knowledgeably interpreting the results. Thus, the design of this program follows the general aim of computer-assisted structure elucidation: to use certain narrow, but powerful, capabilities of the computer so that the expert knowledge of the chemist can be applied more effectively.

ACKNOWLEDGMENT

The financial support of this work by the National Institutes of Health and The Upjohn Co. is gratefully acknowledged.

REFERENCES AND NOTES

(1) Present address: National Center for Biomedical Infrared Spectroscopy, Battelle—Columbus Laboratories, 505 King Avenue, Columbus, OH 43201.
(2) Munk, M. E.; Shelley, C. A.; Woodruff, H. B.; Trulson, M. O. "Computer-Assisted Structure Elucidation". *Fresenius' Z. Anal. Chem.* **1982**, *313*, 473-479.
(3) Shelley, C. A.; Hays, T. R.; Roman, R. V.; Munk, M. E. "An Approach to Automated Partial Structure Expansion". *Anal. Chim. Acta* **1978**, *103*, 121-132.
(4) Carhart, R. E.; Smith, D. H.; Brown, H.; Djerassi, C. "Applications of Artificial Intelligence for Chemical Inference. 17. An Approach to Computer-Assisted Elucidation of Molecular Structure". *J. Am. Chem. Soc.* **1975**, *97*, 5755-5762.
(5) Carhart, R. E.; Smith, D. H.; Gray, N. A. B.; Nourse, J. G.; Djerassi, C. "GENOA: A Computer Program for Structure Elucidation Utilizing Overlapping and Alternative Substructures". *J. Org. Chem.* **1981**, *46*, 1708-1718.
(6) Munk, M. E.; Farkas, M.; Lipkus, A. H.; Christie, B. D. "Computer-Assisted Chemical Structure Analysis". *Mikrochim. Acta* **1986** II, 199-215.
(7) Kudo, Y.; Sasaki, S. "Principle for Exhaustive Enumeration of Unique Structures Consistent with Structural Information". *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 43-49.
(8) Lipkus, A. H.; Munk, M. E. "Combinatorial Problems in Computer-Assisted Structural Interpretation of Carbon-13 NMR Spectra". *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 38-45.
(9) Shelley, C. A.; Munk, M. E. "CASE, A Computer Model of the Structure Elucidation Process". *Anal. Chim. Acta* **1981**, *133*, 507-516.
(10) Lindsay, R. K.; Buchanan, B. G.; Feigenbaum, E. A.; Lederberg, J. *Applications of Artificial Intelligence for Organic Chemistry: The DENDRAL Project*; McGraw-Hill: New York, 1980; pp 65-66.
(11) Shannon, C. E.; Weaver, W. *The Mathematical Theory of Communication*; University of Illinois: Urbana, 1949; pp 48-52.
(12) Wells, M. B. *Elements of Combinatorial Computing*; Pergamon: New York, 1971; Chapter 6.
(13) Ibid., Chapter 4.
(14) Shelley, C. A.; Munk, M. E. "An Approach to the Assignment of Canonical Connection Tables and Topological Symmetry Perception". *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 247-250.
(15) Ichihara, A.; Shiraishi, K.; Sato, H.; Sakamura, S.; Nishiyama, K.; Sakai, R.; Furusaki, A.; Matsumoto, T. "The Structure of Coronatine". *J. Am. Chem. Soc.* **1977**, *99*, 636-637.
(16) Varkony, T. H.; Shiloach, Y.; Smith, D. H. "Computer-Assisted Examination of Chemical Compounds for Structural Similarities". *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 104-111.

Further Development of Structure Generation in the Automated Structure Elucidation System CHEMICS

KIMITO FUNATSU, NOBUYOSHI MIYABAYASHI, and SHIN-ICHI SASAKI*

Laboratory for Chemical Information Science, Toyohashi University of Technology, Tempaku, Toyohashi 440, Japan

Received April 20, 1987

The automated structure elucidation system CHEMICS has been expanded to allow the compound with all or any of the elements C, H, O, N, S, and the halogens to be analyzed. To realize the improvement, a new set of primary, secondary, and tertiary components to be used commonly for both spectral analyses and structure construction was established. Twelve attributes of the components were prepared for the determination of the bonding priorities between the components. This makes it possible to construct a tertiary component by giving an attribute of bonding partner to a secondary component. A new structure generator for the improved system has been developed on the basis of the established components set, using the authors' connectivity stack method. The generator was endowed with a function that takes information about the macrocomponent (substructure to be present or absent designated by a user) into the system; candidate structures with or without macrocomponent are generated in accordance with the information.

INTRODUCTION

We have already reported in a series of publications that CHEMICS, an automated structure elucidation system, can serve for analyzing the structures of organic compounds consisting of various combinations of C, H, and O atoms.[1] One of the most important problems to be solved in the study of automated structure elucidation is how to generate appropriate structure on the basis of given information. To solve the problem, CONGEN introduced a peculiar method based on the graph theory.[2] CASE coped with it through a modified canonicalization method of the connection table.[3] In the authors' CHEMICS, the connectivity stack method in which the connectivity matrix is differently expressed is used.[4] In the present study, we make an attempt to expand the system to allow nine atomic species, including C, H, O, N, S, and the halogens, to be analyzed. To realize this, it is necessary to modify the components by which a full structure is generated, previously stored in a computer, and to improve the generator, the core of CHEMICS.

The practical efficiency of structure elucidation work would be greatly increased if a user can use, in addition to spectral data, substructural data (referred to as "macrocomponent" in terms of CHEMICS) obtained from other sources.[5] In the system proposed in this paper, processing of macrocomponents can be incorporated into the structure generation algorithm by coordinating them closely with each other. Furthermore, macrocomponents desired to be incorporated into the candidate structures of an unknown compound and those desired to be excluded can be separately processed or treated.

ESTABLISHMENT OF COMPONENTS

We have previously reported that CHEMICS requires three hierarchical classes of components, i.e., primary, secondary, and tertiary. The same classification is also adopted in this

0095-2338/88/1628-0018$01.50/0

Table I. Primary Components

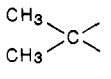
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
CH_3	CH_2	CH	C	OH	O	NH_2	NH	N	SH	S	F	Cl	Br	I

study. A tertiary component is constructed by giving the attribute (afferent nature) of the bonding partner to a secondary component as described later in detail. Primary and tertiary components used are shown in Tables I and II, respectively. Such groupings of components into three classes are useful not only in data analysis but also in performing the generation of component sets, which is required prior to the final structure generation operation. The basic theory for establishing primary and secondary components is described in detail in a previous paper.[6] Here, the number of secondary components is increased from the 63 used in the previous study to 86 in order to carry out spectral data analysis more effectively.

Tertiary Component. A large number of tertiary components would be produced if bonding partners are designated for each bond of each secondary component. A minimum necessary number of tertiary components should be set up that maintains their two major roles, in spectral analysis and as units of structure construction. To accomplish this, descriptors were assumed as shown in Table III with respect to component attributes. We used 12 descriptors, ranging from I to CS, which were numbered from 1 to 12 in order of their priority. The descriptors shown in Table III were added to the secondary components to provide 630 tertiary components, making it possible to identify the attributes of the bonding partners with the bonds of tertiary components (referred to as afferent natures and abbreviated hereafter as AN's). The attributes of root elements of each component, or efferent natures (hereafter abbreviated as EN's) are also presented in Table II.

On the basis of these descriptors and their priority order, the bonding partners with each tertiary component should meet the following four requirements.

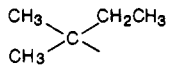
(1) For a single bond of a tertiary component to be connected with another tertiary component through a single linkage, the EN of the former must be equal to the AN of the latter. A remaining single bond of the former can be connected to a component having an EN of a priority rank equal to or below that of the above-mentioned AN. Some examples are



23 AN_{23} = N, EN_{23} = CS

CH_3CH_2- 53 AN_{53} = CS, EN_{53} = CS

The relationship between the ranks of AN_{23} and EN_{53} is $AN_{23} > EN_{53}$, while that between the ranks of AN_{53} and EN_{23} is $AN_{35} = EN_{23}$. There remains a single bond in the **23** component after it is connected to the **53** component. These relationships, therefore, meet requirement 1, permitting the two components to be connected with each other.



I

The possibility of the connection between I and a **577** component is considered below.

577 –O– AN_{577} = CS, EN_{577} = O

One of the single bonds of the **23** component has already been connected with a **53** component (EN = CS). According to the restrictive requirement for AN of the **23** component, the remaining single bond can be connected to a component with an EN equal to N but not to a **577** component. Connection is permitted if the partner is, for example, a **601** component (EN = N) instead of a **577** component.

601 –NH– AN_{601} = CS, EN_{601} = N

(2) A double linkage that connects two components with each other is made by placing a –D– component (**630**) between them.

(3) The single bonds of the **626–629** components (halogens) have attributes of S or below as given in Table III and, therefore, can be connected to any component that meets requirement 1.

(4) An aromatic component (**372–405**) must be connected with at least two aromatic components in addition to itself. Then such a series of the components is defined in order to express the aromaticity, and even if they are connected by single linkage, the series represents an aromatic linkage.

Problems in Establishment of Component. Some structures cannot be generated using these tertiary components alone. There are three reasons as given below, as suggested in our previous paper.[3]

(1) CH_3(CS) is not included in a set of the tertiary components because the *tert*-butyl group, which has already been set up, would be constructed by the use of the CH_3(CS), inevitably resulting in redundancy during the structure generation process.

(2) AN's I–F are not defined for components that have one single bond alone.

(3) Component of H is not defined.

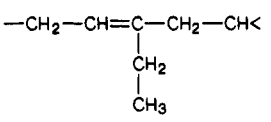
Some of the structures that cannot be generated with restrictions 1–3 are (1) $(CH_3)_2CHCH_3$, CH_3CH_3; (2) CH_3C-H_2—Cl, $CH_3C(=O)$—Cl; (3) $(CH_3)_2NH$, $H_2C=O$, H_2O.

MACROCOMPONENT

A macrocomponent is input in the form of linear combinations of the secondary components. Upon inputting, however, the –D– component and free single bond are omitted.

—CH_2—CH=CH—C= intended substructure CH_2—CH=CH—C= input form

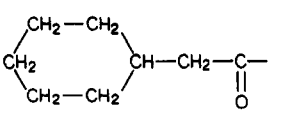
Input of a branched macrocomponent is performed as shown in the example below. A branching position is identified by an asterisk, and it is defined that the position worked with the asterisk is connected to a component which corresponds to the number given next to the asterisk.



intended substructure

CH_2—CH=C*1—CH_2=CH CH_3CH_2*1 input form

Cyclic structures are also represented in the same manner.



intended substructure

CO—CH_2*1 CH_2*2—CH_2—CH_2—CH*1—CH_2—CH_2*2 input form

Positive and Negative Macrocomponents and Input of Two or More of Them. The macrocomponents that may be input by a user can be divided into two groups. Those in one group are desired to be incorporated into the candidate structures

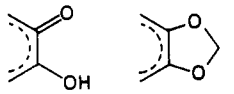
Table II. Tertiary Components

no.	secondary component[a]	efferent nature	afferent nature
1	$(CH_3)_3C$—	CS	S ND N O Y CD CT CS
2	$(CH_3)_2CH$—	CS	S ND N O Y CD CT CS
3	$(CH_3)_2C<$	CS	I Br Cl F S ND N O Y CD CT CS
4	$(CH_3)_2N$—	N	S ND N O Y CD CT CS
5	$(CH_3)_2N(\rightarrow O)$—	N	S ND N O Y CD CT CS
6	$(CH_3)_2C=$	CD	*b*
7	CH_3CH_2—	CS	S ND N O Y CD CT CS
8	$CH_3CH<$	CS	I Br Cl F S ND N O Y CD CT CS
9	$CH_3C\lessdot$	CS	I Br Cl F S ND N O Y CD CT CS
10	CH_3—	CS	ND Y CT
11	CH_3S—	S	S ND N O Y CD CT CS
12	CH_3NH—	N	S ND N O Y CD CT CS
13	$CH_3N<$	N	I Br Cl F S ND N O Y CD CT CS
14	$CH_3N(\rightarrow O)<$	N	I Br Cl F S ND N O Y CD CT CS
15	CH_3O—	O	S ND N O Y CD CT CS
16	$CH_3C(=O)$—	CD	S ND N O Y CD CT CS
17	$CH_3C(=S)$—	CD	S ND N O Y CD CT CS
18	$CH_3C(=NH)$—	CD	S ND N O Y CD CT CS
19	$CH_3C(=N^+=N^-)$—	CD	S ND N O Y CD CT CS
20	$CH_3CH=$	CD	*b*
21	$CH_3C\lessdot$	CD	I Br Cl F S ND N O Y CD CT CS
22	$CH_3S(\rightarrow O)$—	S	S ND N O Y CD CT CS
23	$CH_3S(\rightarrow O)_2$—	S	S ND N O Y CD CT CS
24	—CH_2—	CS	I Br Cl F S ND N O Y CD CT CS
25	>CH—	CS	I Br Cl F S ND N O Y CD CT CS
26	>C<	CS	I Br Cl F S ND N O Y CD CT CS
27	—C≡C—	CT	I Br Cl F S ND N O Y CD CT CS
28	—C≡CH	CT	S ND N O Y CD CT CS
29	—C≡N	CT	S ND N O Y CD CT CS
30	$-N^+\equiv C^-$	ND	S ND N O Y CD CT CS
31	$CH_2=$	CD	*b*
32	O=C=	CD	*b*
33	S=C=	CD	*b*
34	NH=C=	CD	*b*
35	$N^-=N^+=C=$	CD	*b*
36	=C=	CD	*b*
37	—CH=	CD	I Br Cl F S ND N O Y CD CT CS
38	>C=	CD	I Br Cl F S ND N O Y CD CT CS
39	—CHO	CD	S ND N O Y CD CT CS
40	—CHS	CD	S ND N O Y CD CT CS
41	—CH=NH	CD	S ND N O Y CD CT CS
42	$-CH=N^+=N^-$	CD	S ND N O Y CD CT CS
43	>C=O	CD	I Br Cl F S ND N O Y CD CT CS
44	>C=S	CD	I Br Cl F S ND N O Y CD CT CS
45	>C=NH	CD	I Br Cl F S ND N O Y CD CT CS
46	$>C=N^+=N^-$	CD	I Br Cl F S ND N O Y CD CT CS
47	>AC—	Y	I Br Cl F S ND N O Y CD CT CS
48	>ACH—	Y	Y
49	—TPL—[c]	Y	Y
50	—MDO—[d]	Y	Y
51	—AO—	Y	Y
52	—AS—	Y	Y
53	—AN—	Y	Y
54	—ANH—	Y	Y
55	>AN—	Y	I Br Cl F S ND N O Y CD CT CS
56	—ANO—	Y	Y
57	—ASO—	Y	Y
58	$>ASO_2$	Y	Y
59	—N=O	ND	S ND N O Y CD CT CS
60	—N(→O)=O	ND	S ND N O Y CD CT CS
61	—N=S	ND	S ND N O Y CD CT CS
62	—N(→O)=S	ND	S ND N O Y CD CT CS
63	—N=S→O	ND	S ND N O Y CD CT CS
64	—N=NH	ND	S ND N O Y CD CT CS
65	—N(→O)=NH	ND	S ND N O Y CD CT CS
66	$-N=N^+=N^-$	ND	S ND N O Y CD CT CS
67	$-N(\rightarrow O)=N^+=N^-$	ND	S ND N O Y CD CT CS
68	—N=	ND	I Br Cl F S ND N O Y CD CT CS
69	—N(→O)=	ND	I Br Cl F S ND N O Y CD CT CS
70	—S(→O)OH	S	S ND N O Y CD CT CS
71	$-S(\rightarrow O)_2OH$	S	S ND N O Y CD CT CS
72	>S→O	S	I Br Cl F S ND N O Y CD CT CS
73	$-S(\rightarrow O)_2$—	S	I Br Cl F S ND N O Y CD CT CS
74	—OH	O	S ND N O Y CD CT CS
75	—SH	S	S ND N O Y CD CT CS
76	$-NH_2$	N	S ND N O Y CD CT CS
77	—O—	O	I Br Cl F S ND N O Y CD CT CS

Table II (Continued)

no.	secondary component[a]	efferent nature	afferent nature
78	—S—	S	I Br Cl F S ND N O Y CD CT CS
79	—NH—	N	I Br Cl F S ND N O Y CD CT CS
80	>N—	N	I Br Cl F S ND N O Y CD CT CS
81	>N→O	N	I Br Cl F S ND N O Y CD CT CS
82	—F	F	*e*
83	—Cl	Cl	*e*
84	—Br	Br	*e*
85	—I	I	*e*
86	—D—[f]		*b*

[a] AC, aromatic carbon; ACH, aromatic carbon with hydrogen; AO, aromatic oxygen; AS, aromatic sulfur; AN, aromatic nitrogen; ANH, aromatic nitrogen with hydrogen. [b] Implies the linkage with double bond. [c] Denotes structure a below. [d] Denotes structure b below. [e] Implies one of S–CS attributes. [f] Dummy double bond.



a b

Table III. Attributes of Components

priority	symbol	meanings
1	I	iodine
2	Br	bromine
3	Cl	chlorine
4	F	fluorine
5	S	center atom[a] is sulfur
6	ND	center atom is nitrogen that has double bond
7	N	center atom is nitrogen that has only single bonds
8	O	center atom is oxygen
9	Y	center atom is aromatic carbon or nitrogen or sulfur
10	CD	center atom is carbon that has double bond
11	CT	center atom is carbon that has triple bond
12	CS	center atom is carbon that has only single bonds

[a] The atom that has free bond in a component, as underlined: $(CH_3)_2\underline{C}H$-; CH_3-$\underline{S}$-; -$\underline{C{\equiv}C}$-.

of an unknown compound, while those in the other should not be used. The former and the latter are called positive and negative macrocomponents, respectively. The program used here is designed to deal with two or more of them simultaneously. Processing is also possible for macrocomponents that overlap each other. In addition, two or more substructures that do not overlap each other, if known to exist, can be used without any modifications of the substructures. In both cases, structures can be generated that consist of these positive macrocomponents each incorporated at an appropriate position. On the other hand, if two or more substructures are known to be incapable of coexisting with each other, such a requirement can be reflected on the structure generation process by the use of negative macrocomponents. Figure 1 illustrates two input examples using positive and negative components. Example 1 in Figure 1 shows an input format to be used when the positive macrocomponents [(1), (2), and (3)], which may overlap each other, are desired to be incorporated in the structures to be generated while all negative macrocomponents [(1) and (2)] are excluded. Example 2, on the other hand, illustrates an input format to generate candidate structures that contain all of the macrocomponents [(1), (2), and (3)] without overlapping between macrocomponents (1) and (2), while inhibiting the two negative macrocomponents, (1) and (2), from coexisting in any candidate structure.

Degradative Modification and Checking of Macrocomponents. These macrocomponents input as above are degraded into secondary and tertiary components and then used in the processes for generating component sets and the construction of structures (described later in detail). The macrocomponents shown for the above-mentioned input format (example 2 in Figure 1) are degraded as illustrated in Figure 2. The macrocomponent input is converted first into secondary compo-

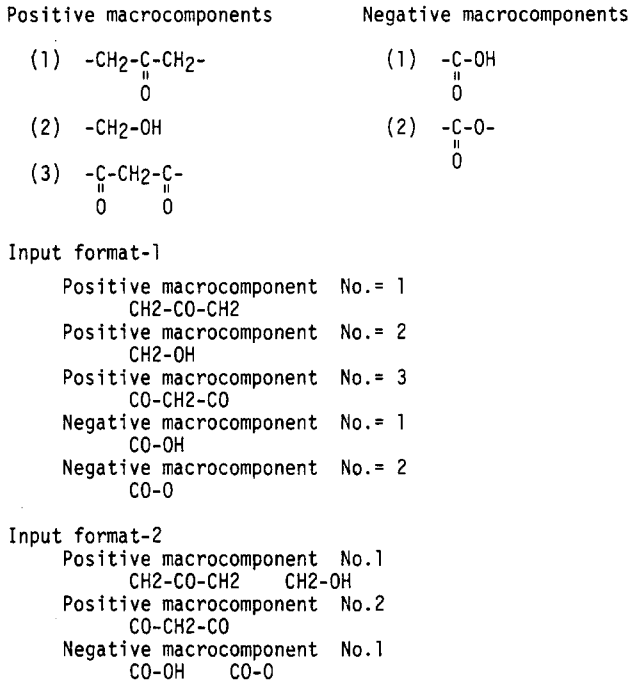


Figure 1. Examples of macrocomponent input format.

nents and the connectivity matrix and then into tertiary components. It is important here that any two macrocomponents should be treated on the same connectivity matrix in order to allow the two macrocomponents to coexist without overlapping at any position (1 in Figure 2). An AN is determined definitely for the CO (**24**) secondary component contained in positive macrocomponent 1 because th two bonds have already been occupied. Specifically, each of these bonds is connected with a component with an EN equal to CS, and therefore the **24** secondary component is assigned to the **335** tertiary component [–CO–(CS)]. On the other hand, as there remains a free single bond in the $-CH_2-$(**24**) secondary component in this macrocomponent, the AN cannot be fixed definitely. However, since one of the two bonds is connected with a >C═O component with an EN equal to CD, this component may have an AN of a priority rank equal to or above CD according to the above-mentioned definitions for tertiary components. Other macrocomponents are also converted into tertiary components by the same procedure as shown in Figure 2. Each of the macrocomponents treated on the same connectivity matrix is called a macrocomponent unit.

Then consistency checking is performed for each macrocomponent unit to assure the following two points. (1) The atomic composition of each macrocomponent unit should stay

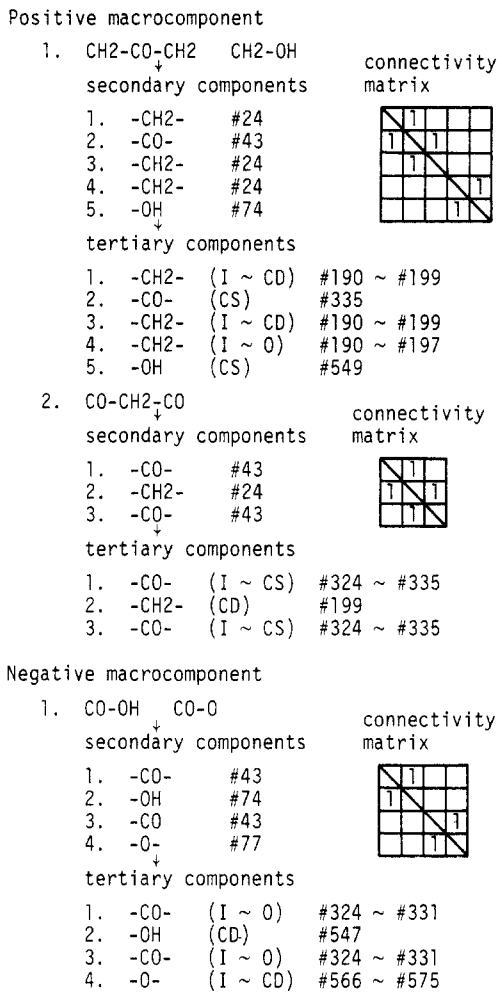


Figure 2. Degradation of macrocomponent.

within the limit fixed by the molecular formula of the test sample. (2) Each tertiary component in the macrocomponents should be included in the set of tertiary components (referred to as "surviving components"), which are selected on the basis of the molecular formula and spectral data of the unknown in CHEMICS. This means that in the CHEMICS system the components selected by a computer always take precedence over the macrocomponents input by the user. A macrocomponent that fails to meet the above two requirements will be rejected even if the user inputs it.

STRUCTURE GENERATION

In the following part of this paper, the structure generation procedure is explained mainly by the use of molecular formula as input data. The smaller the number of the components to be treated, the more simply the procedure can be explained. Therefore, analytical results of spectral data (^{13}C NMR) are used, when necessary, in addition to molecular formulas. The program is designed to select, from the **630** tertiary components defined previously, those that are consistent with the given molecular formula and spectral data. Here again, let us mention that a set of the tertiary components consistent with the molecular formula and spectral data of the unknown is called that of surviving tertiary components and the tertiary ones minus the AN's are called surviving secondary components.

The procedure for spectral analysis and the utilization of the analytical results for the structure generation procedure will be described in detail in our next paper.

Generation of Component Sets. Component sets to be used in structure generation are hierarchically produced at different

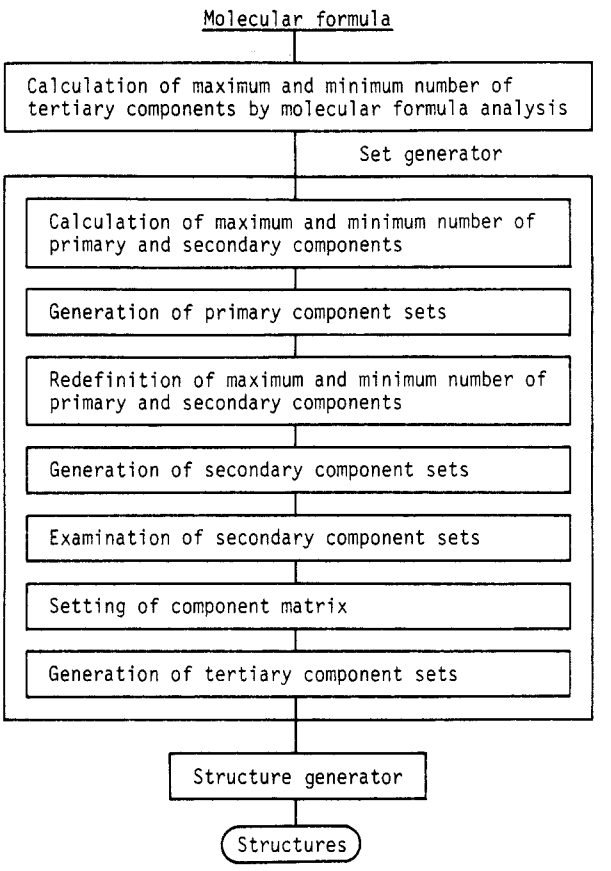


Figure 3. Block diagram of generation of component sets.

levels corresponding to the primary, secondary, and tertiary components on the basis of the analyses of structural information of an unknown.

(A) Primary and Secondary Component Sets. The basic algorithm for generating primary and secondary component sets has already been reported in detail.[6] Here, we present equations to be satisfied for generating primary and secondary component sets, together with some comments on them.

Step 1. On the basis of the atomic composition given by the molecular formula and the index of hydrogen deficiency, the set generator determines the maximum possible number of primary and secondary components as shown in Figure 3. In some cases, the maximum and minimum secondary component vectors may be modified in terms of the maximum value of the surviving tertiary component vectors that are consistent with the molecular formula and spectral data.[7] The minimum secondary component vector is further modified in terms of positive macrocomponents. Specifically, for each of the above-mentioned macrocomponent units, the number of secondary components contained is examined for each kind of secondary component, and the maximum value is used as an element of the corresponding minimum secondary component vector.

Step 2. All primary component sets, P's, that satisfy the equation

$$\sum_{i=1}^{15} X_{ij} P_i = M_j \qquad (j = 1\text{–}9) \tag{1}$$

where P_i represents the ith element of the primary component set, X_{ij} the number of the elements of j in the ith component, and M_j the number of the jth element in the molecular formula input, are generated.

Step 3. By use of the primary component sets, P's, generated in step 2, all secondary component sets, S's, are generated that satisfy eq 2 within the range of the elements of the maximum

Molecular formula

$C_7H_7O_3N$

^{13}C-NMR Data

	chemical shift(ppm)	intensity	multiplicity
1	55.9	49	4
2	114.0	93	2
3	125.7	94	2
4	141.5	3	1
5	164.7	12	1

NO.	TERT-COMP	MAX	MIN	NO.	TERT-COMP	MAX	MIN
1	CH3O- (ND)	1	0	34	>AO (Y)	3	0
2	CH3O- (N)	1	0	35	>AN (Y)	1	0
3	CH3O- (O)	1	0	36	>AN- (O)	1	0
4	CH3O- (Y)	1	0	37	>AN- (Y)	1	0
5	CH3O- (CD)	1	0	38	>AN- (CD)	1	0
6	CH3O- (CS)	1	0	39	>AN- (CS)	1	0
7	-CH< (N)	3	0	40	>ANO (Y)	1	0
8	-N≡C (O)	1	0	41	-N=O (O)	1	0
9	-N≡C (Y)	1	0	42	-N=O (Y)	1	0
10	-N≡C (CD)	1	0	43	-N=O (CD)	1	0
11	O=O= (*)	2	0	44	-NO2 (O)	1	0
12	=O= (*)	2	0	45	-NO2 (Y)	1	0
13	-CH= (ND)	2	0	46	-NO2 (CD)	1	0
14	-CH= (N)	3	0	47	-N= (O)	1	0
15	-CH= (O)	4	0	48	-N= (Y)	1	0
16	-CH= (Y)	4	0	49	-N= (CD)	1	0
17	-CH= (CD)	4	0	50	-NO= (O)	1	0
18	-CH= (CS)	4	0	51	-NO= (Y)	1	0
19	>C= (ND)	2	0	52	-NO= (CD)	1	0
20	>C= (N)	2	0	53	-O- (ND)	2	0
21	>C= (O)	2	0	54	-O- (N)	3	0
22	>C= (Y)	2	0	55	-O- (O)	3	0
23	>C= (CD)	2	0	56	-O- (Y)	3	0
24	>C= (CS)	2	0	57	-O- (CD)	3	0
25	O=C< (ND)	2	0	58	-O- (CS)	3	0
26	O=C< (N)	2	0	59	>N- (O)	1	0
27	O=C< (O)	1	0	60	>N- (Y)	1	0
28	>AC- (ND)	2	0	61	>N- (CD)	1	0
29	>AC- (N)	1	0	62	>N- (CS)	1	0
30	>AC- (O)	2	0	63	-NO< (O)	1	0
31	>AC- (Y)	1	0	64	-NO< (Y)	1	0
32	>AC- (CS)	2	0	65	-NO< (CD)	1	0
33	>ACH (Y)	4	0	66	-NO< (CS)	1	0

Figure 4. Input data and surviving tertiary components.

and minimum secondary components. To perform this, each element of the maximum secondary component vector used for generating these S's was redefined by the following procedures: (i) The number of primary components contained in a surviving secondary component is determined for each kind of primary component, and each corresponding element in the primary component set obtained in step 2 is divided by it. (ii) The smallest of the values of the primary components obtained above is used as the maximum value of the secondary components under the primary component set. By use of this procedure, the primary component set in question is efficiently converted into a secondary component set

$$\sum_{i=1}^{k} Y_{ij}S_i = P_j \qquad (j = 1\text{–}15) \tag{2}$$

where k stands for the number of surviving secondary components, Y_{ij} for the number of the jth primary components contained in the ith secondary component, and S_i for the ith element of the secondary component vector.

Step 4. The secondary component set generated is reviewed from a graph-theoretical point of view[6] and is based on the input spectral data.[7]

(B) Tertiary Component Set. The generation of tertiary component sets is carried out by using the secondary component sets produced above along with surviving tertiary components selected by the molecular formula analysis.

Tertiary components that survived the analysis of the molecular formula and ^{13}C NMR data are, as an example, shown in Figure 4 together with their maximum and minimum values. The surviving tertiary components are converted into secondary component vectors and AN vectors as illustrated in Figure 5. These are used, along with the secondary component sets previously generated, to establish new secondary component

	Sec.comp.	Sec.comp.#	AN
1	CH3O-	15	ND, N, O, Y, CD, CS
2	C≡N-	30	O, Y, CD
3	O=C=	32	*
4	=C=	36	*
5	-CH=	37	ND, N, O, Y, CD, CS
6	>C=	38	ND, N, O, Y, CD, CS
7	>C=O	43	ND, N, O
8	>AC-	47	ND, N, O, Y, CS
9	>ACH	48	Y
10	>AO	51	Y
11	>AN	53	Y
12	>AN-	55	O, Y, CD, CS
13	>ANO	56	Y
14	-N=O	59	O, Y, CD
15	-NO2	60	O, Y, CD
16	-N=	68	O, Y, CD
17	-NO=	69	O, Y, CD
18	-O-	77	ND, N, O, Y, CD, CS
19	>N-	80	O, Y, CD, CS
20	-NO<	81	O, Y, CD, CS

Figure 5. Conversion of the surviving tertiary components into the secondary components and the AN's.

vectors and a component matrix. The elements of the new vectors are represented by the secondary component numbers, and their dimensions are expressed by the number of secondary components in the sets. These elements and dimensions are not changed until all tertiary component sets corresponding to these secondary component sets are completely generated.

(1) Establishment of Component Matrix. The component matrix consists of columns identical with the secondary component vectors and rows corresponding to their AN's. From the surviving tertiary components, secondary components are selected that are equal to the elements of the secondary component vector, and the elements of the component matrix that correspond to their AN's are fixed at unity. For example, it is assumed here that a secondary component set of ($-OCH_3$, >AC–, >AC–, >ACH, >ACH, >ACH, >ACH, $-NO_2$) is obtained. In such a case, for example, $-OCH_3$ (secondary component **15**) in this secondary component set is selected from the results shown in Figure 5, and the component matrix elements corresponding to the AN's that it may have are fixed at unity. Figure 6a shows a component matrix that was determined for this secondary component set. If the secondary component vector contains no secondary component that has an EN equal to the AN relevant to the unity element in the component matrix, all elements of the column corresponding to the AN in this component matrix are replaced with zeros (Figure 6b).

(2) Generation of Tertiary Component Set. An AN vector that makes a pair with the secondary component vector is used to generate a tertiary component set. Such an AN vector is scanned to provide all possible tertiary component sets that correspond to the secondary component set. Only such AN's that give unity elements in corresponding rows of the component matrix are adopted as the elements of this AN vector. Under these conditions, the AN vector is scanned successively from the AN of the highest priority rank to others of lower ranks. Figure 7 illustrates the process of scanning the AN vector based on the component matrix shown in Figure 6b. If for an element there is an identical element on the left within the secondary component vector during this AN vector scanning process, the corresponding element in the AN vector has an attribute that is equal to or below that of the element on the left. This procedure can prevent identical tertiary component sets from being generated repeatedly.

(3) Processing of Macrocomponent for Tertiary Component Generation. When two or more positive macrocomponent units are already input, the component matrix is modified as described below, using one of the units, before the tertiary

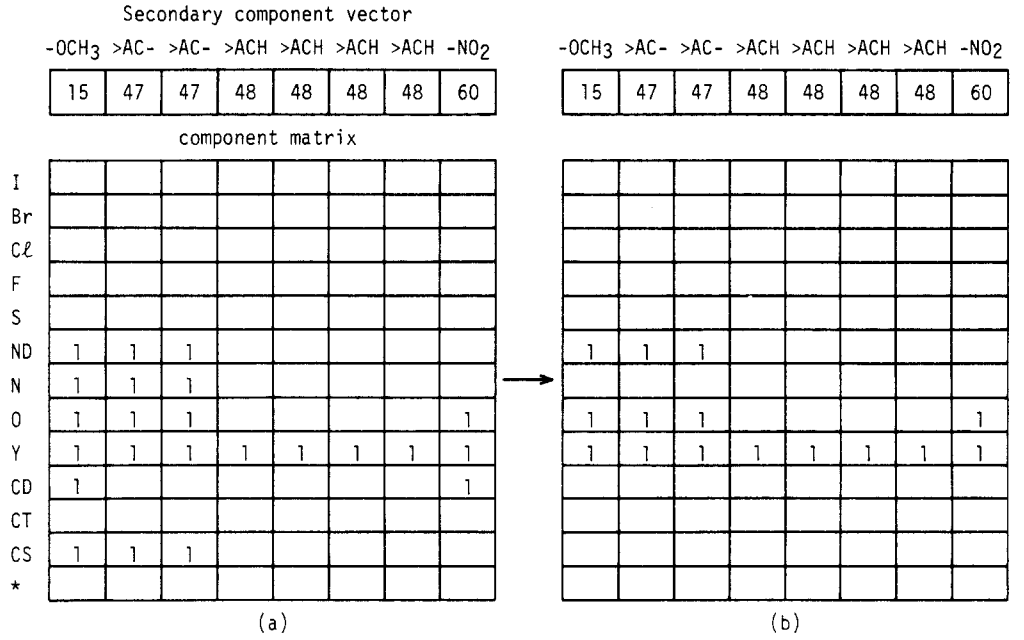


Figure 6. Establishment of component matrix.

	-OCH3	>AC-	>AC-	>ACH	>ACH	>ACH	>ACH	-NO2
1	ND,	ND,	ND,	Y,	Y,	Y,	Y,	O
2	ND,	ND,	ND,	Y,	Y,	Y,	Y,	Y
3	ND,	ND,	O,	Y,	Y,	Y,	Y,	O
4	ND,	ND,	O,	Y,	Y,	Y,	Y,	Y
5	ND,	ND,	Y,	Y,	Y,	Y,	Y,	O
6	ND,	ND,	Y,	Y,	Y,	Y,	Y,	Y
7	ND,	O,	O,	Y,	Y,	Y,	Y,	O
8	ND,	O,	O,	Y,	Y,	Y,	Y,	Y
9	ND,	O,	Y,	Y,	Y,	Y,	Y,	O
10	ND,	O,	Y,	Y,	Y,	Y,	Y,	Y
11	ND,	Y,	Y,	Y,	Y,	Y,	Y,	O
12	ND,	Y,	Y,	Y,	Y,	Y,	Y,	Y
13	O,	Y,	Y,	Y,	Y,	Y,	Y,	O
14	O,	Y,	Y,	Y,	Y,	Y,	Y,	Y
15	Y,	Y,	Y,	Y,	Y,	Y,	Y,	O
16	Y,	Y,	Y,	Y,	Y,	Y,	Y,	Y

Figure 7. Scanning of AN vector.

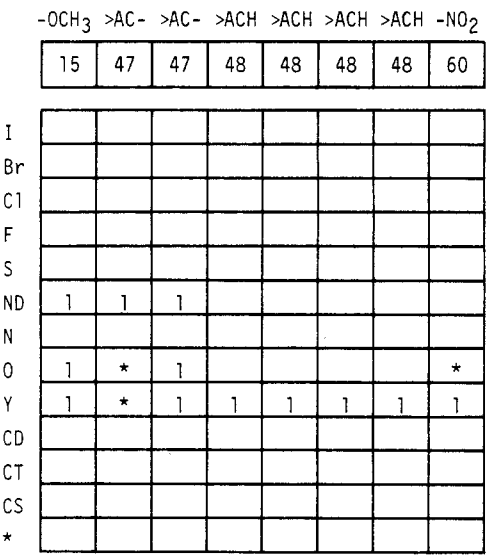


Figure 8. Modification of component matrix by macrocomponent.

component set generation procedure is performed. Here, it is assumed that >AC–NO_2 is used as an input macrocomponent. This macrocomponent unit is degraded into the level of tertiary components by the above-mentioned procedure.

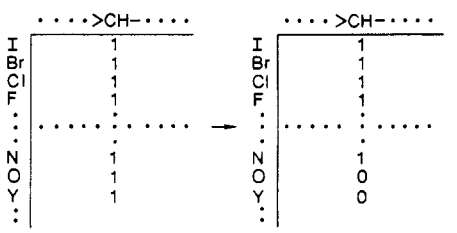
>AC–(ND) –NO_2(Y)

(i) Unnecessary elements in the component matrix are replaced with zeros in order to fix the bonding relationship among the secondary components that are contained in the macrocomponent. In this example, an AN can be determined definitely for each of the component. All asterisks in the component matrix that correspond to the unnecessary tertiary components are replaced with zeros as illustrated in Figure 8. The above operation allows this macrocomponent to be contained in every tertiary component set generated.

In addition, the following rules are applied in dealing with macrocomponents.

(ii) Only when the macrocomponent contains a component for which an AN can be fixed definitely is the AN on the right allowed to be higher in priority ranking than the one on the left, even if there is an identical component on the left in the second component vector.

(iii) When the AN's of the components in the macrocomponent cannot be determined definitely, on the other hand, only the elements of the component matrix that do not correspond to them are replaced with zeros. If, for instance, the component without definite AN is >CH=(I–N), all of the elements from O to CS are accordingly replaced with zeros.



(iv) If a similar kind of secondary component is contained in the macrocomponent, the one for which the number of possible AN's is smaller must be on the left. Thus >CH–(N), >CH(I–D) and >CH–(I–N) are arranged

	····>CH-	>CH-	>CH-····
I		1	1
Br		1	1
Cl		1	1
F		1	1
S		1	1
ND		1	1
N	1		1
O			
Y			
CD			
CT			
CS			
*			

(4) Checking of Tertiary Component Set. By the above operations, all tertiary component sets that contain one positive macrocomponent unit and are consistent with all the members of surviving tertiary components are generated completely, with any two of them being not identical with each other. These sets are further checked with respect to the following requirements. Only the sets that meet them are then sent to the structure construction step.

(i) Each tertiary component contained in the positive macrocomponent units that were not used in generating the tertiary component set (including those for which AN's cannot be fixed definitely) should have a counterpart to make a pair in the tertiary component set.

(ii) The number of tertiary components for each kind contained in the set should not exceed the maximum number of the corresponding surviving tertiary components.

(iii) The set should contain tertiary components of the minimum values among the surviving components, and their number should be at least equal to the value.

(iv) Given a tertiary component with an AN in the set, at least one tertiary component with an EN equal to that AN should exist in the set in addition to the former.

(v) The AN and EN of the highest priority rank in the set should be equal to each other.

(vi) When the AN and EN of the lowest priority rank in the set are compared with each other, the former should be equal to the latter or should be of a rank higher than that of the latter.

(vii) The number of single bonds of a tertiary component should be smaller than the total number of tertiary components having an EN that is lower in priority ranking than the AN of that tertiary component.

(viii) The total number of bonds of the tertiary components having the same EN should be greater than the number of tertiary components that have an AN equal to that EN.

(ix) The total number of single bonds of the tertiary components having the same EN should be greater than the number of tertiary components that have an AN equal to that EN and have single bond.

(x) The component set should be consistent with the input spectral data.[7]

Structure Generation. The structure construction step is designed to generate all possible structures that consist of positive macrocomponents but do not contain negative ones. The algorithm for this structure construction is based on the connectivity stack method.[4] For this study, various check routines were set up to allow the connectivity stack method to work efficiently, and the set reduction method (see Appendix) was improved to simplify the processing of macrocomponents.

The multiple linkage can be dealt with by several components, and thus zero or one is given to the stack.

(A) Simplification of Isomorphism Check. Isomorphism checking is an important part of the connectivity stack method for efficient structure generation. In general, however, this check represents the major time-consuming step in this method because it is necessary to carry out exhaustive permutations

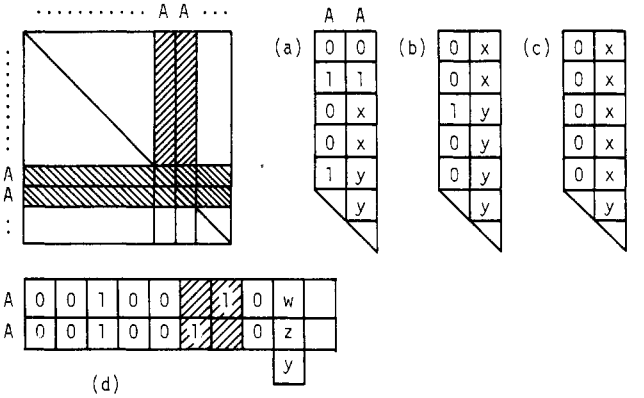


Figure 9. Cases where isomorphic structures are generated inevitably.

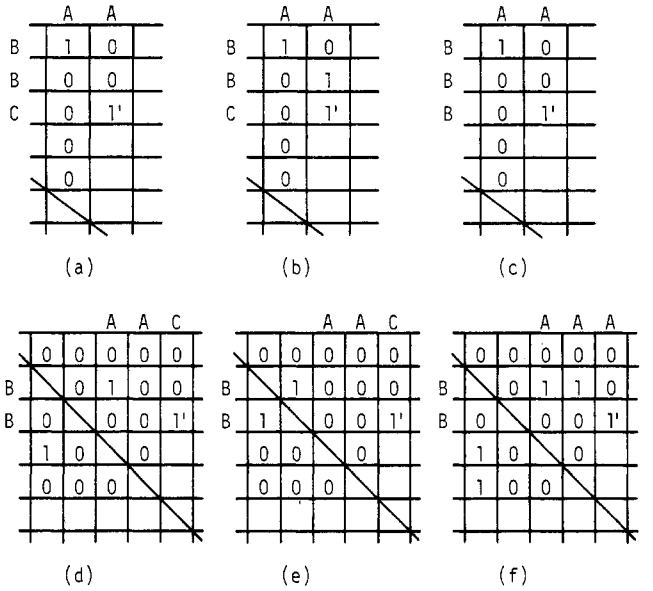


Figure 10. Cases where isomorphic structures are not generated.

with respect to the same kind of components contained in the component set. For this study, some new operations as shown below have recently been worked out to avoid the isomorphism checks as far as possible and incorporated in the stack method.

(i) Isomorphism checking is not performed if it is known in advance that isomorphic structures will be generated. Figure 9 illustrates some cases where isomorphic structures are generated inevitably. In parts a, b, and c of Figure 9, generation of a linkage at the element denoted by x will always result in generation of isomorphic structures. In these cases, the linkage jumps x and extends to y. For parts a, b, and c of Figure 9, the requirements for excluding the possibility of isomorphism are summarized below.

(a) If the ith row, which has a linkage in the column on the right, and rows above it are completely identical with the column on the left, the $i + 1$ linkage should be in the $i + 1$ row below it in the column on the left.

(b) The first linkage that occurs in the column on the right should not be above the first linkage in the column on the left.

(c) If there is no linkage in the column on the left, there should not be a linkage above y. However, the number of linkages generated at y is not restricted. For Figure 9d, the requirement for avoiding isomorphism is as follows.

(d) If the upper and lower rows are identical with each other in the columns on the left of the shaded portion, the possibility of a inkage to z depends on w: If $w = 0$, a linkage is not produced. If $w = 1$, a linkage is produced.

(ii) No isomorphism checks are performed if it is clear that no isomorphic structures are generated. Such cases are given below.

(a) Component A in a column of the connectivity matrix accompanies another component of the same type existing in the column on the left, and the first linkage is made in the column that contains A. This does not apply if a component in the row accompanies a component of the same type existing in a row above it. Thus, in Figure 10a, no isomorphic structures are generated if 1′ is the first linkage to A. In parts b and c of Figure 10, on the other hand, isomorphic structures may be generated because 1′ is not the first linkage to A.

(b) Component B in a row of the connectivity matrix accompanies a component of the same type existing in a row above it, and the first linkage is made in the column that contains B. This does not apply if a component of the same type as one of those in this row is contained in a column on the left. Thus, in Figure 10d, no isomorphic structures are generated if 1′ is the first linkage to B. Isomorphic structures may be generated, however, if 1′ is generated as shown in part e or f of Figure 10.

(c) The components in a column of the connectivity matrix accompany no components of the same type in the columns on the right, and the components in a row accompany no components of the same type in the rows above it.

(B) Generation of Linkage and Requirement for It. As stated previously, restrictions concerning the priority ranks of AN's and EN's are imposed on the linkages between tertiary components, which serve as basic units in constructing structures by CHEMICS. The formation of a linkage between two components, A and B, depends on whether the AN's of each of them include EN's of the other in a manner that meets the restrictive requirements given previously. In the structure generator of CHEMICS, the element at the growing head of the connectivity stack that is being constructed is checked in terms of these restrictions, and a linkage to the element is generated if the requirements are met. This inspection judges among the following three cases: (1) linkage is possible, (2) linkage is impossible, and (3) linkage is absolutely impossible on the connectivity matrix up from the tertiary component set in question. In the third case, the relevant element of the connectivity stack is marked so that it will not be checked repeatedly (described later in detail).

(C) Absolute Linkage and Absolute Nonlinkage. A linkage that is maintained throughout the structure generating process and that which should not be generated during the process are respectively referred to as an absolute linkage and an absolute nonlinkage. They can be established under the conditions presented below.

Absolute Linkage. (i) A component set containing a component with an AN includes only one other component that has an EN equal to that AN. (ii) A component set containing a component with an EN includes only one other component that has an AN equal to or higher in priority ranking than that EN. To distinguish this linkage from the others, "−1" is given as the element in the connectivity stack.

Absolute Nonlinkage. (i) This corresponds to the case that the check concerning "generation of linkage and requirement for it" for the element in a connectivity stack ends at (3). (ii) Absolute nonlinkage also refers to a connectivity stack element relevant to a component in which the number of remaining bonds becomes zero as a result of the formation of an absolute linkage. The absolute nonlinkage in such a case is expressed by "−2" in the stack so that it is distinguished from no linkage expressed by zero.

Figure 11 illustrates a tertiary component set that is generated on the basis of the input data given in Figure 4, along with the identification of absolute linkage and absolute nonlinkage to the relevant connectivity matrix. As these linkages are identified, it becomes unnecessary to perform various checks of the relevant elements in the connectivity stack. As a result, the number of stack dimensions to be actually treated diminishes from 28 to 15, allowing the processing to be performed efficiently.

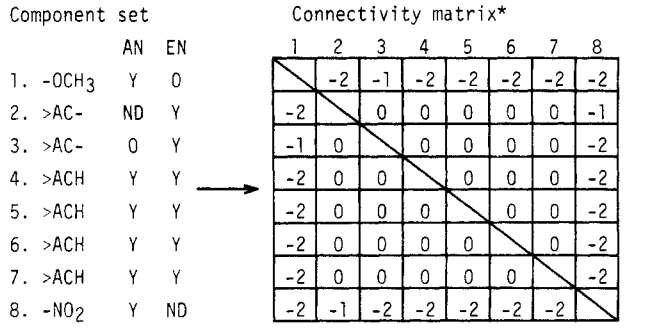


Figure 11. Arrangement of absolute and nonabsolute linkages.

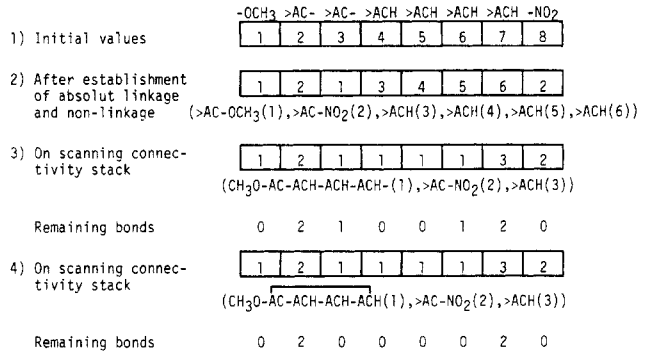


Figure 12. Checking vector for separated structures.

(D) Checking of Separated Structure. A structure required by CHEMICS should be given in the form of a complete graph. It is necessary, therefore, to find separated structures in the course of the structure generation process. To simplify this checking, a separated structure checking vector is used that shows the progress of the structure construction process. Such a vector is illustrated in Figure 12. The figures in the vector represent serial numbers given in relation to a connection graph. It is judged that there exists a separated structure when the total number of bonds remaining in the components with the same serial number becomes zero in this vector. The structure shown in Figure 12(3) is judged not to be separated, whereas that in (4) is judged to be separated and through subsequent operation will be pruned.

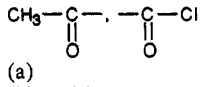
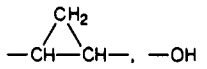
(E) Checking of Generated Structure. The structures generated as above are checked to determine whether they meet the following requirements. Those meeting them are adopted as final candidates that are consistent with the molecular formula and spectral data of the unknown.

(i) All elements in the separated structure checking vector should be 1, and the number of bonds remaining in each component should be zero.

(ii) In the case of a structure containing aromatic components, all aromatic components should be incorporated in the form of ring members, and one definite substructure consisting of aromatic components should have $4n + 2$ or $4n$ π-electrons ($n = 1, 2, 3, \ldots$). The number of π-electrons, however, should not be 4 or 8.

(iii) Rings consisting of aromatic components should have

Table IV. Input Data and Output Results

	molecular formula	macrocomponents	no. of structures	CPU time
1	C_6H_9NO	(a)	37491	47 h, 4 min, 48 s
		(b) positive 1 >C=O positive 2 —CH=CH_2	131	30 min, 26 s
		(c) positive 1 —CH=CH_2 negative 1 —OH	1613	36 h, 5 min, 20 s
2	C_5H_7NO	(a)	7075	6 h, 38 min, 9 s
		(b) positive 1 >C=O positive 2 —CH=CH_2 negative 1 NH negative 2 NH_2	13	6 min, 20 s
3	C_4H_7NO	(a)	802	23 min, 56 s
		(b) positive 1 —NH—CO— negative 1 NH_2—CO—	5	34 s
4	$C_4H_5O_2Cl$	(a)	907	14 min, 0 s
		(b) positive 1		
			1	6 s
5	C_3H_7NO	(a)	87	52 s
		(b) positive 1 —NH_2 negative 1		
			25	44 s

$4n + 1$, $4n + 2$, or $4n + 3$ π-electrons ($n = 0, 1, 2, \ldots$).

(iv) Any two rings consisting of aromatic components should not contain three or more aromatic components in common.

(F) Processing of Macrocomponent in Structure Generation Process. The existence of positive macrocomponent or the absence of negative macrocomponent is checked by a set reduction method modified by the authors (see Appendix), a method for substructure retrieval.

The CPU time for set reduction will increase if the checking of macrocomponents is performed only after the candidate structure is generated. The checking, therefore, is conducted as follows.

Each macrocomponent unit is always defined in the form of a submatrix in the connectivity matrix. As shown in the example in Figure 13a, when tertiary components are fixed on a connectivity matrix and a positive macrocomponent of AC–NO_2 is input, the latter is degraded into >AC–(ND) and –NO_2(Y). This is represented by the eight-row, eight-column minor matrix at the top of the left-hand side. In terms of the connectivity stack method, this positive macrocomponent should be contained in a structure represented by the stack that is extended to A. Thus, at the time when the connectivity stack reaches A (referred to as the checking point), a check is made to determine whether the positive macrocomponent exists there. If it exists, the subsequent operation of the stack method is continued, and if not, pruned. This checking method is still time-consuming. The tertiary components consisting of the input positive macrocomponent are transferred to the right in the connectivity matrix, as shown in Figure 13b. As a result, checking point A, which existed at the tail end of the stack before the transfer, comes to the head of the stack after the transfer. Thus, checking can be performed much earlier, resulting in pruning of unnecessary branches as well as a big decrease in the number of set reduction operations to be performed.

When two or more positive macrocomponents are input, the tertiary components in them are rearranged. An example is shown in Figure 13c. Suppose a positive macrocomponent of –ACH–ACH– is input in addition to >AC–NO_2; the sequence of tertiary components in them is represented as shown in the figure, with a checking point set up at A for >AC–NO_2 and at B for –ACH–ACH–.

Operations for negative macrocomponent are conducted in the same manner as for positive ones. The operations are only done for the negative macrocomponents that are possible to be defined in the tertiary component set.

RESULTS AND DISCUSSION

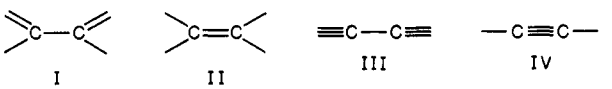
Table IV shows results of operation of the structure generator that is developed in the present study. Operation is performed on the basis of a molecular formula and either with or without assumed macrocomponents. It is revealed that in all cases input of macrocomponent results in a large decrease in CPU time and the number of candidates. The structure generator and macrocomponent operation procedure developed here are expected to serve effectively for the processing of data on substructures available from the chemist in his/her structure elucidation work. At present, however, the macrocomponent to be used should be expressed by using the secondary components defined in this report. Further studies are required to improve the system so that a structure of any given skeleton can be input in a more flexible format. Operation of the present program was carried out by using MV/2000DC supplied by Data General Japan Corp.

APPENDIX

Improvement in Set Reduction Method. The set reduction program, which is used in the CHEMICS system and performs the core role in the macrocomponent processing for structure generation, is based on an algorithm proposed by Sussenguth.[8] This algorithm, however, has two undesirable points as follows.

Paragraphs 1 and 2 below illustrate such problems together with solutions presented by the authors.

(1) Substructures I and III below cannot be distinguished from substructures II and IV, respectively.[9]



I II III IV

To avoid this inconvenience, sets of linkages are generated separately for each bond type. To simplify the representation of structures, furthermore, a hydrogen atom is not treated as a node but regarded as an attribute of a non-hydrogen atom

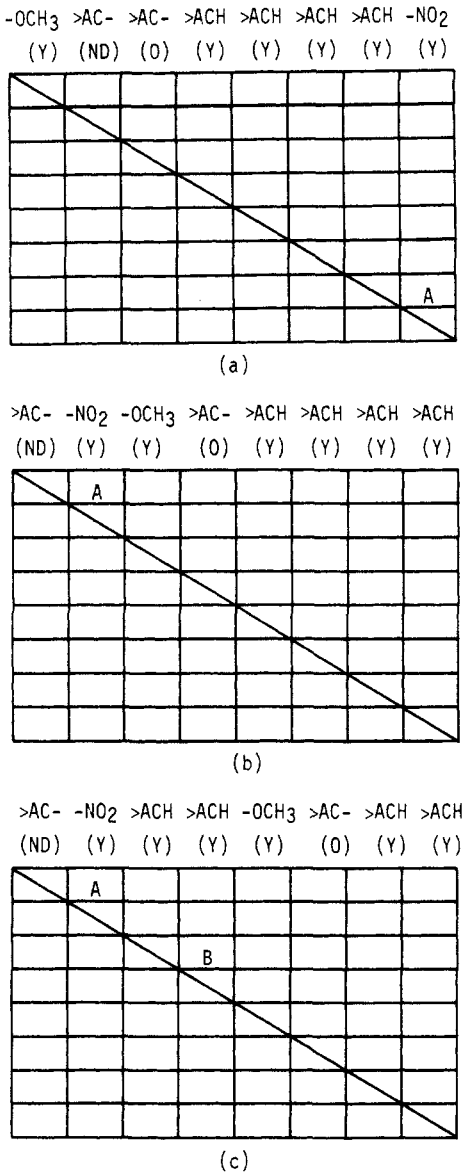
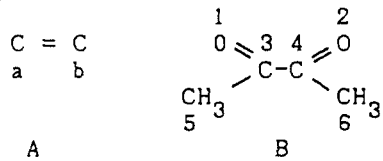


Figure 13. Rearrangement of checking point for macrocomponent.

to which it is connected. For example 1, a correct result, with substructure A not contained in substructure B, is obtained when this improved procedure is used.

Example 1



			set
Node value C	(a, b)	(3, 4, 5, 6)	1
Branch double	(a, b)	(1, 2, 3, 4)	2
Degree 1	(a, b)	(1, 2, 3, 4, 5, 6)	3
Hydrogen 0	(a, b)	(1, 2, 3, 4, 5, 6)	4
partition 1~4	(a, b)	(3, 4)	5
connectivity 5 double	(a, b)	(1, 2)	6
partition 5~6	(a, b)	(0)	7

The set for the full structure, set 7, becomes empty, leading to a desirable result with substructure A not contained in substructure B.

(2) In some cases, atoms in a substructure cannot be assigned to those in the full structure as shown in example 2.

Example 2

			set
Node value C	(b)	(2, 4, 6)	1
O	(a, b)	(1, 5)	2
Branch single	(a, b, c)	(1, 2, 3, 4, 5, 6)	3
Degree 1	(a, c)	(1, 2, 3, 4, 5, 6)	4
2	(b)	(2, 3, 4)	5
Hydrogen 0	(b)	(1, 2, 3, 4, 5, 6)	6
1	(a, c)	(1, 2, 4, 5, 6)	7
Partition 1~7	(a, c)	(1, 5)	8
	(b)	(2, 4)	9
connectivity			
8 single	(b)	(2, 4)	10
9 single	(a, c)	(1, 3, 5)	11
partition 8~11	(a, c)	(1, 5)	12
	(b)	(2, 4)	13
assignment	(a)	(1)	14
connectivity			
14 single	(b)	(2)	15
partition 12~15	(a)	(1)	16
	(b)	(2)	17
	(c)	(5)	18

This occurs because set 12 and set 14 are multiplied in negative forms in order to allow the set reduction process to converge earlier in carrying out the multiplication among sets 12–15. To avoid this inconvenience, new sets of linkages are regenerated for sets 16–18, and then multiplication among these sets is conducted.

Connectivity			
16 single	(b)	(2)	19
17 single	(a, c)	(1, 3)	20
18 single	(b)	(4)	21
partition 16~21	(a)	(1)	22
	(b)	(-)	23
	(c)	(-)	24

Consequently, it becomes impossible to define a set of atoms in full structure B that correspond to b and c in substructure A. Thus, a desirable result will be obtained when substructure A is not contained in full structure B.

In addition, the following improvements were made to allow the set reduction method to serve for the processing of macrocomponents performed during the structure construction. (1) Sets of node values are generated by using secondary components instead of non-hydrogen atoms. (2) Sets of branch values are not generated because only one type of bond is used in CHEMICS. (3) Sets concerning the number of ring members are not generated either. (4) Sets concerning linkages are not generated separately for each linkage type.

REFERENCES AND NOTES

(1) Sasaki, S.; Abe, H.; Hirota, Y.; Kudo, Y.; Ochiai, S.; Saito, K.; Yamasaki, T. *J. Chem. Inf. Comput. Sci.* **1978**, *18*, 211. Abe, H.; Fujiwara, I.; Nishimura, T.; Okuyama, T.; Kida, T.; Sasaki, S. *Comput. Enhanced Spectrosc.* **1983**, *1*, 55.
(2) Schelley, C. A.; Woodruff, H. B.; Shelling, C. R.; Munk, M. E. *Computer Assisted Structure Elucidation*; Smith, D. H., Ed.; ACS Symposium Series 54; American Chemical Society: Washington, DC, 1977; p 92.
(3) (a) Masinter, L. M.; Sridharan, N. S.; Lederberg, J.; Smith, D. H. *J. Am. Chem. Soc.* **1974**, *96*, 7714. (b) Masinter, L. M.; Sridharan, N. S.; Carhart, R. E.; Smith, D. H. *Ibid.* **1974**, *96*, 7714.
(4) Kudo, T.; Sasaki, S. *J. Chem. Doc.* **1974**, *14*, 200. Kudo, Y.; Sasaki, S. *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 43.
(5) Sasaki, S.; Fujiwara, I.; Abe, H.; Yamasaki, T. *Anal. Chim. Acta* **1984**, *122*, 87.
(6) Abe, H.; Okuyama, T.; Fujiwara, I.; Sasaki, S. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 220.
(7) Funatsu, K.; Ohta, Y.; Sasaki, S., submitted for publication.
(8) Sussenguth, E. H. *J. Chem. Doc.* **1965**, *5*, 36.
(9) Jochelson, N.; Mohr, C. M.; Reid, R. C. *J. Chem. Doc.* **1968**, *8*, 113.