

Figure 1. Flow chart of the NOE analysis.

## ALGORITHM

NOE data are applied after making clear the assignment relation between all atoms and signals by analysis of 2D-NMR. What "NOE data" means, in this paper, is the increasing intensity of the signal (%) calculated from the NOE chart, and it is input on the basis of the number of  $^1\text{H}$ -NMR peaks of the sample. That is, the number of irradiated signals, the number of signals on which NOE is observed, and the increasing intensity (%) together make one set to be input (Figure 1, step 1). After input of the NOE data, the program performs calculation of all candidate structures in turn, but a user can select the structure for the calculation.

CHEMICS gives the information of the component unit as shown in Tables 1 and 2. As examples are the following:

(1) The number of components which constitute the candidate structure (Table 1, (1)) are given.

(2) The number of secondary components which constitute the candidate structure (Table 1, (2)) are an attribute of the component as the bonding partner with which the secondary component should be connected (Table 1 (3)).<sup>1</sup>

(3) The number of tertiary components which constitute the candidate structure (Table 1, (4)) are a connectivity matrix of the tertiary component. Elements of the connectivity matrix, 1 or -1, indicate the bonding of each component and 0 or -2 indicate nonbonding (Table 1, (5)).<sup>1</sup>

(4) When the center atom of the tertiary component is carbon and the  $^{13}\text{C}$ -NMR signal is assigned to it uniquely, 1 is given in Table 1, (6). If not so, the possible number of assignments is shown in Table 1, (7), and the information of the assignment is written in the next columns (Table 1, (8)).

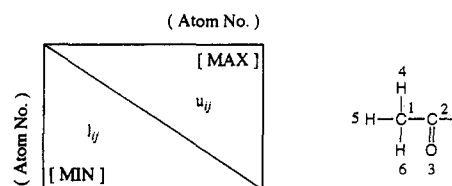
(5) NM matrices of proton and carbon NMR analyses for the tertiary component<sup>2</sup> show the possibility of assignment patterns of  $^{13}\text{C}$ - or  $^1\text{H}$ -NMR signals to every component (Table 2).

(6) Data of C-H COSY (path 1) are given.

Table 2. Example of NM-Matrix

NM-MatrixT( $^1\text{H}$ )								NM-MatrixT( $^{13}\text{C}$ )											
Signal no.								Signal no.											
1 2 3 4 5 6 7 8								1 2 3 4 5 6 7 8 9 10 11 12											
26 <sup>a</sup> 6 0 0 0 0 0 0 <sup>b</sup>								26 <sup>a</sup> 0 0 0 0 0 0 1 0 2 0 0 0 <sup>c</sup>											
3 0 0 3 0 0 0 0								134 1 0 0 0 0 0 0 0 0 0 1 0 0											
134 0 0 0 0 0 3 0 0								173 0 0 0 1 0 0 0 0 0 0 1 0											
173 0 0 0 3 0 0 0 0								199 0 0 0 0 0 1 0 0 0 0 0 0											
199 0 2 0 0 0 0 0 0								0 0 0 0 0 0 0 1 0 0 0 0											
0 0 0 0 2 0 0 0								201 0 0 0 0 0 1 0 0 0 0 0 0											
201 0 2 0 0 0 0 0 0								0 0 0 0 0 0 0 1 0 0 0 0											
0 0 2 0 0 0 0 0								0 0 0 0 0 0 0 0 0 0 0 1											
0 0 0 0 2 0 0 0																			

<sup>a</sup> Number of tertiary component. <sup>b</sup> Pattern of  $^1\text{H}$ -NMR signal assignments. Each element means the number of hydrogens assisted to each signal. <sup>c</sup> Pattern of  $^{13}\text{C}$ -NMR signal assignments. Each element means the number of carbons assigned to each signal.



	1	2	3	4	5	6
1	0.00	1.53	2.38	1.11	1.11	1.11
2	1.51	0.00	1.21	2.17	2.17	2.17
3	2.33	1.17	0.00	3.27	3.27	3.27
4	1.09	2.14	2.46	0.00	1.81	1.81
5	1.09	2.14	2.46	1.78	0.00	1.81
6	1.09	2.14	2.46	1.78	1.76	0.00

Figure 2. Distance matrix.

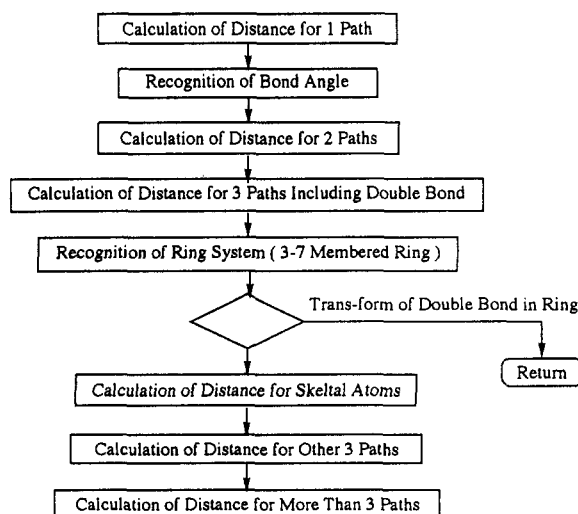


Figure 3. Creation of distance matrix.

After the data provided by CHEMICS are translated into the necessary form, it recognizes automatically the existence of the geometrical isomer for the candidates and it begins to build up the three-dimensional structures from all geometrical

isomers. When the user has the geometrical information of the unknown, the program does not make the calculation for isomers inconsistent with the information (Figure 1, step 3). The geometrical isomers of the candidate structures selected in this manner are converted to the three-dimensional structures.

The distance matrix such as (all atom)  $\times$  (all atom) on the upper left in Figure 2, has the maximum value  $u_{ij}$  and the minimum value  $l_{ij}$  at the upper right and the lower left of the matrix, respectively. For example, the structure on the upper right in Figure 2 is expressed as the matrix on the lower part in Figure 2. The distance matrix is established by the algorithm shown in Figure 3, and the standard values based on the atomic hybridization and a kind of bond are used as the bond length and the bond angle. The distance between atoms through three paths around the double bond, when the information of the geometrical isomer is specified, is defined on the basis of the information. That is, in the case of a cis-isomer, the dihedral angle  $\phi = 0$ ,  $d = u$  as the maximum value, and  $d = l$  as the minimum value in eq 1.

$$d_{il} = (p1 - p2 \cos \phi)^{1/2}$$

$$p1 = d_{ij}^2 + d_{jk}^2 + d_{kl}^2 - 2d_{ij}d_{jk} \cos \theta_1 - 2d_{jk}d_{kl} \cos \theta_2 + 2d_{ij}d_{kl} \cos \theta_1 \cos \theta_2 \quad (1)$$

$$p2 = 2d_{ij}d_{kl} \sin \theta_1 \sin \theta_2$$

$$\theta_1 = L_{ijk}, \quad \theta_2 = L_{jkl}$$

$$\text{max: } d_{ij} \leftarrow u_{ij} \quad \text{min: } d_{ij} \leftarrow l_{ij}$$

In the case of the trans-isomer,  $\phi = \pi$ ,  $d = u$  as the maximum value, and  $d = l$  as the minimum value. The distance between atoms through three paths in other cases other than the cases mentioned above are calculated by eq 1. At this time, for the dihedral angle  $\phi = q$ , the maximum value =  $u$  for the maximum distance calculation, and for  $\phi = 0$ , the minimum value = 1 for the minimum distance calculation.

After the distances between all atoms through three paths are defined, the distance between atoms through more than three paths are defined by eq 2. That is, the maximum value

$$u_{im} = u_{il} + u_{lm}$$

$$l_{im} = \max(\text{sum of van der Waals radii}, l_{il} - u_{lm}) \quad (2)$$

is equal to the sum of the distance for three paths and the maximum value for the fourth path and minimum value are chosen as the larger value between value of the distance for three paths minus two maximum values of the fourth path and the sum of the van der Waals radii of the atoms.

In the same manner, the path is increased one by one until the distances between all atoms are obtained.

Recognition of the ring structure is worthy of special mention here. Rough distances between skeletal atoms through more than two paths for the nonaromatic rings such as six- and seven-membered rings are calculated by MM2PP,<sup>6</sup> and the values are defined in advance as data in the program. If the ring structure is found by the ring perception program, the maximum and minimum values of distance are defined by these data. In the case of aromatic rings, the distance between each atom, which is the skeletal atom and the atom bonded to the skeleton, can be defined to have a narrow range between the maximum and the minimum values. Furthermore, when the double bond is in the ring of less than seven members and

it is transformed, this geometrical isomer is discarded and the next isomer is examined. If the isomer has a cumulene structure in the ring, it is also discarded.

In the following step, a check is made for the distance matrix by NOE data (Figure 1, step 5). First, a ratio  $r_{ij}$  of the distance between hydrogens (on which NOE data are observed) is obtained by  $I_{ij}$ , the increasing signal intensity (%) in Figure 1, step 1. That is, suppose that a ratio  $r_{ab}$  of the distance between a pair of hydrogens ( $H_a, H_b$ ) is equal to 1.00 and the other ratios of distances for other pairs are obtained by eq 3 on the basis of  $I_{ab}$ .

$$r_{ij} = (I_{ab}/I_{ij})^{1/6} \quad (3)$$

For example

$$\begin{array}{ccc} i_{ab} = 12.9\%, & I_{cd} = 9.8\%, & I_{ef} = 5.5\% \\ \downarrow & \downarrow & \downarrow \\ 1.00 \rightarrow & (12.9/9.8)^{1/6} & (12.9/5.5)^{1/6} \\ & = 1.05 & = 1.15 \\ r_{ab} = 1.00 & r_{cd} = 1.05 & r_{ef} = 1.15 \end{array}$$

Using  $r_{ij}$  obtained above, in the case of the maximum value of the distance  $u_{ij}$ , a maximum value of the quotient obtained from the division of the distance between each atom  $u_{ij}$  by the ratio of distance  $r_{ij}$  is defined as a proportional constant, and the maximum value of the distance  $u'_{ij}$  is obtained by multiplying each ratio of distance  $r_{ij}$  by the proportional constant. In the case of the minimum value of distance  $l_{ij}$ , a maximum value of the quotient obtained from the division of the distance between each atom  $l_{ij}$  by the ratio of the distance  $r_{ij}$  is defined as a proportional constant, and the minimum value of distance  $l'_{ij}$  is obtained by multiplying each ratio of distance  $r_{ij}$  by the proportional constant (eq 4).

( $i, j$ : proton pair on which NOE data are observed)

$$R_{\min} = \min(u_{ij}/r_{ij})$$

$$u'_{ij} = R_{\min} r_{ij}$$

$$R_{\max} = \max(l_{ij}/r_{ij})$$

$$l'_{ij} = R_{\max} r_{ij} \quad (4)$$

In this processing, although the width between  $u_{ij}$  and  $l_{ij}$  in the distance matrix is restricted as shown in Figure 4, step 3, and an inconsistent structure, such as that shown in Figure 4, step 4, is discarded because it conflicts with the NOE data.

Generation and minimization of the three-dimensional coordinates are executed by the algorithm in the refs 3, 6, 8 and 9. After the three-dimensional coordinates are obtained, for correction of the preliminary coordinates parallel and rotational displacement is done before minimization. For example, in the case of a double bond, a double bond atom and the atom bonding to it are configured by the standard bond length on the plane, and the side chain extended from the atom of the  $\alpha$  position of the double bond is shifted by parallel and rotational displacement the same as that atom is done. In the case of a double bond in a ring structure, the double bond atom and the atom bonding to it are shifted to the plane which is formed by two double bond atoms and the centroid of the ring structure. The same processing mentioned above is applied to a triple bond, cumulene, and aromatic structure. Hydrogen atoms in a methyl, methylene, and methine group or methyl groups in *tert*-butyl and isopropyl groups and so on are shifted by rotational displacement as

Table 3. NMR Data of  $\beta$ -Ionone

$^{13}\text{C}$ -NMR Data							
no.	chemical shift	intensity	multiplicity	no.	chemical shift	intensity	multiplicity
1	197.6	10.0	1	7	33.7	10.0	1
2	142.3	10.0	2	8	33.1	10.0	3
3	135.7	10.0	1	9	28.3	20.0	4
4	135.2	10.0	1	10	26.8	10.0	4
5	131.2	10.0	2	11	21.3	10.0	4
6	39.4	10.0	3	12	18.3	10.0	3

$^1\text{H}$ -NMR (90 MHz) Data							
no.	chemical shift		intensity	no.	chemical shift		intensity
1	1.10	1.10	24.0	5	2.10	2.10	8.0
2	1.50	1.50	8.0	6	2.30	2.30	12.0
3	1.70	1.70	8.0	7	6.20	6.20	4.0
4	1.80	1.80	12.0	8	7.30	7.30	4.0

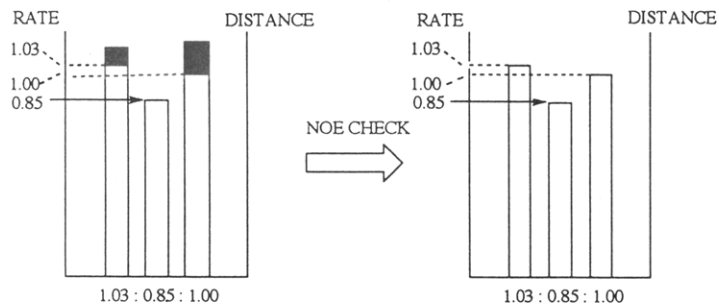
  

NOE Data							
no.	irr signal	obs signal	increase (%)	no.	irr signal	obs signal	increase (%)
1	1	8	12.9	5	4	7	7.2
2	1	7	9.8	6	4	5	2.9
3	1	2	5.5	7	6	8	8.7
4	4	8	5.0	8	6	7	5.2

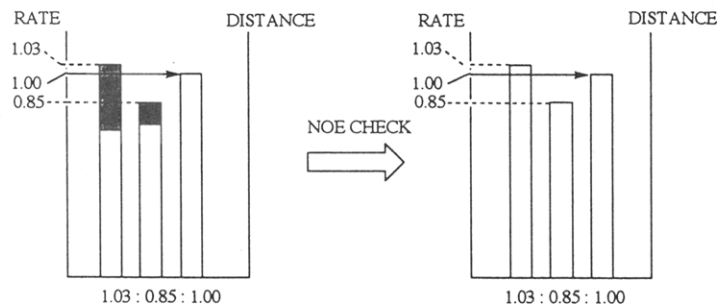
C-H COSY signal no.		H-H COSY signal no.		C-C connec. inf. signal no.		C-H COSY signal no.		H-H COSY signal no.		C-C connec. inf. signal no.	
H	C					H	C				
1	9	2	3	1	10	5	8				
2	6	3	5	4	11	6	10				
3	12	7	8	7	9	7	2				
4	11			1	2	8	5				

## MAXIMUM

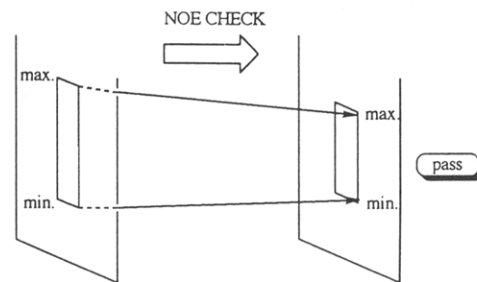


[ 1 ]

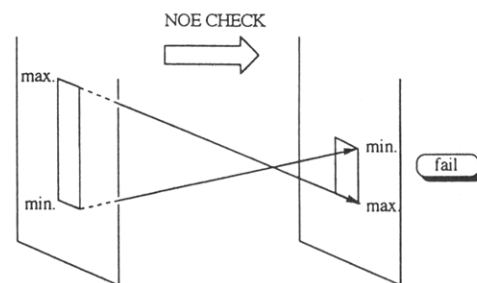
## MINIMUM



[ 2 ]



[ 3 ]



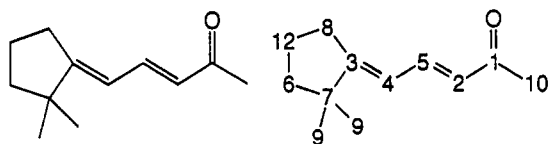
[ 4 ]

Figure 4. Examination of distance matrix by NOE data.

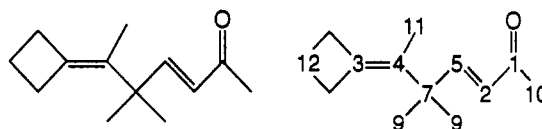
hydrogens or methyl groups being put away from each atom or group as evenly as possible. The coordinates are corrected as mentioned above and then are minimized. The minimiza-

tion is discontinued at the point at which the  $E$  value of the nonlinear minimizing function changes from a decrease to an increase.<sup>9</sup> The corrected coordinates are adapted as the final

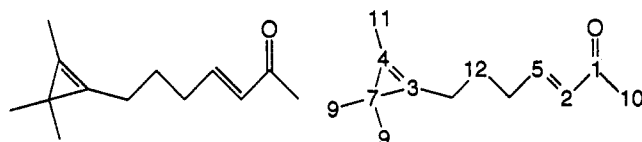
Structure 1



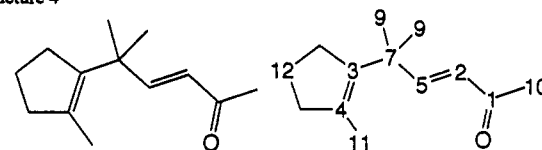
Structure 2



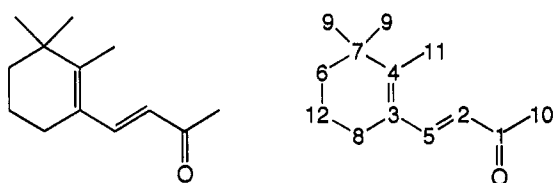
Structure 3



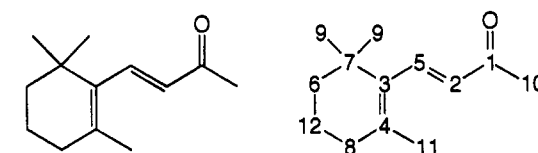
Structure 4



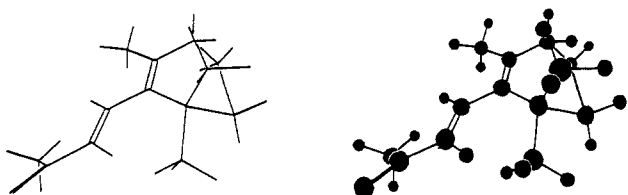
Structure 5



Structure 6



**Figure 5.** Candidate structures generated by CHEMICS for  $\beta$ -ionone as an unknown. Candidate structure and the corresponding  $^{13}\text{C}$ -NMR signal assignment.



**Figure 6.** Three-dimensional structure of  $\beta$ -ionone. (output from 3D-MOL<sup>6</sup> wire frame and ball-stick).

coordinates.

In the case of the derivation of the coordinates from the distance matrix restricted by NOE data, since the candidate structures, being unlike the true structure, although successful in the examination of Figure 1, step 5, are restricted by NOE data, it is expected that many pairs of atoms violate the van der Waals radii of each other. Thus, the candidate structure having more than a certain number of pairs of atoms which violate van der Waals radii is discarded as being inconsistent with NOE data. Concretely, a structure in which more than 1% of the total pair of atoms in the structure violate van der Waals radius of the others is output onto a file with the coordinates; then, a user deletes that structure on the basis of his judgment. Further the user can change the number of 1% arbitrarily.

## RESULTS

The output file is a Molfile format.

The present program was developed on superminicomputer MV-10000 (Data General Corp.) and written in FORTRAN 77. When NMR data of  $\beta$ -ionone (Table 3) are input, the six candidate structures in Figure 5 are obtained with the information of the NMR signal assignment by CHEMICS. Although complete assignments are not made for structures

2–4 ( $\text{C}^6$ ,  $\text{C}^8$ ) the structure is not changed by replacement of  $\text{C}^6$  and  $\text{C}^8$  on structure 2, so that the candidate structures to be processed by NOE data are eight. Each candidate structure has four geometrical isomers, because it has two double bonds. With the trans form of the double bond in the rings of structures 3–6 being discarded, twenty geometrical structures are obtained from eight candidate structures.

The results for each structure are summarized in Table 4. For structure 3, this structure should be discarded from the candidates because all structures passing the check of geometrical restriction fail in the check of van der Waals radius violation. Structure 1 is discarded in a case such that carbons bonded to  $\text{C}^3=\text{C}^4$  are a trans form of each other. For other structures, the cis form of carbons bonded to  $\text{C}^2=\text{C}^5$  have relatively high  $E$  values or are discarded by the check of van der Waals radius violation, and thus trans forms give satisfactory results.

From the above results, in examination of the candidate structures by analysis of NOE data it is supposed that seven geometrical isomers are promising among twenty structures, and a correct answer is included in them.

The correct three-dimensional structure has a result consistent with NOE data except that only a pair of hydrogens in the dimethyl group violates van der Waals radii.

## DISCUSSION

Results mentioned above seem to be generally satisfactory in the sense that *examination of candidate structures* is one of the objects of this program. As for the examination by NOE data, the expression that *it is expected right or not right* should be used, and it seems to be better not to decide upon whether it is successful in examination or not. In the minimization process in Figure 1, step 9, the final coordinates

Table 4. Result for  $\beta$ -Ionone

structure						possibility
1	1-2 = 5-4	cis	7-3 = 4-5	cis	$E = 33.9707$	-
	1-2 = 5-4	trans	7-3 = 4-5	cis	$E = 32.1104$	+
	1-2 = 5-4	cis	7-3 = 4-5	trans	fail in van der Waals radius check	-
	1-2 = 5-4	trans	7-3 = 4-5	trans	fail in van der Waals radius check	-
2	1-2 = 5-7	cis	6-3 = 4-7	cis	$E = 33.9509$	-
	1-2 = 5-7	trans	6-3 = 4-7	cis	$E = 33.2508$	+
	1-2 = 5-7	cis	6-3 = 4-7	trans	$E = 34.4470$	-
	1-2 = 5-7	trans	6-3 = 4-7	trans	$E = 33.1699$	+
3	1-2 = 5-6	cis	7-3 = 4-7	cis	fail in van der Waals radius check	-
	1-2 = 5-6	trans	7-3 = 4-7	cis	fail in van der Waals radius check	-
	1-2 = 5-8	cis	7-3 = 4-7	cis	fail in van der Waals radius check	-
	1-2 = 5-8	trans	7-3 = 4-7	cis	fail in van der Waals radius check	-
4	1-2 = 5-7	cis	7-3 = 4-6	trans	$E = 30.400$	-
	1-2 = 5-7	trans	7-3 = 4-6	trans	$E = 29.7179$	+
	1-2 = 5-7	cis	6-3 = 4-8	cis	$E = 30.2188$	-
	1-2 = 5-7	trans	6-3 = 4-8	cis	$E = 30.4076$	+
5	1-2 = 5-3	cis	5-3 = 4-7	trans	$E = 35.8782$	-
	1-2 = 5-3	trans	5-3 = 4-7	trans	$E = 40.2084$	+
6	1-2 = 5-3	cis	5-3 = 4-8	trans	fail in van der Waals radius check	-
	1-2 = 5-3	trans	5-3 = 4-8	trans	$E = 31.809$	+

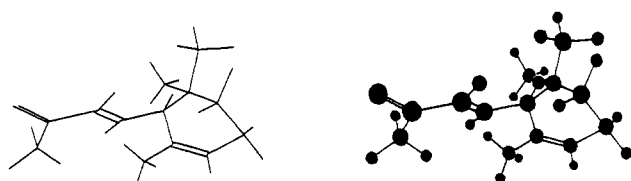
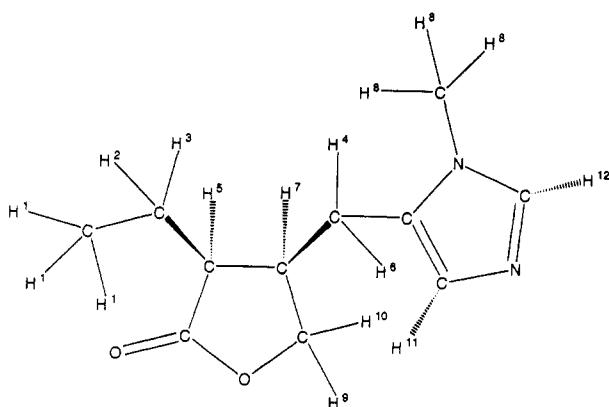
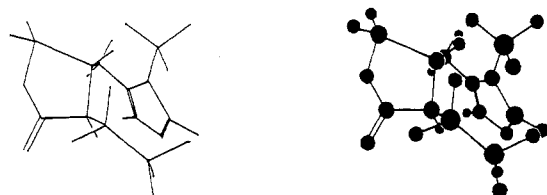
Figure 7. Three-dimensional structure of  $\alpha$ -ionone. (output from 3D-MOL<sup>6</sup> wire frame and ball-stick).

Figure 8. Pilocarpine.

Figure 9. Three-dimensional structure of pilocarpine. (output from 3D-MOL<sup>6</sup> wire frame and ball-stick).

of the structure are obtained by correction of the coordinates at the point at which the  $E$  value is minimum, but the  $E$  value could converge only when the total number of atoms is less than a few. If that number is more than a few, the  $E$  value does not converge, and it comes into a problem where the minimization stops. In the present program, the minimization calculation stops at the first minimum. However, the number of pairs of atoms which violate the van der Waals radii of each other changes somewhat, because the conformation of the

Table 5. NOE Data of Pilocarpine (500 MHz)

no.	irr signal	obs signal	increase (%)
1	11	9	1.3
2	11	7	3.3
3	10	12	7.2
4	10	9	2.7
5	10	7	5.6
6	10	5	2.4
7	9	12	6.89
8	9	11	3.0
9	9	10	2.4
10	9	4	1.8
11	8	12	17.9
12	8	7	1.8
13	8	6	1.8
14	8	4	1.8
15	7	12	8.0
16	7	11	6.6
17	7	10	4.3
18	7	5	10.1
19	6	12	9.1
20	6	11	1.1
21	6	4	25.6
22	6	2	5.4
23	5	12	7.0
24	5	11	1.1
25	5	10	2.0
26	5	7	9.5
27	5	3	2.8
28	5	2	1.8
29	5	1	1.1
30	4	12	8.0
31	4	11	2.6
32	4	9	2.2
33	4	7	2.0
34	4	6	22.3
35	3	12	8.6
36	3	5	3.1
37	3	2	22.2
38	3	1	14.6
39	2	12	7.7
40	2	6	4.9
41	2	5	2.3
42	2	3	20.1
43	1	12	6.9
44	1	5	3.6
45	1	3	2.7

structure is slightly different, dependent on the number of the minimization processes. This fact causes a problem also in the *analysis of conformation*. Thus, only a rough conformation

might be obtained by the present program.

One enantiomer is generated by operation of  $\alpha$ -ionone, as shown in Figure 7. As configuration is restricted only by distance, it seems to make it possible to determine a diastereomer, but it is impossible to make the distinction of an enantiomer at the present. However, the decision about a diastereomer is based on only NOE data, so that it is difficult to decide on diastereomer without NOE data which definitely restrict the stereoisomerism. For example, as shown in the case of pilocarpine in Figure 9, a correct structure or its enantiomer is not obtained when the information of an asymmetric center on C<sup>5</sup> is not given. This problem seems to be solved by introduction of a *stereoisomeric structure generator*,<sup>10</sup> which enables one to generate all possible stereoisomers using only the topological information of structure.

As mentioned above, good results are obtained for the objects of this work, *examination of candidate structure* and *generation of three-dimensional structure*, although the problem of improvement in the minimization and the treatment of an enantiomer still remains.

#### REFERENCES AND NOTES

- (1) Funatsu, K.; Miyabayashi, N.; Sasaki, S. Further Development of Structure Generation in Automated Structure Elucidation System

CHEMICS. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 18.

- (2) Funatsu, K.; Susuta, Y.; Sasaki, S. Introduction of Two-Dimensional NMR Spectral Information to an Automated Structure Elucidation System, CHEMICS. Utilization of 2D-INADEQUATE Information. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 11.
- (3) Crippen, G. M. *Distance Geometry and Conformational Calculations. Chemometrics Research Studies Series 1*; Wiley: New York, 1981.
- (4) Kline, A. D.; Braun, W.; Wüthrich, K. Studies by Proton Nuclear Magnetic Resonance and Distance Geometry of The Solution Conformation of The  $\alpha$ -Amylase Inhibitor Tendamistat. *J. Mol. Biol.* **1986**, *189*, 377. Brunger, A. T.; Clore, G. M.; Gronenborn, A. M.; Karplus, M. Three-Dimensional Structure of Proteins Determined by Molecular Dynamics with Interproton Distance Restraints: Application to Crambin. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 3801.
- (5) Yamasaki, T.; Abe, H.; Kudo, Y.; Sasaki, S. CHEMICS: A Computer Program System for Structure Elucidation of Organic Compounds. *ACS Symp. Ser.* **1977**, *54*, 108.
- (6) This software package was provided by TORAY System Center Corp. in Japan.
- (7) Fletcher, R.; Reeves, C. M. Function Minimization by Conjugate Gradients. *Comput. J.* **1964**, *7*, 149.
- (8) Powell, M. J. D. An Iterative Method for Finding Stationary Values of a Function of Several Variables. *Comput. J.* **1962**, *5*, 147.
- (9) Klopman, G.; McGonigal, M. Computer Simulation of Physical-Chemical Properties of Organic Molecules. 1. Molecular System Identification. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 48.
- (10) Abe, H.; Hayasaka, M.; Miyashita, Y.; Sasaki, S. Generation of Stereoisomeric Structure Using Topological Information Alone. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 216.