Recent Advances in the Automated Structure Elucidation System, CHEMICS. Utilization of Two-Dimensional NMR Spectral Information and Development of Peripheral Functions for Examination of Candidates

Kimito Funatsu* and Shin-ichi Sasaki

Department of Knowledge-Based Information Engineering, Toyohashi University of Technology, Tempaku, Toyohashi 44l, Japan

Received October 10, 1995[⊗]

A program for applying two-dimensional NMR spectroscopic data from $^{1}H-^{13}C$ COSY (one-bond C−H correlations)/$^{1}H-^{1}H$ COSY (three bond H−H correlations) and 2D-INADEQUATE (one-bond C-C correlations) experiments has been developed and introduced into CHEMICS. The main concepts employed in this study consists of making a more accurate assignment of ^{13}C and ^{1}H chemical shifts to carbons and hydrogens in the sample structure at the data analysis step in CHEMICS followed by the generation of more probable structures as candidates using $^{13}C-^{13}C$ and $^{1}H-^{1}H$ coupling information. The detailed algorithm are described. In addition, current configuration of CHEMICS including peripheral functions for examination of candidates by mass and ^{13}C-NMR spectral prediction and the brief overview are described.

1. INTRODUCTION

The automated structure elucidation system for organic compounds, CHEMICS, presents all possible structures that are consistent with an unknown's spectroscopic data and molecular formula. As spectroscopic data of the unknown IR, ^{1}H-NMR, and ^{13}C-NMR spectra have hitherto been employed in CHEMICS.[1,2] For structure elucidation of organic compounds with a complex structure, two-dimensional (2D) NMR measurement has been made particularly in these days. Therefore, in addition to those traditional spectroscopic data, the information provided by 2D-NMR technique was introduced into CHEMICS system. However, only the results of 2D-INADEQUATE experiment was utilized in the previous work.[3] Munk et al. have reported on the application of 2D-NMR spectral data to their structure elucidation systems, CASE[4] and its successor SESAMI.[5] In the system, data from $^{1}H-^{13}C$ COSY (one-bond C−H correlations)/$^{1}H-^{1}H$ COSY (three bond H−H correlations) experiments are handled as well as data from 2D-INADEQUATE (one-bond C-C correlations) experiment. In this application, two steps are required to generate the structure: a computer generates fragmental substructures compatible with 2D NMR signal connectivity information, and then it constructs candidate structures using those substructures. On the other hand, in our approach, a computer directly and automatically generates the candidates under the intercalated relationship between the 2D NMR analysis and the structure generation procedures, as will be described later, without generating fragmental substructures as in the SESAMI system.

In this paper, in order to expand CHEMICS we also have noted information obtainable from $^{1}H-^{13}C$ COSY, $^{1}H-^{1}H$ COSY experiments as well as 2D-INADEQUATE experiment.

A series of procedures in the present CHEMICS without 2D-INADEQUATE analysis is as follows (Figure 1). (1) Input spectroscopic data and a molecular formula of the unknown. (2) The 630 substructures (called components in the system) are already stored in a computer.[2] They are necessary and sufficient to construct any neutral chemical structure containing C, H, O, N, S, and halogen atoms. There are three class levels, i.e., primary, secondary, and tertiary components, depending on the size, attributes, and characteristics of the component. Components not consistent with the unknowns of the data are discarded. (3) If NMR data

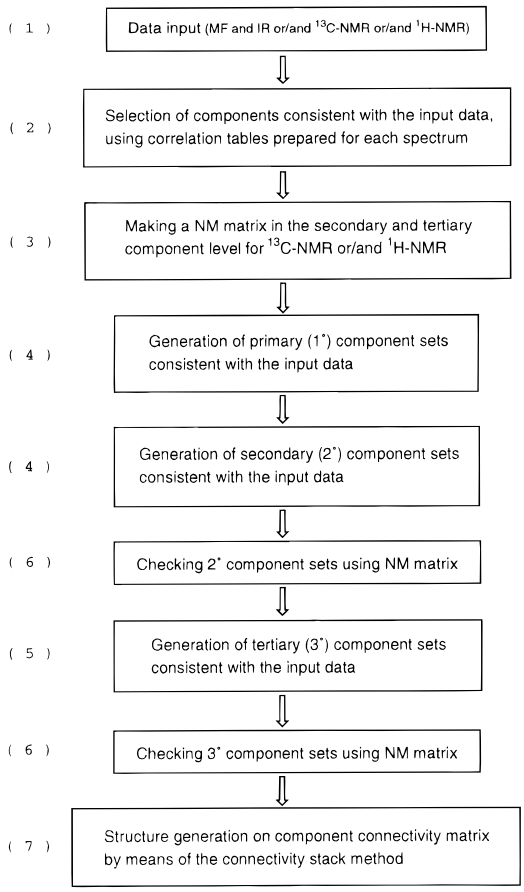


Figure 1. Block diagram of CHEMICS (the numbers in parentheses correspond to the ones that appeared in the Introduction).

[⊗] Abstract published in *Advance ACS Abstracts,* March 1, 1996.

0095-2338/96/1636-0190$12.00/0

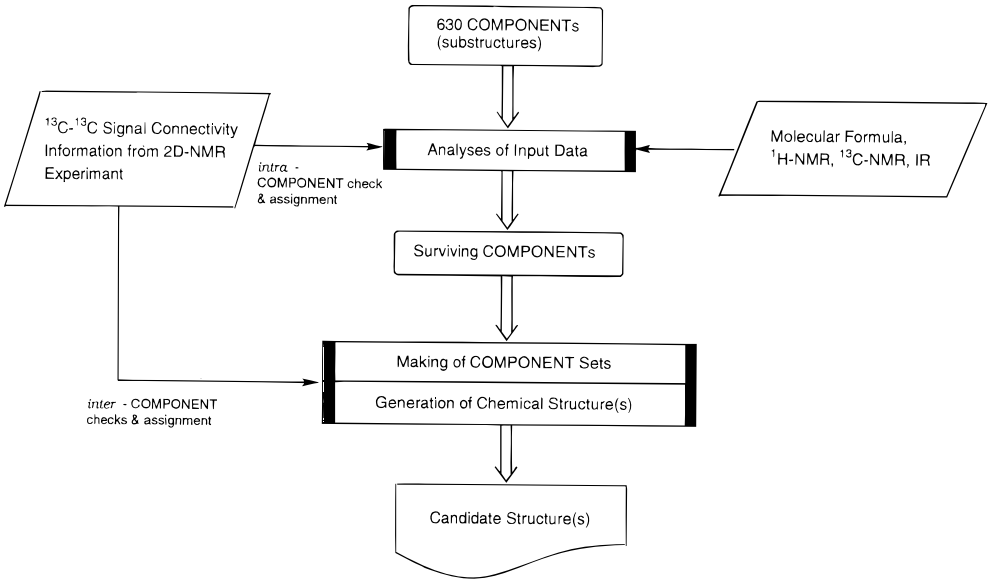


Figure 2. Flowchart of CHEMICS and relation with 2D-NMR data analysis.

are already input, all possible combinations of which component corresponds to which NMR signals are calculated for each secondary and tertiary components. The results are summarized as the NM matrix.[6] (4) Possible sets of secondary components that satisfy spectroscopic information are obtained. (5) From each set of secondary components obtained at step 4, sets of tertiary components are obtained. These partly satisfy the spectroscopic information, and they will be used for structure construction. (6) Each set of tertiary components is examined to determine whether it is consistent with all of the spectroscopic data on the bases of the NM submatrix constructed by the NM matrix elements, corresponding to each component in the set.[6] (7) From each set of tertiary components thus obtained, all possible candidate structures without any deficiency or duplication are constructed on the connectivity matrix by use of the connectivity stack technique.[2,7] Here, the connectivity matrix stands for the connective relationships of tertiary components to construct full structure(s).

We took notice that the NM matrix of CHEMICS contains intensive information about possible assignments of the signals to components and newly attached the function automatically exchanging connectivity information between signals into the connectivity between the corresponding components and generating candidate structures under the restrictive condition to satisfy that connectivity.

The NM matrix summarizes which component corresponds to which NMR signal using the so-called correlation table of the component on the range of the NMR chemical shift. The NM submatrix correlates components in a set of tertiary components with the corresponding carbon and hydrogen atoms in a sample structure to ^{13}C- and ^{1}H-NMR signals by the introduction of coupling information of the ^{13}C−^{13}C, ^{1}H−^{13}C, and ^{1}H−^{1}H nuclei to NM submatrix. The clearer the assignment correlation becomes, the fewer the number of candidates respond.

The concept of CHEMICS system consists in never missing correct solution. On the other hand, a large number of candidate structures are generated which are not inconsistent with input data. Therefore, examination functions for candidates are required.

Most of the NMR elucidation softwares can provide substructural information of the unknown but never generate all possible whole structures. However, after generation of candidate whole structure, the commercially available NMR elucidation and analysis softwares are powerful to optimize 3D-structure.

In the last part of this paper (Appendix), the brief overview of current configuration of CHEMICS and, in particular, the examination functions assisted by mass and ^{13}C-NMR spectral prediction are described with some results.

2D-NMR data which can be handled with analytical functions so far developed are ^{13}C−^{13}C signal connectivity information (There are two ways: direct information obtained either from the 2D-INADEQUATE method, long range C−H COSY or indirect information from combination of long range C−H COSY and H−H COSY.), ^{13}C−^{1}H signal connectivity information, and ^{1}H−^{1}H signal connectivity information. For convenience of explanation from now on, description will begin with the analysis of complete ^{13}C−^{13}C signal connectivity information.

2. ANALYSIS OF CONNECTIVITY INFORMATION OF ^{13}C−^{13}C SIGNAL

Analytical processing of the ^{13}C−^{13}C signal connectivity information in CHEMICS is roughly divided into two stages (Figure 2): Namely, the first one is the part which examines carbon−carbon connectivity *within the component* and performs assignment processing when components are chosen, and the second one is the part which examines carbon−carbon connectivity *between components* at the time of structure generation while performing assignment processing. What made this flow more precisely correspond to analytical processing in CHEMICS is shown in Figure 3. 2D-NMR data analysis is divided into four stages in dependence on the position where the processing is conducted. Stage 1 takes the part of analytical processing within the component mentioned in Figure 2, and the remaining three stages share analytical processing among components. In stage 2, we obtained the corresponding relation between the component and ^{13}C-NMR signal, for each component set, from input signal connectivity information and correla-

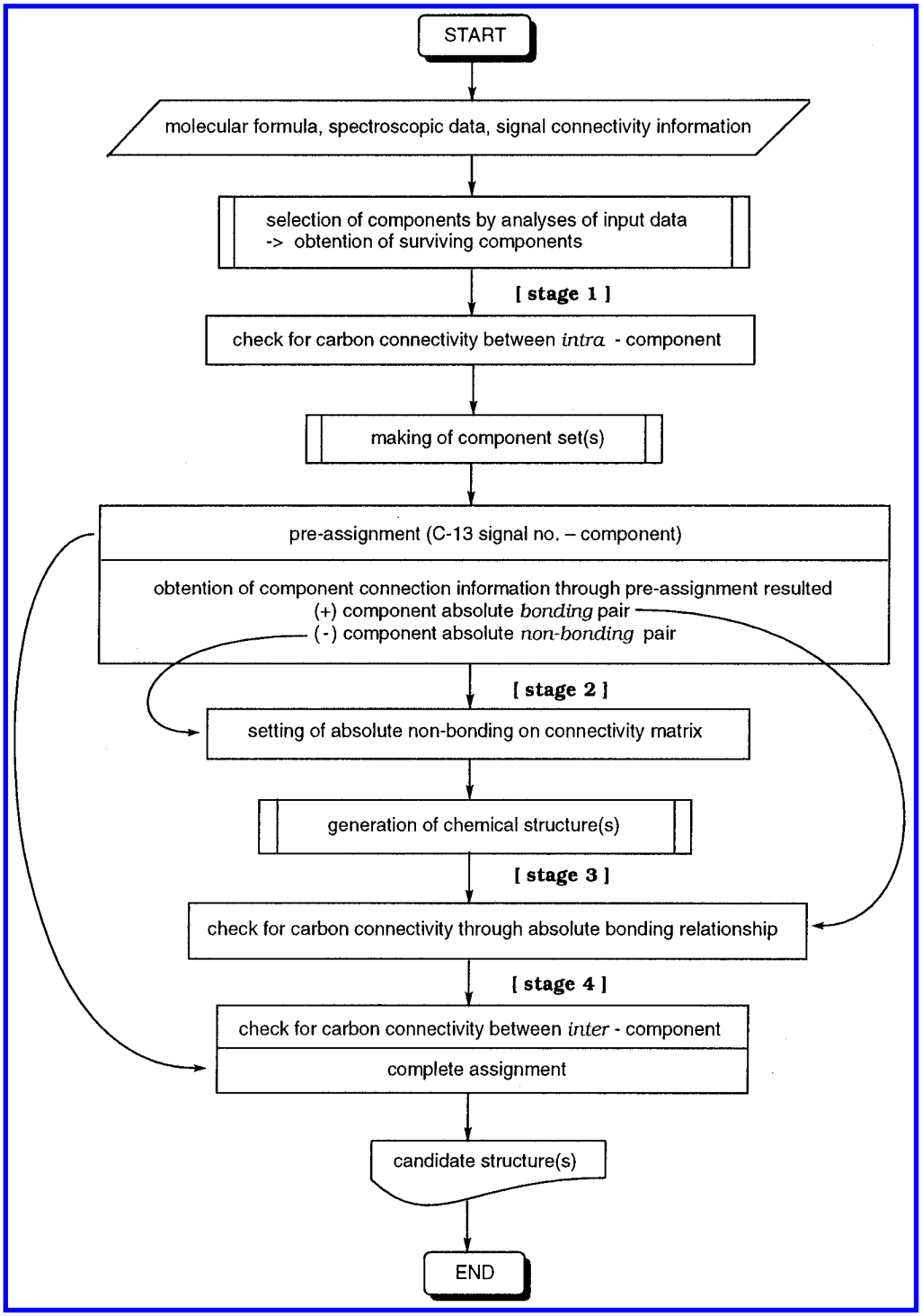


Figure 3. Flowchart of 2D-NMR data analysis.

tion table (this relation is called preassignment information), and acquire information regarding connectivity restriction between components prior to structure construction. In other words, these are the absolute connectivity relation that two components are certainly connected and the absolute non-connectivity relation that they can never be connected. The absolute nonconnectivity relation is employed in stage 2 before the structure generation by connectivity stack method, whereas the absolute connectivity relation is used in stage 3 after the structure generation, both as the condition to examine connectivity of the component. For structures which survived in stage 3, component connectivities not yet examined are investigated on the ground of the previous assignment information and $^{13}C-^{13}C$ signal connectivity information, and complete assignment processing is also attempted at the same time.

Here we go describing details of $^{13}C-^{13}C$ signal connectivity information processing while following the process of structure elucidation for an unknown sample which satisfies ^{1}H- and ^{13}C-NMR spectral data and $^{13}C-^{13}C$ signal connectivity information shown by paired signal numbers attached to ^{13}C-NMR signal which are shown in Table 1. It is sufficient here only to enter the 2D-NMR information just as the pair of ^{13}C-NMR signal numbers, and we like to leave putting emphasis that input of information like partial structure is never required.

First, description is given to the examination processing within components in stage 1. When a molecular formula

Table 1. Spectral Data of L-Menthol[a,b]

^{1}H-NMR data/ppm (rel intensity) 90 MHz	
0.80 (16)	1.55 (8)
0.90 (16)	1.95 (4)
0.93 (16)	2.31 (4)
1.10 (4)	3.28 (4)
1.30 (4)	4.46 (4)
^{13}C-NMR data/ppm (rel intensity) multiplicity	
1. 71.3 (13) d	6. 25.8 (14) d
2. 50.1 (17) d	7. 23.2 (13) t
3. 45.1 (12) t	8. 22.2 (12) q
4. 34.6 (11) t	9. 21.0 (12) q
5. 31.7 (21) d	10. 16.1 (15) q
^{13}C−^{13}C signal connectivity information	
1−2	4−5
1−3	4−7
2−6	5−8
2−7	6−9
3−5	6−10

[a] Nakanishi, K. Superconductivity FT-NMR. *Kodans*ha **1986**, 126−127, 152−155. [b] Molecular formula: $C_{10}H_{20}O$.

Table 2. Surviving Tertiary Components

*$(CH_3)_2CH-$ (O)	$CH_2<$ (CD)	−OH (CD)
*$\underline{(CH_3)_2CH-}$ (CD)	$CH_2<$ (CS)	−OH (CS)
*$(CH_3)_2CH-$ (CS)		
	−CH< (O)	−O− (CD)
*CH_3CH_2- (CD)	−CH< (CD)	−O− (CS)
*$\underline{CH_3CH_2-}$ (CS)	−CH< (CS)	
*$\underline{CH_3CH<}$ (O)	−CH═ (O)	
*$\underline{CH_3CH<}$ (CD)	−CH═ (CD)	
*$CH_3CH<$ (CS)	−CH═ (CS)	

is put in together with ^{1}H- and ^{13}C-NMR data, 20 components shown in Table 2 survived, as they do not conflict with these data, out of 630 tertiary components along the procedure of data analysis which have so far been mentioned in the previous papers.[1,2,6] Among these, components marked with an asterisk in Table 2, which contained a carbon−carbon bond in the component, were examined with ^{13}C−^{13}C signal connectivity information entered, and four underlined were rejected. For instance, as shown in Figure 4a, six kinds of signal assignment to no. 1 $CH_3CH<$(O) are considered as possible patterns based on the correlation table. This is caused by the fact that these have satisfied none of the signal pairs shown in Table 1. On the contrary, there are six kinds of possible assignment in the case of $CH_3CH<$(CS) as shown in Figure 4b, and only three underlined have satisfied the input signal pairs and cleared this examination. Also at the same time, signal no. 5 or no. 6 has been assigned to the carbon in >CH− of this component. Following this, component sets which do not conflict with the molecular formula and 1D NMR spectral data are generated on the basis of 16 tertiary components survived. Figure 5 shows one of the generated tertiary component sets and preassignment information. In other words, this shows that the carbon of >CH− in $(CH_3)_2CH$(CS) is uniquely assigned to ^{13}C-NMR signal no. 6, whereas $-CH_2-$(CS) to either of no. 3, no. 4, or no. 7. This information of preassignment is obtained through the process shown in Figure 5. In short, early assignment relation (Figure 5, top), component B cleared in stage 1, for instance, is possibly assigned to no. 5 or no. 6, but a unique correspondence between component A and signal no. 6 eliminates correspondence between other

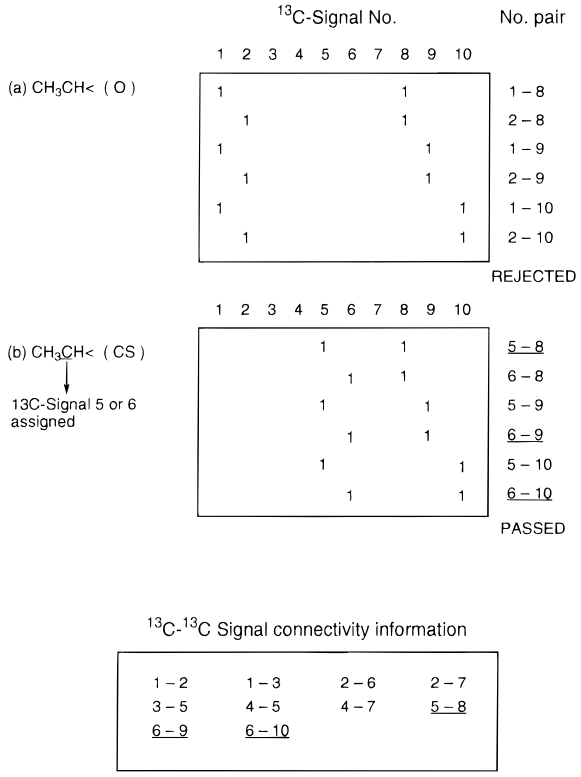


Figure 4. Intracomponent examination and assignment processing at stage 1.

components and signal no. 6. From the unique assignment relation (component B with signal no. 5) newly generated by the processing, assignment for E becomes no. 1 and no. 2, giving the final information of preassignment.

Regarding the tertiary component set shown in Figure 5, the connectivity relation between components A and B, for instance, is set by using part 1 as shown in Figure 6. And connectivity relations of A with C_1, C_2, and C_3 give all the same combinations of the corresponding signal numbers as 6−3, 6−4, and 6−7. Therefore, region of three connectivity elements under the same condition was collected as part 2. For a total of 11 parts in addition, it is determined whether setting of the absolute connectivity is possible or that of the absolute nonconnectivity is possible or uncertain as shown in Table 3. Preassignment information obtained before and ^{13}C−^{13}C signal connectivity information are utilized for this operation. Regarding part 2, for instance, though three possibilities are obtained as the signal pair from the preassignment information, this part is regarded to be in the relation of absolute nonconnectivity (shown with −) because neither of these pairs exists in the connectivity information entered. Similar relations were obtained for parts 1, 6, and 9, too. Part 11 has a possible signal pair, on the contrary to this, but this pair is set as the absolute connectivity (shown with +) because it has satisfied the input data. By the way, connectivity between parts in uncertain connectivity relation is examined in the last stage 4.

Following determination of absolute connectivity and nonconnectivity relations from ^{13}C−^{13}C signal connectivity information with the procedure mentioned above, examination is conducted in checks 2 and 3 which correspond to before and after the structure generation by connectivity stack method in the next stage as shown in Figure 3. In the connectivity matrix on the left side in Figure 7, −2 was

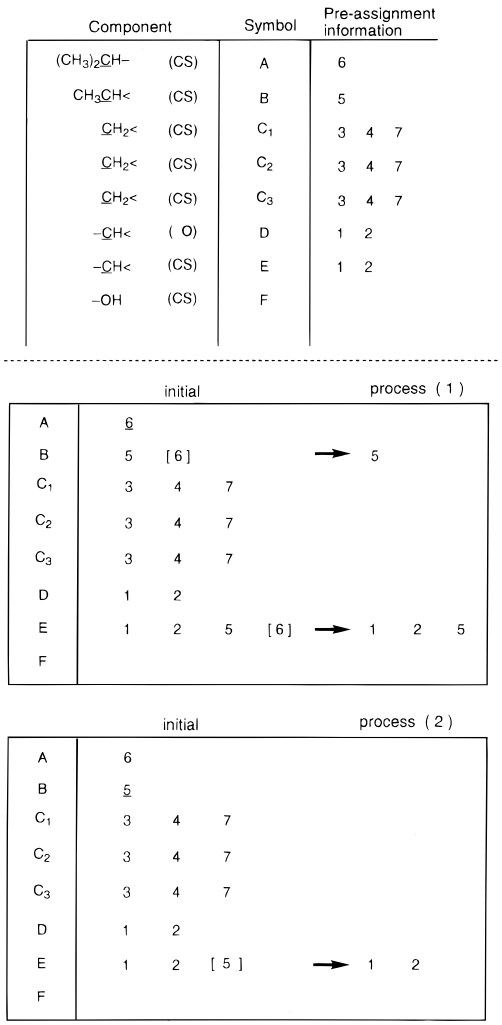


Figure 5. Flow of processing to obtain preassignment information.

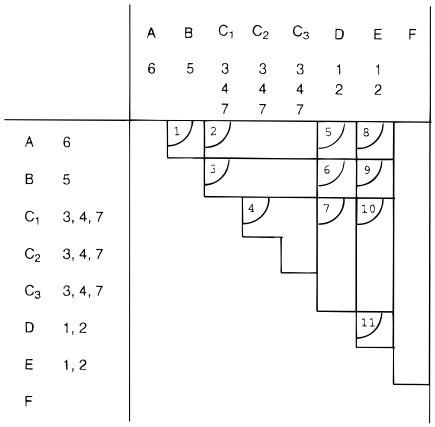


Figure 6. Connectivity matrix divided into 11 portions according to combination of component pairs.

settled for parts 1, 2, 6, and 9 to express the absolute nonconnectivity. This is done as stage 2. Similarly on the right of Figure 7 (which corresponds to stage 3), structure generation by the connectivity method has finished, and here examination if absolute connectivity exists is conducted for part 11 previously obtained. In this example, five candidate structures were generated according to setting in stage 2, and three of them were further accepted according to stage 3.

In the last stage 4, the final examination of the carbon connectivities between components is done while carrying

Table 3. Component Connectivity Information

part no.	pair of signal no.	component connection[a]
1	6−5	−
2	6−3	
	6−4	−
	6−7	
3		uncertain
4		uncertain
5		uncertain
6	5−1	
	5−2	−
7		uncertain
8		uncertain
9	5−1	−
	5−2	
10		uncertain
11	1−2	+

$^{13}C-^{13}C$ signal connectivity information

1−2	1−3	2−6	2−7
3−5	4−5	4−7	5−8
6−9	6−10		

[a] +, absolute connectivity; −, absolute nonconnectivity.

the assignment processing for the survived three candidate structures. Shown in Figure 8 are those three candidate structures, the corresponding connectivity matrices, and ^{13}C signal number assigned to each component in ( ). In this case, signals no. 4, no. 3, and no. 7 have been assigned to component connectivities of C_1, C_2, and C_3, respectively. Because connectivity between C_1 and C_3 had been corresponding to the connectivity of signals 4 and 7, these were uniquely determined.

On this ground, examination is successively carried on if signal pairs which correspond to all component connectivities in the structure have satisfied the signal connectivity information entered. For instance, concerning the structure shown at the top in Figure 8, the pair of signals no. 7 and no. 1 corresponding to the connectivity of components C_3 and E is compared with the signal connectivity information, but this structure gives no connectivity information with agreement in this case, causing removal from candidates. Similarly, the structure at the bottom in Figure 8 brought a conflict in the connectivity relation between components C_2 and E or C_3 and D and was eliminated from candidates. The second structure in Figure 8 did not give such conflict and was accepted.

Collected in Figure 9 was the result of executing CHEMICS with input of the $^{13}C-^{13}C$ signal connectivity information. The left column shows input data and the right output result. Assigned results of the ^{13}C NMR signal are simultaneously output for any example.

An example shown in Figure 9a is the executed result of CHEMICS for isomers of thujopsene. Input data are the molecular formula ($C_{15}H_{24}$), ^{13}C-NMR data, and $^{13}C-^{13}C$ signal connectivity information. In this case, convergence was possible into seven including the correct answer. By the way, 4450 candidates are obtained without $^{13}C-^{13}C$ signal connectivity information. An example shown in Figure 9b is the result for input of the complete $^{13}C-^{13}C$ signal connectivity information in addition to the molecular formula ($C_{15}H_{22}O_3$) and ^{13}C-NMR data. The number of output candidate structures is 3, and these three candidates are to be objectively output even from the viewpoint of ^{13}C-NMR signal connectivity information put out at the same

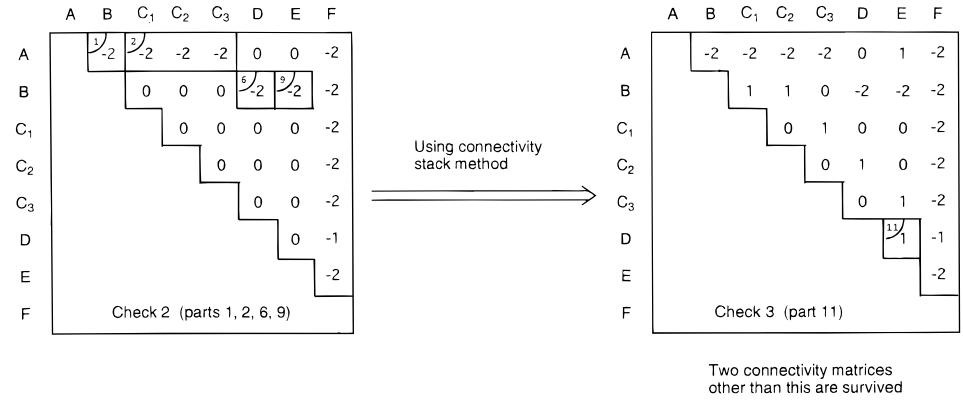


Figure 7. Connectivity matrices before (left) and after (right) use of the connectivity stack method. (−2, absolute nonconnectivity; −1, absolute connectivity).

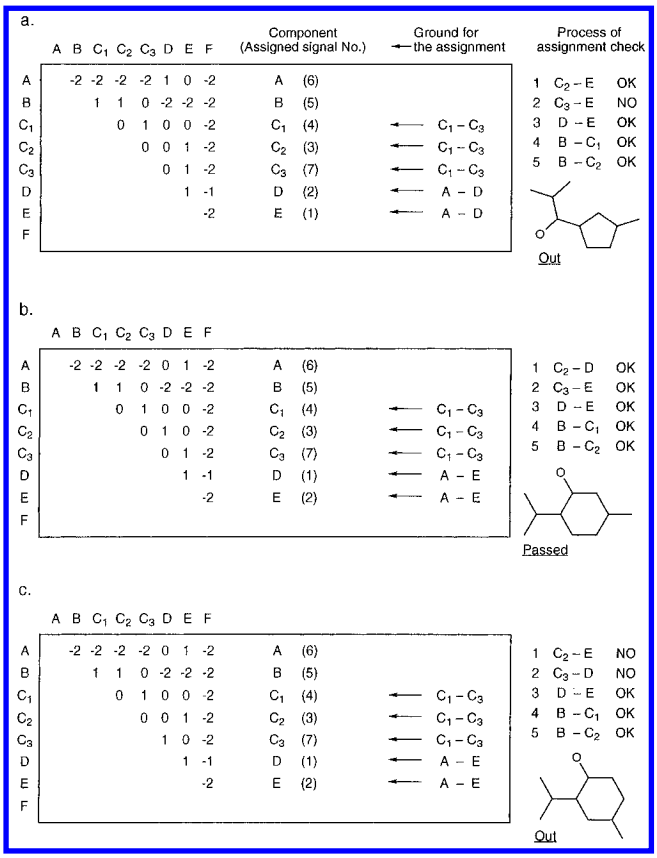


Figure 8. Examination for connectivity matrix to obtain complete assignment information of L-menthol.

time. If referring to its IR spectrum, however, it can be considered as a five-membered ring lactone at a high possibility, allowing us then to reach the correct structure which is underlined. Shown in Figure 9c is also an example that made use of a partial ^{13}C-NMR signal connectivity information. Ambiguity in assignment rises due to the partial connectivity information, and, correspondingly, the number of candidate structures is relatively large also. But these are reasonable. We have shown some other examples. Besides, connectivity relations of saturated carbons are automatically obtained from the ^{13}C−^{13}C signal connectivity information to speed execution, to which macrocomponent processing mentioned in literature[2] and elsewhere is performed. In Figure 9d, macrocomponent information acquired on the way has also been shown in particular.

Procedure of analytical processing for the ^{13}C−^{13}C signal connectivity information mentioned here is applied to analysis of ^{1}H−^{1}H signal connectivity information and the like which are easily available in practice.

Disposal for Partial ^{13}C−^{13}C Signal Connectivity Information. When nothing but partial ^{1}H−^{1}H signal connectivity information is at hand as the input information, examination processing in stages 2 and 3 cannot be conducted, considering existence of signal pairs unobserved yet. Therefore, it has been devised so that examination is conducted for connectivities between carbons in correspondence, only when ^{13}C−^{13}C signal connectivity information has been obtained for all methyl groups (signal with multiplicity of 4), by using the signal connectivity information regarding these.

3. ANALYSIS OF ^{13}C−^{1}H SIGNAL CONNECTIVITY INFORMATION

Analyses of ^{1}H- and ^{13}C-NMR data are independently performed from each other by means of the correlation table, and no consideration has been taken for mutual assignment relationship between respective NM matrices expressing the analytical result (namely, enumerated possibility of assignment). In short, they are information independent from each other. If a connectivity relation between ^{1}H- and ^{13}C-NMR signals is given, however, restraint naturally takes place in the assignment relation enumerated on both NM matrices, causing clarification of assignment. Clarification of assignment in this stage brings elimination of components which lost assignment possibility therewith and also effectively influences conformity examination for secondary and tertiary component sets. Shown in Figure 10 was an example for the process to clarify assignment on the NM matrix of a certain component by using ^{13}C−^{1}H signal connectivity information.

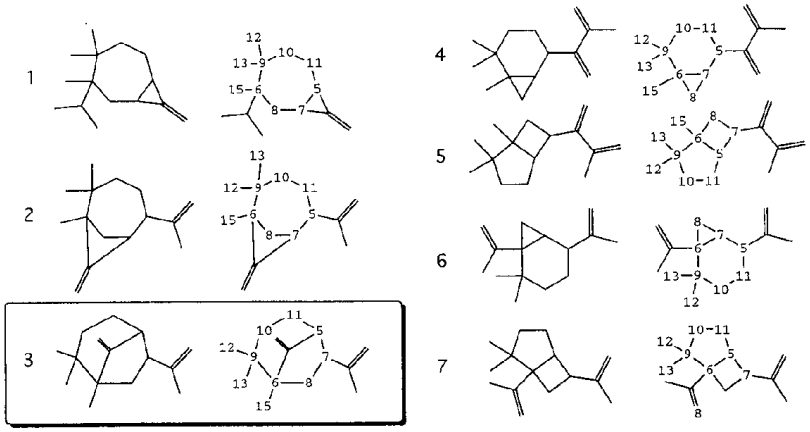
Here we describe the process of clarifying assignment using data of C−H COSY shown in Table 4 in addition to ^{1}H- and ^{13}C-NMR data already shown in Table 1. Prior to this, we put up some of the data of C−H COSY shown in Table 4. Here, signal pairs of C and H obtained from the C−H COSY chart are represented as the pair of the corresponding signal numbers which have already been attached to each signal.

Now, through analyses of the molecular formula ($C_{10}H_{20}O$) and ^{1}H- and ^{13}C-NMR data, 20 tertiary components survive as shown in Table 2. For instance, as the evidence for survival of $-CH_2-$(CS) in these, 32 kinds of combinations are enumerated as shown on the left in Figure 10 for possibility of assignment to ^{1}H-NMR signal in the process

a) $C_{15}H_{24}$ Thujopsen isomer

	^{13}C-NMR			^{13}C-^{13}C SIGNAL
	PPM	Rel. int.	Mult.	Connectivity
1.	160.1	8.0	1	5 7
2.	149.9	8.0	1	5 11
3.	108.7	22.5	3	6 8
4.	101.7	20.0	3	6 9
5.	49.9	22.0	2	6 15
6.	49.7	9.0	1	7 8
7.	47.0	18.0	2	9 10
8.	40.4	22.0	3	9 12
9.	38.8	12.0	1	9 13
10.	35.21	18.5	3	10 11
11.	32.5	23.0	3	
12.	25.7	23.0	4	
13.	23.5	18.0	4	
14.	21.3	16.0	4	
15.	18.1	18.0	4	

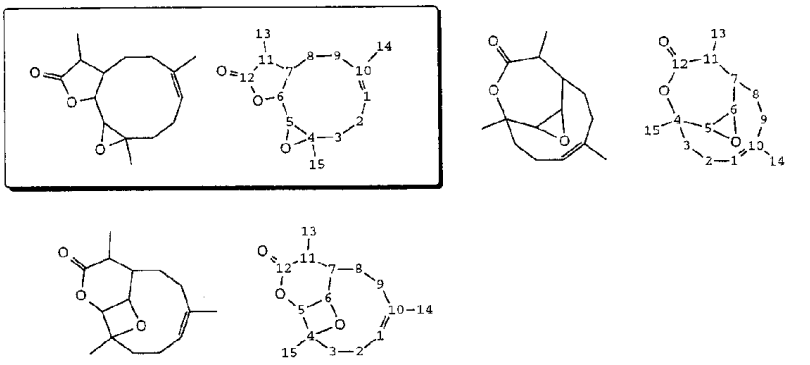
(R. Benn, J. Klein, Magn. Res. Chem., 24, 638 (1986))



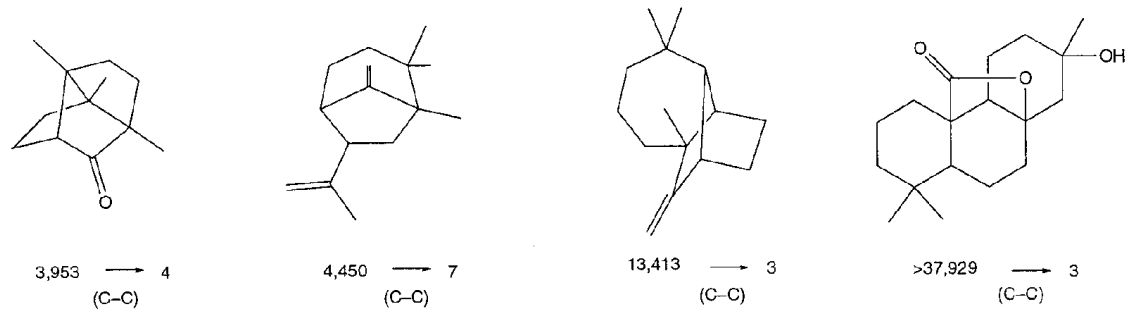
b) $C_{15}H_{22}O_3$ 11,13-Dihydroparthenolide

	^{13}C-NMR		^{13}C-^{13}C SIGNAL
	PPM	Mult.	Connectivity
1.	125.00	2	1 2
2.	24.00	3	1 10
3.	36.60	3	2 3
4.	61.30	1	3 4
5.	66.60	2	4 5
6.	82.00	2	4 15
7.	51.80	2	5 6
8.	29.70	3	6 7
9.	41.00	3	7 8
10.	134.40	1	7 11
11.	42.30	2	8 9
12.	177.20	1	9 10
13.	13.10	4	10 14
14.	16.70	4	11 12
15.	17.00	4	11 13

(F.J. Parodi et al., SPECTROSCOPY LETTERS, 20 (1987) 445)



c)



3,953 → 4 (C–C)

4,450 → 7 (C–C)

13,413 → 3 (C–C)

>37,929 → 3 (C–C)

1D ^{1}H– and ^{13}C–NMR + *partial* ^{13}C–^{13}C signal connectivity

d) $C_{15}H_{22}O_3$

	^{13}C-NMR		^{13}C-^{13}C SIGNAL
	PPM	Mult.	Connectivity
1.	185.7	1	1 5
2.	177.2	1	1 6
3.	155.0	2	3 6
4.	151.2	1	4 5
5.	127.6	1	4 7
6.	125.0	2	2 10
7.	80.9	2	3 9
8.	53.1	2	4 9
9.	41.0	1	5 15
10.	40.4	2	7 8
11.	37.2	3	8 10
12.	22.3	3	9 11
13.	24.6	4	8 12
14.	12.0	4	9 13
15.	10.2	4	11 12
			10 14

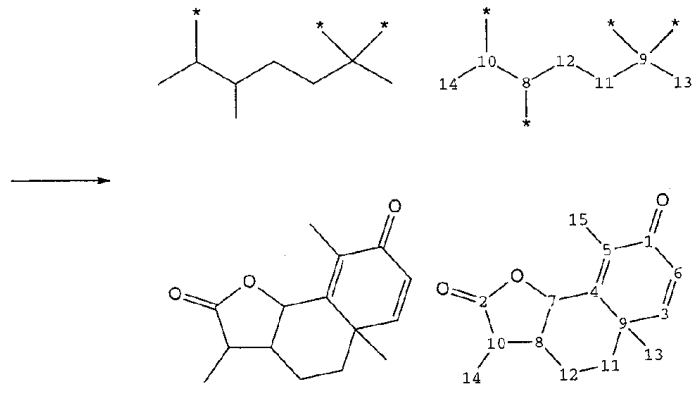


Figure 9. Execution example of CHEMICS using ^{13}C−^{13}C signal connectivity information.

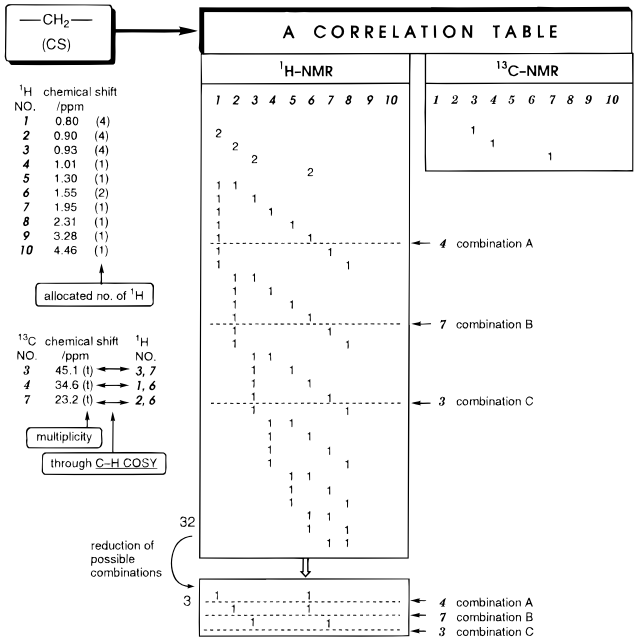


Figure 10. Utility example of CHEMICS using $^{13}C-^{1}H$ signal connectivity information at component level.

Table 4. C−H COSY Data of L-Menthol

pair of signal no. $^{13}C-^{1}H$	pair of signal no. $^{13}C-^{1}H$
4−1	5−5
10−1	4−6
7−2	7−6
8−2	3−7
3−3	6−8
9−3	1−9
2−4	

of analysis, and stored in the NM matrix. The column corresponds to each signal number of ^{1}H-NMR, and elements of the matrix are the hydrogen number assigned to that signal. On the other hand, possibility of assignment to ^{13}C-NMR signal is obtained as three kinds as shown on the right in Figure 10. Basis for survival of $-CH_2-(CS)$ is due to the possibility of assignment above, but these assignment relations to ^{1}H- and ^{13}C-NMR signals are mutually independent. Here let us consider the role of C−H COSY. Assignment of $-CH_2-(CS)$ to the ^{13}C-NMR signal is no. 3, no. 4, and no. 7. According to C−H COSY, these ^{13}C-NMR signals at no.3, no. 4, and no. 7 are to be connective with ^{1}H-NMR signals (3, 7), (1, 6), and (2, 6), respectively. Accordingly, assignment of hydrogen in $-CH_2-(CS)$ to the ^{1}H-NMR signal is effective only for these pairs of (3, 7), (1, 6), and (2, 6), and all other possibilities of assignment are abandoned as they conflict with the C−H COSY data. Here, the possible combination is eventually compressed to the NM matrix shown at the bottom in Figure 10. At this time, assigned elements in each row between the both NM matrices have been corresponded 1:1. Here, if effective assignment of hydrogen does not exist, this component itself is discarded as it conflicts with the input data. This assignment information is employed also for consistency check of the component set. With this sort of operation, mutual correlation is conducted for results of ^{1}H- and ^{13}C-NMR analyses, causing reduction of assignment possibility and screening of components. In this sample, the number of survived components could be reduced from 20 to 10 by making use of C−H COSY. Because of this reduction of components and reduction of frame for assignment possibility, reduction of candidate structural formula (219 → 85) and operating time.

Shown in Figure 11 is an example of execution with regard to utilization of C−H COSY. Partial $^{13}C-^{13}C$ signal connectivity information has also been utilized in this example.

4. ANALYSIS OF $^{1}H-^{1}H$ SIGNAL CONNECTIVITY INFORMATION

Coupling in C−H COSY is known to take place not only through chemical connectivity between carbon atoms connected with proton but also between spatially vicinal protons themselves. Thus, direct use of all H−H COSY data are hardly possible in CHEMICS at present because it is not so easy to generate connectivity between just corresponding carbon atoms from that information. However, only when an analyst can confirm coupling of protons themselves which are in vicinal relation in H−H COSY, it is possible to accept the signal pairs as the input data and carry automated analytical processing.

When separation between signals is poor in 1D ^{1}H-NMR, it is not supposed to possibly get too useful information even with H−H COSY. However, H−H COSY gives a high NMR sensitivity for proton, so it is often measured owing to merits including measurement time enough to be reduced and sample concentration enough to be low. Further, when separation between signals is good, the coupling relationship can easily be found from 1D ^{1}H-NMR, for instance, even without relying on H−H COSY measurement. In short, analysts often apply analytical processing to information of H−H COSY as easily available data. Therefore, investigation was made regarding to what extent the analytical process can be applied in CHEMICS, the automated structure elucidation system. Besides, these data are to be called $^{1}H-^{1}H$ signal connectivity information.

A. Acquisition of Input Data. Utilized as the input data is the signal pair corresponding to the cross peak of vicinal protons confirmed in H−H COSY. In short, a number has been given to each signal of ^{1}H-NMR beforehand, and the pair of numbers for the signal pair in question is utilized as the input data. However, each signal must be what has been clarified as the signal group caused by each proton so that it can be obtained by measuring C−H COSY or the like.

B. Relationship with Whole System. $^{1}H-^{1}H$ signal connectivity information is utilized in the level of candidate structure examination. Whenever one structure is composed, for components containing proton connected with carbon in the candidate structure, connectivity between carbons within a component or between components is respectively examined.

C. Algorithm of Analysis. As for the algorithm, it is identical to the analytical method for the partial $^{13}C-^{13}C$ signal connectivity information. When the $^{1}H-^{1}H$ signal connectivity information is utilized alone, it has been programmed so as to investigate, in the level of candidate structure examination, connectivity between carbons within a component and between components in the sense of indirectness.

(1) Without $^{13}C-^{1}H$ signal connectivity information entered: This case means stand-alone processing of $^{1}H-^{1}H$ signal connectivity information. In a certain candidate

structure, first for connectivities between all carbon atoms with vicinal proton, combination is determined between corresponding signals based on the assignment information for NM submatrix of ^{1}H-NMR at levels within a component and between components. When that satisfying ^{1}H−^{1}H signal connectivity information (input data) is not found in signal pairs obtained here, those structures are removed.

(2) With ^{13}C−^{1}H signal connectivity information entered: The ^{1}H−^{1}H signal connectivity information can be converted into the ^{13}C−^{13}C signal connectivity information by using with the ^{13}C−^{1}H signal connectivity information jointly. It has been assumed that the relationship of complete correlation between signals of ^{1}H- and ^{13}C-NMR spectra has been obtained, causing ^{1}H−^{1}H signal connectivity information to be automatically transformed into the ^{13}C−^{13}C signal connectivity information with ease, and utilized as the partial ^{13}C−^{13}C signal connectivity information. It is needless to say that the above-mentioned stand alone ^{1}H−^{1}H signal connectivity information is processed in parallel.

5. MACROCOMPONENT PROCESSING ACCOMPANIED BY NMR ASSIGNMENT INFORMATION

In some cases of structure elucidation for unknown samples, we can know particular partial structures and signal assignment information of NMR spectrum corresponding thereto. It has become possible in CHEMICS to reflect this information as macrocomponent + NMR assignment information.

The particular partial structure and NMR signal assignment information are either (1) automatically acquired by the system or (2) manually given to the system by the user. The way of consideration in these two cases is described below.

A. Case of Automated Acquisition. Connectivity of saturated carbons can sometimes be obtained automatically as the partial structure from ^{13}C−^{13}C signal connectivity information and multiplicity of ^{13}C-NMR. It is limited however to the case when the carbon number in molecular formula equals the ^{13}C-NMR signal number. Respectively recognized are the following:

(1) as a methyl group when a ^{13}C-NMR signal with a multiplicity of 4 forms coupling with another ^{13}C-NMR signal,

(2) as a methylene group when a ^{13}C-NMR signal with a multiplicity of 3 forms coupling with two different ^{13}C-NMR signals,

(3) as a methyne group when a ^{13}C-NMR signal with a multiplicity of 2 forms coupling with three different ^{13}C-NMR signals, and

(4) as a quadric carbon when a ^{13}C-NMR signal with a multiplicity of 1 forms coupling with four different ^{13}C-NMR signals.

Accordingly, ^{13}C−^{13}C signal connectivity information can be transformed into partial connectivity of saturated carbon. It turns out in this stage that the NMR signal has already been assigned to that partial structure. It is an important fact through this that clarification of assigning NMR signal to each component has been in progress. When ^{13}C−^{1}H signal connectivity information exists, in addition, it is meant that assignment of both ^{13}C- and ^{1}H-NMR data have been clarified.

B. Case of Manual Input by Users. Sometimes partial structures and the assignment relation of NMR signal there to may manually be given to the system by the user. Input is performed conversationally. In this occasion, assignment of the NMR signal for each component which composes the input substructure (macrocomponent) is conducted, with regard to component−NMR signal assignment relation collected for an NM matrix which has been obtained by the system on the ground of the NMR correlation table, by selecting one out of those possible combinations. If hardly selectable, input is not absolutely needed. Assignment of the NMR signal can be made for both ^{13}C and ^{1}H, and information thus entered is automatically transformed into ^{13}C−^{13}C or ^{1}H−^{1}H signal connectivity information. Further, when ^{13}C−^{1}H signal connectivity information has been put in, corresponding relationship of both ^{13}C- and ^{1}H-NMR data are displayed at the same time, causing aid of assignment task of the NMR signal to the partial structure.

Also, when coupling between signals has been clearly known though partial structure itself is not clear, it can be coped with entering a substructure being provisionally put in once and then eliminating the entered substructure after the assignment task is performed for that substructure. Through this operation, it is finally sufficient to have entered the connectivity relation only. As mentioned here, even when 2D-NMR information has not been obtained, pseudo ^{13}C−^{13}C or ^{1}H−^{1}H signal connectivity information is derived by the user, and it is to be employed in CHEMICS by means of procedures mentioned in sections 2 and 4.

6. EXAMPLE OF EXECUTING CHEMICS WHEN B-IONON IS TAKEN AS AN UNKNOWN

The input data were collected in Figure 12. The ^{1}H−^{1}H signal connectivity information is automatically transformed into a new partial ^{13}C−^{13}C signal connectivity information by using the ^{13}C−^{1}H signal connectivity information and added to the ^{13}C−^{13}C signal connectivity information already entered. To the ^{1}H−^{1}H signal connectivity information itself, stand-alone processing of examination is conducted before the last examination using the ^{13}C−^{13}C signal connectivity information in stage 4. In short, whenever a structure is composed, examination using the ^{1}H−^{1}H signal connectivity information is conducted through connectivity between carbon atoms in the structure on levels within component and between components, respectively (section 4.C). In this example of execution, the substructure (macrocomponent) attached with the ^{13}C-NMR assignment information has been entered as shown at the bottom right in Figure 12. Input of this information is conversationally performed, and a part of that operation was shown in Figure 13. The assignment is made in the unit of component while looking at the NMR data. Candidates of assignable signal numbers are presented in accordance with the correlation table, so the user is enough to choose among them. Whenever assigned, a corresponding part goes successively revised (Figure 13, underlined parts). From this signal assignment information, a pseudo ^{13}C−^{13}C signal connectivity information 1−2 is obtained and utilized. Although only the ^{13}C-NMR signal was assigned here, the ^{1}H-NMR signal can also be assigned with a similar operation. From the input data in Figure 12, structural formulas of six candidates were put out accompanied by NMR signal assignment information (Figure 14).

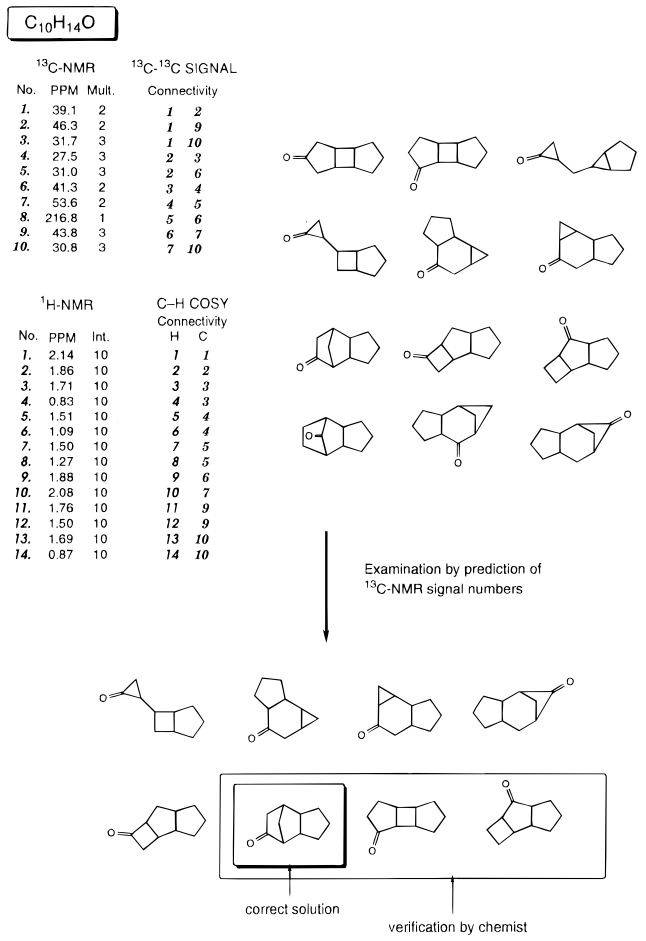


Figure 11. Execution example of CHEMICS utilizing C−H COSY data. Examination by prediction of ^{13}C-NMR signal numbers was also conducted. Enclosed with broken line frame are candidate structures eventually survived with detailed analysis of ^{1}H-NMR by the user.

Two-dimensional NMR data to enter CHEMICS are only pairs of signal numbers as we have shown here, processing thereafter is completely carried by CHEMICS. What we must particularly touch here is the fact that, if one acquires signal number pairs by looking at 2D-NMR chart, there is a fear of making a mistake due to points of complication or ambiguity. For this point, it is required to have a scheme of connecting CHEMICS directly to the NMR instrument and obtaining signal number pairs of input data for CHEMICS while processing raw data. Study concerning this is now in progress.

7. EXPERIMENTAL SECTION

CHEMICS and the related peripheral functions (see Appendix) are written in FORTRAN 77. Earlier program development was on VAXstation 3100. Currently program development is continuing on several kinds of Unix workstations.

In principle, there is no size limitation of compound analysis with CHEMICS and the related modules. But, at present, program memory size limits the maximum size of a compound to 50 carbons. Of course, the size is changeable, dependent on hardware performance.

The CHEMICS system including its peripheral functions are now available from commercial vender (Nagase & Co., Ltd.) in Japan.

APPENDIX: ON THE PRESENT STATUS OF CHEMICS

Improvements and developments of new functions of CHEMICS have been made even while writing this paper. We have described only major parts of trunk and periphery of CHEMICS of this main issue because it is overwhelming to include all aspects of these in this paper, though it is hoped they will be described by those in engagement of the development occurring at the present time. As for others than those written above, only an outline is mentioned here supplementarily, and its detail is to be described in another chance.

Shown in Figure 15 was a schematic diagram of comparatively recent CHEMICS. Shown in the brokenlined frame was the core function, so to speak. Some new functions have been attached around this in recent years to support CHEMICS structure analysis from various standpoints. In short, these are the following:

(1) 2D-NMR spectral data analysis (^{13}C−^{13}C, ^{13}C−^{1}H, and ^{1}H−^{1}H signal connectivity information, and HMBC data),

(2) acquisition of macrocomponent with IR spectral data analysis using symbolic logic (IRASSL).[8]

(3) acquisition of macrocomponent with NMR spectral data analysis to which correlation between ^{1}H- and ^{13}C-NMR chemical shifts (PASTEL) was applied,[9]

(4) examination with prediction of ^{13}C-NMR chemical shift and its assignment,[10]

(5) examination with prediction of ^{13}C-NMR signal number,[11]

(6) examination with mass spectral prediction,

(7) elucidation of three-dimensional candidate structure with NOE data,[12] and

(8) counterpart elucidation.[13]

Roles of (4) and (6) in the above-mentioned as the examination program are to be more or less touched upon.

Execution was made here for nicotine regarded as the unknown sample. Shown in Table 5 were input data. As a macrocomponent (or "host part"),[13] the 3-substitute pyridyl group was put in. As the result, seven candidate structures shown in Figure 16 were presented by CHEMICS. Incidentally, prediction for ^{13}C-NMR signal number is made for these candidate structures based on evaluation for the environmental equivalence of each carbon in the structure, and candidates, nos. 2, 5, and 7 which disagree with measured number of 10 (all were predicted as 9) can be removed.

Now, prediction of mass spectrum (*m/z* value only) is made for seven candidates generated a while ago, and a sort of ranking is given for candidate structures through comparison with the measured data. Prediction of *m/z* values for each candidate structure is carried by application of "fragmentation rules". "Fragmentation rules" has been described as an external file or program. Two examples of this description were shown in Figures 17 and 18, respectively. In the form of an external file, "fragmentation rules" is represented with digital strings in accordance with a certain grammar. Shown in Figure 17 was part of the rules relating to carbonyl. Contents of this file can easily be added and modified by the user, whereas for the rules which cannot be handled with this description, the form of a FORTRAN program was adopted. Figure 18 shows a part of the searching routine of the McLafferty rearrangement. Written in this has been the procedure in which, if structural conditions which allow

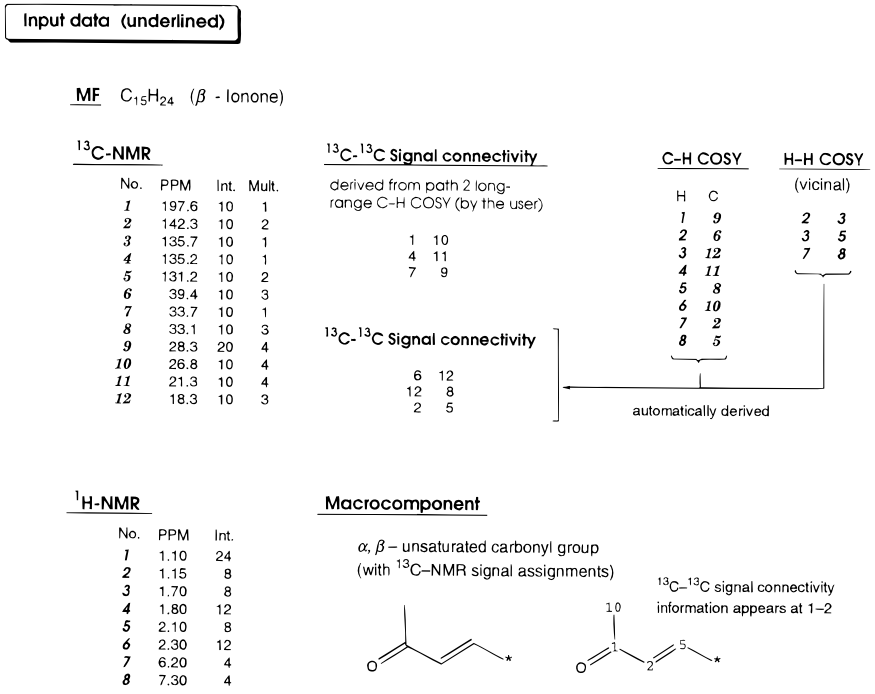
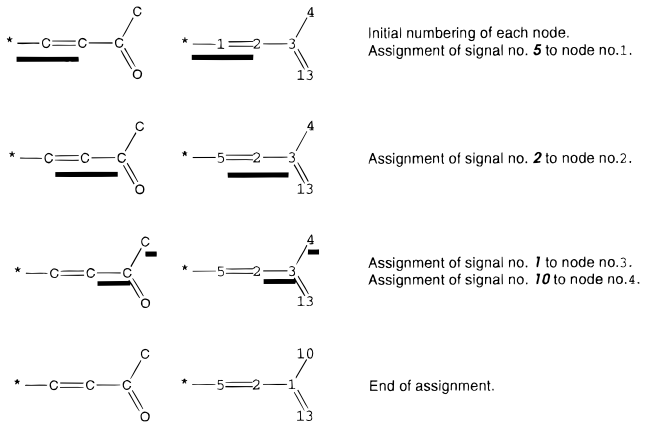
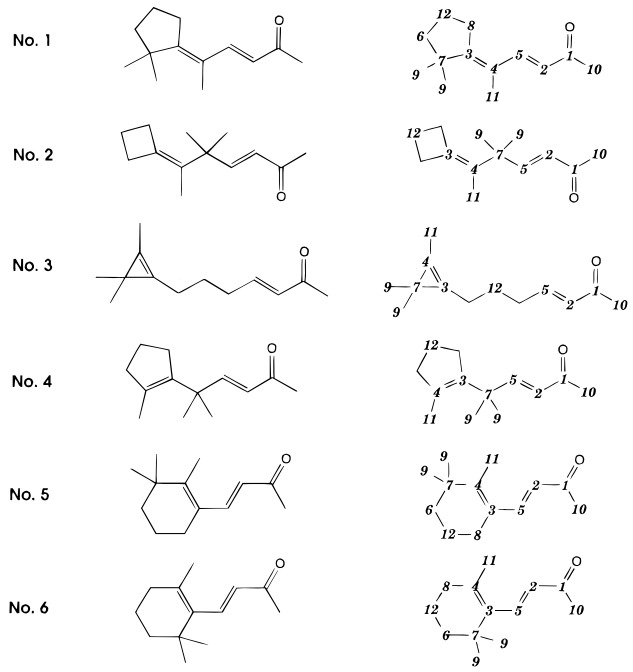


Figure 12. Input data for β-ionone regarded as the unknown.



13C-NMR chemical shift	No.	1H-NMR chemical shift -- 90 MHz
197.60 (C)	1	
142.30 (CH)	2	6.20 – (1H)
135.70 (C)	3	
135.20 (C)	4	
131.20 (CH)	5	7.30 – (1H)
39.40 (CH2)	6	1.50 – (2H)
33.70 (C)	7	
33.10 (CH2)	8	2.10 – (2H)
28.30 (CH3)	9	1.10 – (6H)
26.80 (CH3)	10	2.30 – (3H)
21.30 (CH3)	11	1.80 – (3H)
18.30 (CH2)	12	1.70 – (2H)

Figure 13. Input of macrocomponent with ^{13}C-NMR signal assignment.



13C-NMR chemical shift	No.	1H-NMR chemical shift -- 90 MHz
197.60 (C)	1	
142.30 (CH)	2	6.20 – (1H)
135.70 (C)	3	
135.20 (C)	4	
131.20 (CH)	5	7.30 – (1H)
39.40 (CH2)	6	1.50 – (2H)
33.70 (C)	7	
33.10 (CH2)	8	2.10 – (2H)
28.30 (CH3)	9	1.10 – (6H)
26.80 (CH3)	10	2.30 – (3H)
21.30 (CH3)	11	1.80 – (3H)
18.30 (CH2)	12	1.70 – (2H)

Figure 14. Candidate structures corresponding to input data Figure 12. (Shown in the right column are assignment information with ^{13}C-NMR signal number. No number for a carbon means that the carbon has not yet clearly assigned.)

application of McLafferty rearrangement become fully ready, the corresponding fragmentation is caused according to the rule. On the ground of "fragmentation rules" described in forms as above, *m/z* values are predicted. Incidentally, contents of some books have been incorporated in "fragmentation rules". Now, when *m/z* values for seven candidate structures shown in Figure 16 are predicted, the values are compared with the measured *m/z*, and the proportion of the number of *m/z* values agreeing with the measured versus the number of predicted *m/z* values is calculated as a sort of prediction factor for each candidate. This factor of prediction for each candidate was collected in Table 6. In the candidate structure no. 4, for instance, the number of predicted *m/z* values is 14 but that of m/z values thereof agreed with the measured is 10. Therefore the prediction becomes 71.4%. It may be regarded that the larger this value is the closer to

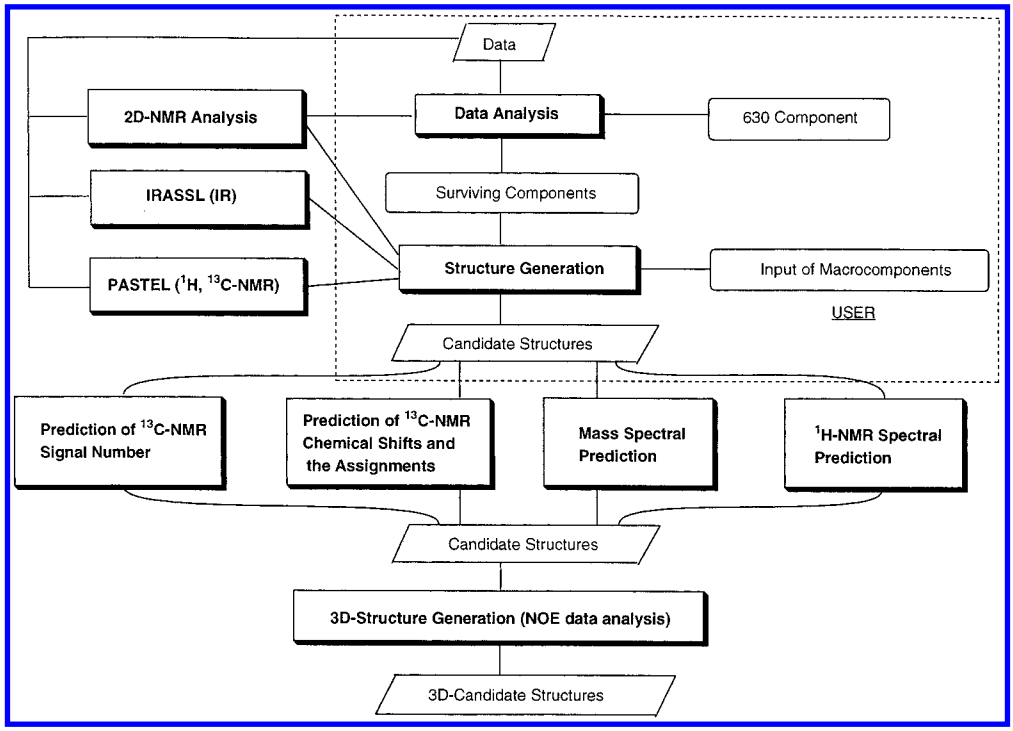


Figure 15. Construction diagram of recent CHEMICS.

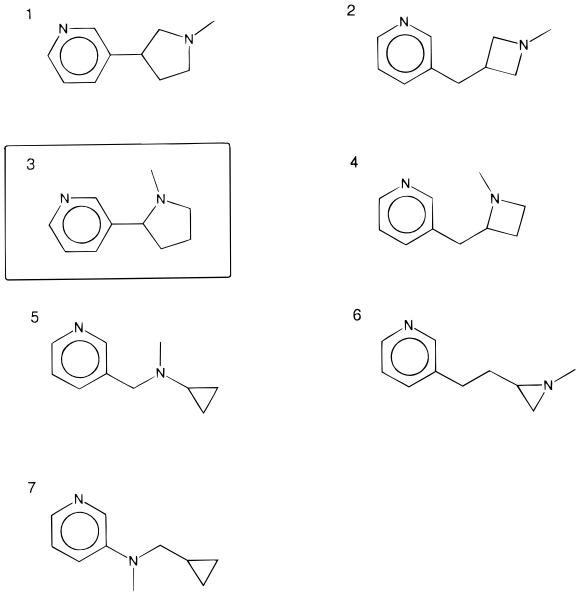


Figure 16. Candidate structures obtained with CHEMICS by using data shown in Table 5.

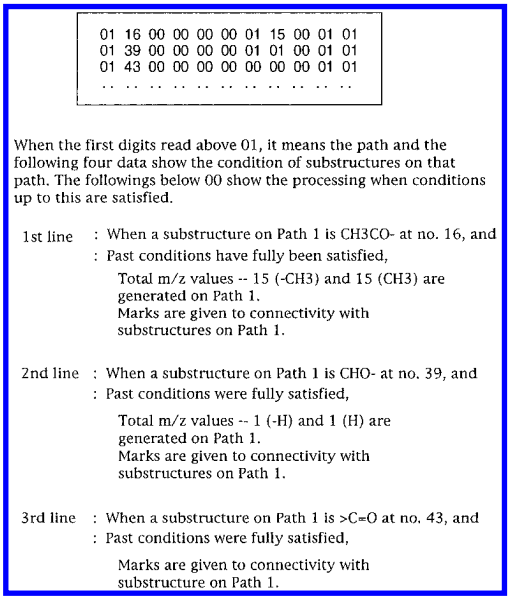


Figure 17. A part of the list of fragmentation rules for carbonyl group (external file form).

the true structure it is. Also for every candidate, collected together in Table 6 is the sum of each signal intensity of *m/z* values in a measured chart which agreed with the predicted. In this system, it has been programmed so as to predict important signals. It may be said therefore that the larger this value is the closer to the true structure it is. A graph which employed these two values was shown in Figure 19. If we judge from the character of each value, it can be regarded that the structure becomes closer to the truth as it goes toward.

Prediction of the ^{13}C-NMR chemical shift values was made for structures nos. 1 and 3 regarded to have a high possibility in using this examination program and a performed operation to squeeze the candidates through comparison with measured values.

Bangov reported investigations on the application of the linear relationship between the ^{13}C-NMR chemical shift and charge density for automation of the signal and structural assignment and ranking of candidate structures on the basis of standard approximation error of relationship.[14] Obviously, its reliability depends on the reliability of the computational scheme for charge density.

In our work, prediction of the ^{13}C-NMR chemical shift value is carried by means of index files which are arranged in correlation between the chemical shift values and the corresponding substructures up to ϵ level of the maximum environment which were induced from a large amount of measured data.[10] And the predicted chemical shift values are assigned to each carbon in the structure so that the sum of differences between the measured values and the chemical shift values is smallest in consideration of its multiplicity.

```
loop 1 ~ size of connectivity matrix
  if sec. component = 43 (>C=O)                              then   [ 1 ]

loop 1 ~ size of connectivity matrix
    if there exists a component connected with [ 1 ]         then
      if sec. component  = 24 (-CH2-) or  ...                then   [ 2 ]

loop 1 ~ size of connectivity matrix
    if there exists a component connected with [ 2 ]         then
      if sec. component  = 24 (-CH2-) or  ...                then   [ 3 ]

loop 1 ~ size of connectivity matrix
    if there exists a component connected with [ 3 ]         then
      if a component connected with [ 3 ] ≠ [ 2 ]            then
        if sec. component  = 24 (-CH2-) or  ...              then   [ 4 ]

          call subroutine (4 to 10 membered ring check)

        if  [ 1 ], [ 2 ], [ 3 ]  do not exist on the same ring or
            [ 2 ], [ 3 ], [ 4 ]  do not exist on the same ring  then

          - McLafferty rearrangement is probable -

           do cut the bond between [ 2 ]  and [ 3 ]
              and check the number of fragments
              on connectivity matrix

           if two fragments exist on the connectibity matrix     then

             calculate m/e values of      fragment including  [ 2 ] and
                                          fragment including  [ 3 ]
             move 1 from m/e value of     fragment including  [ 2 ]
                       to  m/e value of   fragment including  [ 3 ]
             move H from fragment including  [ 2 ]
                       to    fragment including  [ 3 ]
             add information that mass njmber 1 move to [ 1 ] from [ 4 ]

           endif
         endif
           endif
         endif
       endif
    endloop
       endif
     endif
    endloop
       endif
     endif
    endloop
       endif
    endloop
```

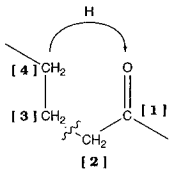


Figure 18. A part of McLafferty rearrangement search routine (in Fortran program form).

Here, the candidate with the smallest sum of these differences is regarded as the true answer. Figures 20 and 21 show the correspondence between predicted and measured values for the above candidates nos. 1 and 3, respectively. For candidate no. 1, correspondence at signal numbers 2 and 5 is poor, resulting in the sum of differences between predicted and measured values being 64.4. For candidate no. 3, on the other hand, both show a good agreement, and the sum of differences is 22.0, being much smaller than that of candidate structure no. 1. Accordingly, candidate structure no. 3, which is the true answer, could be taken up as the final candidate structure without difficulty.

Of course, the reliabily of prediction of ^{13}C-NMR chemical shift value depends on the quality and the quantity of the ^{13}C-NMR database from which the index files are derived. However, the merit of this method is to be able to get sufficient results from the experimental point of view without a complicated theoretical background.

Here, both examination programs have been made so as to be usable even singularly with data in Molfile form as the input.

By the way, now that prediction of the ^{13}C-NMR chemical shift value has become possible, it will be natural that the prediction of the ^{1}H-NMR spectrum is helped along with it. With regard to this, it has been so arranged that mean ^{1}H-NMR chemical shift values of the candidate are predicted by using the index files of chemical shift/substructure which are induced from ^{1}H-NMR spectral database containing signal assignment, information like appropriate coupling constants of the candidate is arranged in addition, and then evaluation and ranking of candidate structures can automatically be made by comparison with measured ^{1}H-NMR spectral data. Furthermore, we have developed in parallel a program supporting ^{1}H-NMR signal assignment to the structure by utilizing this program, which becomes necessary in cases including construction of ^{1}H-NMR database.[15]

While it is variously devised to raise utility as mentioned above, a problem due to structure generation, namely,

Table 5. Spectral Data of Nicotine[a]

chemical shift (ppm)	int.	chemical shift (ppm)	int.
	^{1}H-NMR Data (90 MHz)		
8.5	46	3.2	16
8.4	38	3.1	20
7.7	18	2.5	28
7.6	25	2.3	41
7.2	25	2.1	114
7.1	17	1.9	21
3.3	19	1.8	53

chemical shift (ppm)	int.	mult.	chemical shift (ppm)	int.	mult.
		^{13}C-NMR Data			
22.6	28	3	123.4	32	2
35.3	32	3	134.8	33	2
40.2	32	4	138.8	9	1
56.9	40	3	148.4	25	2
68.3	36	2	149.4	22	2

wave number (cm^{-1})	%T	wave number (cm^{-1})	%
	IR Data		
3680	88	1460	30
3400	58	1420	18
3080	65	1320	30
3020	50	1020	22
2960	12	920	64
2910	28	820	28
1590	58	620	60
1570	32		

[a] Molecular formula $C_{10}H_{14}N_2$

Table 6. Rate of Prediction for *m/z* Value for Each Candidate Structure and Sum of Observed Signal Intensities Corresponding to Predicted *m/z* Values

	candidate structure						
	1	2	3	4	5	6	7
prediction rate of *m/z* value	100.0	76.9	100.0	71.4	62.9	60.0	50.0
sum of intensities	176	69	189	78	26	20	22

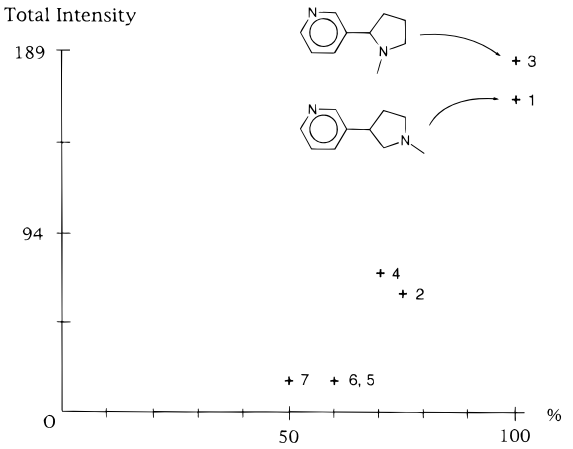


Figure 19. Ranking of candidate structures by mass spectral prediction.

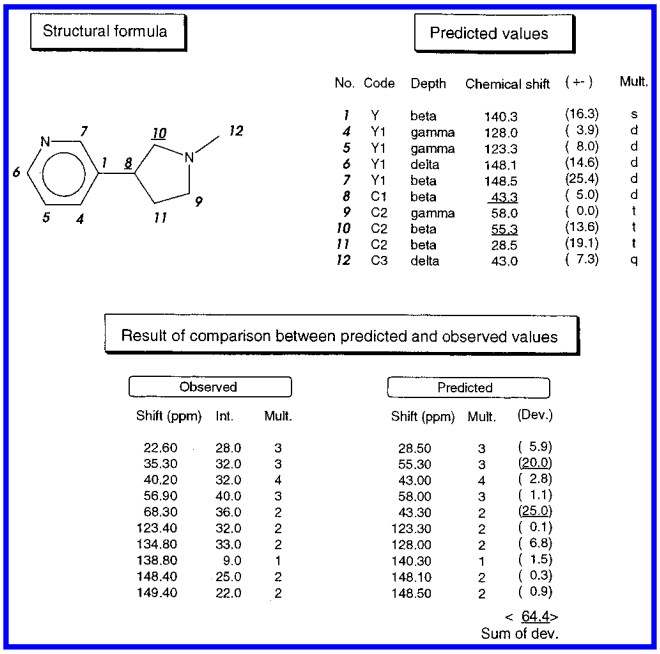


Figure 20. Prediction of the ^{13}C-NMR chemical shift values for the candidate structure no. 1 and its comparison with the measured values.

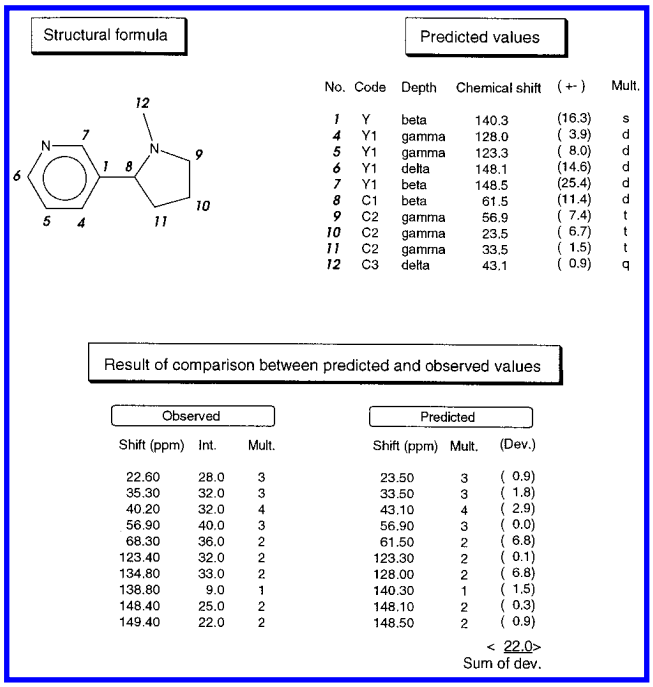


Figure 21. Prediction of the ^{13}C-NMR chemical shift values for the candidate structure no. 3 and its comparison with the measured values.

essential conditions for input of time and molecular formula needed for structure generation, has gradually occupied a growing weight as the next subject. Though performance of computers is advancing year after year, we face the former problem now of handling what has a large molecular formula. An essential settlement for this problem is not easy as long as the present algorithm of structure generation is being employed, but it is possible to generate candidate structures in the sequence from those of and near the true answer and let them be put out as an alternative method there to. In short, this is the scheme in which an analytically meaningful sequence is attached to components surviving at the spectral analysis by some means, component sets are made in the sequence from that more easily accepted by a consistency check with NM matrix for the component set while choosing the component in this sequence, and remarkable candidate structures including the true answer are allowed to generate in an early stage. This also shows the merit that the sequence is given to candidate structures from this, though of course, making study of the candidate structure easier. In order to attach a meaningful sequence in view of spectral analysis for the component here, we employed a sort of frequency distribution function for each component which had been induced from ^{1}H- and ^{13}C-NMR spectral database beforehand. Though we have to pay attention to congeniality between contents of the database and the unknown sample, it appears in general that the true answer and candidate structures in its vicinity have already been generated. In a certain case, the true answer may be generated last, without considering that mentioned above, but generated first if employing the function for this time. This function has already been realized on CHEMICS and presents relatively good results.[16]

CHEMICS was developed by enumerating all candidate structures while containing the true answer, for the purpose of automated structure analysis, at no conflict with the input information as its role. Many functions have been attached to that purpose, rising to the state regarded as practical.

Up to this point in time we have only discussed the structure of problems to be solved in utilizing a computer for the structure elucidation while giving an explanation of CHEMICS and its periphery. Eventually, how to let the computer execute must be determined by one, and how and to what extent in accordance with the philosophy of chemists when they cope with the structure analysis. Though we stated the use of the computer, it was never intended to conduct anything in different from the basic philosophy of problem solution in chemistry. The only difference is that the intuition, high-level, and ultralogic process of problem solution which functions only with people is an object that is difficult to logicalize for the computer, and it is hardly hoped that the computer will do this as well. And this may be better because the scene of computer utilization and the role people should play are made clear. Now provided that the problem solution is formulated, the computer executes the whole processing with the possible highest speed under that premise. Here, it can be well expected that exhaustiveness, which is one of the characters possessed by the computer, sometimes presents an unexpected answer with a full of surprise as well as presumable answers. Logic which is different from human intuition supplies materials of intuition and association in the opposite way and renders us an opportunity for evolution to an intellectual task in higher level.

REFERENCES AND NOTES

(1) Abe, H.; Fujiwara, I.; Nishimura, T.; Okuyama, T.; Kida, T.; Sasaki, S. Recent Advances in the Structure Elucidation System, CHEMICS. *Comput. Enhanced Spectrosc.* **1983**, *1*, 55−62.
(2) Funatsu, K.; Miyabayashi, N.; Sasaki, S. Further Development of Structure Generation in the Automated Structure Elucidation System, CHEMICS. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 18−28.
(3) 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*, 6−11.
(4) Christie, B. D.; Munk, M. E. The Application of two-Dimensional Nuclear Magnetic Resonance Spectroscopy in Computer-Assisted Structure Elucidation. *Anal. Chim. Acta* 1987, **200**, 347−361.

(5) Christie, B. D.; Munk, M. E. The Role of Two-Dimensional Nuclear Magnetic Resonance Spectroscopy in Computer-Enhanced Structure Elucidation. *J. Am. Chem. Soc.* **1991**, *113*, 3750−3757.
(6) Yamasaki, T.; Abe, H.; Kudo, Y.; Sasaki, S. CHEMICS: A Computer Program System for Structure Elucidation of Organic Compounds. *ACS Symp. Ser.* **1977**, *54*, 108−125.
(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) Funatsu, K.; Susuta, Y.; Sasaki, S. Application of Infrared Data Analysis Based on Synbolic Logic in Automated Structure Elucidation by CHEMICS. *Anal. Chim. Acta* **1989**, *220*, 155−169.
(9) Funatsu, K.; Del Carpio, C. A.; Sasaki, S. Quantitative Examination of the Relationship between 1H and 13C-NMR Chemical Shifts Applied to Structure Elucidation. *Comput. Enhanced Spectrosc.* **1986**, *3*, 119−131. Funatsu, K.; Del Carpio, C. A.; Sasaki, S. Computer-Assisted Structure Elucidation Based on the Interdependent Analysis of ^{1}H- and ^{13}C-NMR Spectra. *Comput. Enhanced Spectrosc.* **1986**, *3*, 133−140.
(10) Funatsu, K.; Eguchi, K.; Sasaki, S. Development of a System for 13C-NMR Chemical Shift Prediction with Aid of Index-file and for Ranking of Candidate Structures. In *Proceedings of the 16th Symposium of Chemical Information*; Tukihara, T., Ed.; Tokushima University: Tokushima, Japan, **1993**; pp 81−84.
(11) Funatsu, K.; Katsumi, H.; Sasaki, S. Computer Proguram for predinting the number of 13C-NMR Signals Based on Chemical Structure. *Comput. Enhanced Spectrosc.* **1986**, *3*, 87−90.
(12) Funatsu, K.; Nishizaki, M.; Sasaki, S. Introduction of NOE DAta to an Automated Structure Elucidation System, CHEMICS. Three-Dimensional Structure Elucidation Using the Distance Geometry Method. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 745−751.
(13) Funatsu, K.; Susuta, Y.; Sasaki, S. Application of the automated structure elucidation system (CHEMICS) to the chemistry of natural products. *Pure Appl. Chem.* **1989**, *61*, 609−612.
(14) Bangov, I. P. Use of the ^{13}C-NMR Chemical Shift/Charge Density Linear Relationship for Recognition and Ranking of Chemical Structures. *Anal. Chim. Acta* **1988**, *209*, 29−43.
(15) Funatsu, K.; Acharya, B. P.; Sasaki, S. Application of a Digital 1H-NMR Spectrum to the Survival Test of Substructures and the Assignment Test. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 735−744.
(16) Hayami, K.; Funatsu, K.; Sasaki, S. Improvement in the structure generation of the automated structure elucidation system for organic compounds, CHEMICS. *Bunseki Kagaku* **1993**, *42*, 369−374.

CI950152R