separated into different columns. Any AESET pointed upward by a V- or \$-AESET is also a joint position for a side chain. A locant Q is assigned for the AESET at a space in the V or the \$ column. This AESET must have a locant in another column. This column is traced down to the last locant. The original \$-AESET is collocated to the position following this last locant. If it is a V- or U-AESET instead of a \$-AESET, a portion of the AETBL is inverted and inserted here. The portion is either {SI...V}, {TI...V}, or {RI...U}, or (Q){I...V} where the Q-AESET is not included. The inversion requires both the inversion of the AESET order and the inversion of the connection sequence. The latter is accomplished by exchanging the connectivity number in the o datum with one in the xyz triad.

Any locant S not at AESET(-1) signals the collocation of the portion to a lower position in the AETBL. This collocation also requires the inversion. The last locant on such a S column is changed to a locant Q, and the portion {SI...I}(Q) is inverted and inserted into the place to be found in the manner stated above. The locant S is changed to a locant I after the collocation.

In the last stage of operation \hat{L} , ring locants are gathered together as demonstrated in Table IX with the process (B-F linkage) \rightarrow (new R ring). If there exists AESET without ring locants between the A-AESET and the B-AESET of the base ring, they are collocated to the position following the C-AESET. This collocation requires no inversion or data modification.

The operation \hat{R} arranges AESET of each ring and arc in the required order which conforms with the atom-numbering scheme of the SEFLIN system. There are four elements of operation listed in Table XI.

The test of the structural hierarchy rules is made in the final operation \hat{H} . The "terminal ring first" strategy is already fulfilled by the preceding operation. The remaining problem is whether the first ring is the right one. If it is not, then a proper exchange of the base ring and the arc or a proper inversion of the whole AETBL is put into action. The system [R5)A2B)] in Table XI is changed to [R4)A3B)] because the base ring should be the smaller one. The inversion of the whole AETBL is conducted by inverting the order of AESET and exchanging the connection numbers. The portions in T columns and in Q columns are moved as intact units. Lastly, the hierarchy test is conducted on side chains in each T column and on those Q columns attached to the same ring.

At the exit of the collocation subroutine, the numbering of AESET is brought back into the natural order.⁷ Actually, after the collocation, the topmost AESET may have a numbering such as -12. Then, any datum with a connection

number -12 in the AETBL is to be changed to -1. This process is made with a reference table.

ACTIVITY TABLE

Activity indexes of atoms in the molecule for a certain chemical inference are listed in a table called the activity table (ACTTBL). Actually, ACTTBL is the general term for any matrix of data for the inference. Data such as the π -electron density distribution, the relative susceptibility for a nucleophilic reagent, and measures related to the bond strain and steric hindrance are mostly generated from SEFLIN through its Group-KTBL or AETBL. Physical properties for atoms in the molecule, such as the coupling constants and chemical shifts in ¹H NMR spectroscopy, are filed in the mass storage for retrieval with the SEFLIN. The construction of the latter type of data belongs to the discipline of database management. The ACTTBL for the former type of activity data is constructed in an analogous format as the Group-KTBL and the AETBL. Data in the t field are for the corresponding skeletal atoms or groups. Data in the xyz triad are for substitution groups and for the interatomic or intergroup linkages expressed with the connection numbers. In some cases, auxiliary tables are used to list activity indexes for atoms and linkages in those multiatom substitution groups. Sometimes, a small ACTTBL to include a limited portion of the AETBL is used in the inference. The generation of such an ACTTBL and the application for the given chemical inference rely on the experience of chemists at the present stage of development. It is possible that the intelligence of chemists may be systematized into a unified scheme to be embedded into the computer program. Then, it may become possible to conduct chemical studies on the whole population of the huge amount of chemical information utilizing the exactness and the theoretically limitlessness of the machine memory.

REFERENCES AND NOTES

- (1) C-H. Lin, "SEFLIN—Separate Feature Linear Notation System for Chemical Compounds", *J. Chem. Inf. Comput. Sci.*, preceding paper in this issue.
- (2) W. E. Brugger, A. J. Stuper, and P. C. Jurs, "Generation of Descriptors from Molecular Structure", *J. Chem. Inf. Comput. Sci.*, **16**, 105 (1976).
- (3) In the first feature of SEFLIN, symbols X, I, G, F, E, and H appear only with the symbol V. The symbol Q appears with symbols V and P. An equal symbol [=] in the format corresponds to an inactive corner of the atom and the dot symbol [.] corresponds to a p orbital or a π bond. The format is derived from the octet theory of organic compounds.
- (4) The symbol [0] for the z datum illustrates the termination. The general meaning of this symbol in the Group-KTBL and in the AETBL is a hydrogen atom.
- (5) J. H. Hendrickson, "Systematic Synthesis Design. III. The Scope of the Problem", *J. Am. Chem. Soc.*, **97**, 5763 (1975).
- (6) A specific linear notation system for ¹H NMR spectra is presented in: H. Skolnik, "A Computerized System for Storing, Retrieving, and Correlating NMR Data", *Appl. Spectrosc.*, **26**, 173 (1972).
- (7) AESET numbering at the entrance to this subroutine is orderly but is not necessarily consecutive. This is especially so when the AETBL corresponds to a product of the reaction such as the combination of two molecules, the elimination of a part of the molecule, etc. The collocated AETBL leaving the subroutine, however, is properly numbered in the natural order.