# Introduction of Two-Dimensional NMR Spectral Information to an Automated Structure Elucidation System, CHEMICS. Utilization of 2D-INADEQUATE Information

KIMITO FUNATSU,* YUTAKA SUSUTA, and SHIN-ICHI SASAKI

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

A program for applying two-dimensional double-quantum coherence $^{13}C$ NMR spectroscopic data (2D-INADEQUATE data) has been introduced into CHEMICS, a system for automated structure elucidation. The main concepts employed in this study consist of making a more accurate assignment of $^{13}C$ chemical shifts to carbons in the sample structure at the data analysis step in CHEMICS followed by the generation of more probable structures as candidates using $^{13}C$–$^{13}C$ coupling information. A detailed algorithm is described together with the structure elucidation of some of the terpenoids. Application of the 2D-INADEQUATE technique to CHEMICS results has drastically decreased the number of probable candidate structures in comparison with the results afforded by the present system.

## INTRODUCTION

The automated structure elucidation system of organic compounds, CHEMICS, presents all possible structures that are consistent with an unknown's spectroscopic data and molecular formula.[1,2] As spectroscopic data, IR, $^1H$ NMR, $^{13}C$ NMR spectra have hitherto been employed. In addition to these, the information provided by 2D NMR technique has now been introduced into the CHEMICS system. Munk et al. have reported on the application of 2D NMR spectral data to their structure elucidation system, CASE.[3] 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 CASE system.

As a first trial, we have noted information obtainable from 2D-INADEQUATE experiments among the 2D NMR techniques.[4] The technique shows a pair of signals that connect through the coupling between the $^{13}C$–$^{13}C$ nuclei. Another technique, H–H COSY, including coupling information through space as well as the bond, cannot directly be applied to structure generation in CHEMICS, because the structure is represented in the topological level in the system. Thus, only 2D-INADEQUATE information has been introduced into the data analysis in CHEMICS.

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 chemical structure. 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 eliminated. (3) If NMR data 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.[5] (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

**Table I.** Spectral Data of L-Menthol[a] ($C_{10}H_{20}O$)

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

[a] Nakanishi, K. *Superconductivity FT-NMR*; Kodansha: Tokyo, 1986; pp 126–127.

to determine whether it is consistent with all of the spectroscopic data on the basis of the NM submatrix constructed by the NM matrix elements, corresponding to each component in the set.[5] (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.[2,6] Here, the connectivity matrix stands for the connective relationships of tertiary components to construct full structure(s).

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 atoms in a sample structure to $^{13}C$ NMR signal by the introduction of coupling information of the $^{13}C$–$^{13}C$ nuclei to the NM submatrix. The clearer the assignment correlation becomes, the fewer the number of candidates are responded.

Information from 2D-INADEQUATE has been introduced into CHEMICS. The analytical results are excellent, as described later.

## METHOD

Here, the process of structure elucidation is shown for L-menthol as an unknown sample. In this case, first the molecular formula and spectroscopic data of $^1H$ and $^{13}C$ NMR of the sample are input. They are shown in Table I. According to the present CHEMICS (without 2D-INADEQUATE technique), 20 tertiary components are given as components that are consistent with the sample's data, as shown in Table II. Then 219 candidate structures are constructed on the basis of those 20 components referring to its molecular formula and spectral data.
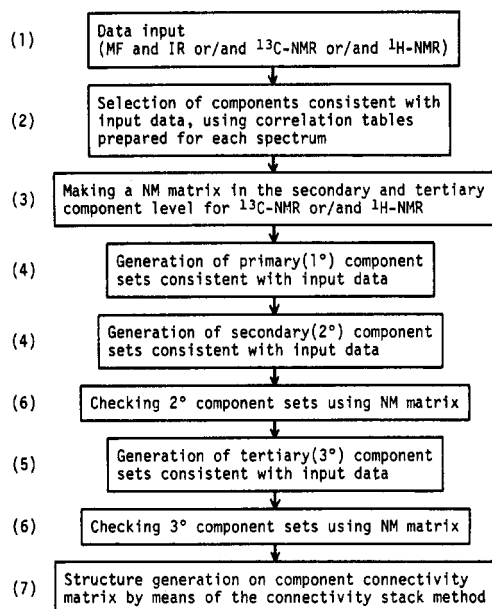
UTILIZATION OF 2D-INADEQUATE INFORMATION

*J. Chem. Inf. Comput. Sci., Vol. 29, No. 1, 1989* **7**



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

**Table II.** Components Surviving Spectral Analysis of CHEMICS

| no. | tertiary component | afferent nature[a] | no. | tertiary component | afferent nature[a] |
|---|---|---|---|---|---|
| 1 | isopro— | O | 11 | :CH— | O |
| 2 | isopro— | CD | 12 | :CH— | CD |
| 3 | isopro— | CS | 13 | :CH— | CS |
| 4 | $CH_3CH_2$— | CD | 14 | —CH= | O |
| 5 | $CH_3CH_2$— | CS | 15 | —CH= | CD |
| 6 | :$CHCH_3$ | O | 16 | —CH= | CS |
| 7 | :$CHCH_3$ | CD | 17 | —OH | CD |
| 8 | :$CHCH_3$ | CS | 18 | —OH | CS |
| 9 | $CH_2$< | CD | 19 | —O— | CD |
| 10 | $CH_2$< | CS | 20 | —O— | CS |

[a] Implies attribute of connection partner[3]: CD, $sp^2$ carbon; CS, $sp^3$ carbon; O, $sp^3$ oxygen.

**Table III.** Signal–Signal Connectivity Information of L-Menthol[a]

| | | | | | signal no. | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| connectivity with signal no. | 2 | 1 | 1 | 5 | 3 | 2 | 2 | 5 | 6 | 6 |
| | 3 | 6 | 5 | 7 | 4 | 9 | 4 | | | |
| | | 7 | | | | 8 | 10 | | | |

[a] Nakanishi, K. *Superconductivity FT-NMR*; Kodansha: Tokyo, 1986; pp 152–155.

In addition to the present analysis, the analytical process incorporating 2D-INADEQUATE information is described, focusing on the correlation between components selected from the correlation table of components and chemical shifts and each $^{13}C$ NMR signal of the sample. The overall flow of the 2D-INADEQUATE analysis is shown in Figure 2. For simplicity, the analysis is described by dividing the flow into four stages.

The 2D-INADEQUATE technique gives coupling information between the $^{13}C$–$^{13}C$ nuclei (hereafter referred to as signal–signal connectivity information) of the sample, as shown in Table III. The information is effectively used in each stage of the analysis.

**Stage 1.** Among the respective NM submatrixes produced from the sets of secondary as well as tertiary components, the one satisfying the signal–signal connectivity information was selected. In this examination (check 1), the components possessing plural different carbon atoms, belonging to different levels of secondary and tertiary components, are taken into consideration. The components $(CH_3)_2CH$– and $CH_3CH$< and others correspond to this. The signals corresponding to
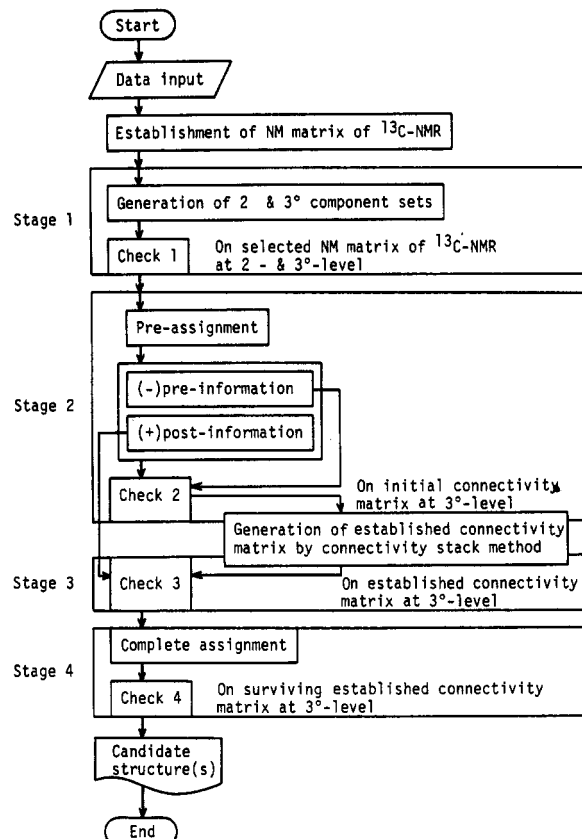


**Figure 2.** Flow chart of the 2D NMR spectral analysis.



**Figure 3.** NM submatrix for a certain secondary component set.

**Table IV.** Surviving Tertiary Component Set

| component type | tertiary component | afferent nature[a] | no. of |
|---|---|---|---|
| A | isopro– | CS | 1 |
| B | :$CHCH_3$ | CS | 1 |
| C | –$CH_2$– | CS | 3 |
| D | :CH– | O | 1 |
| E | :CH– | CS | 1 |
| F | –OH | CS | 1 |

[a] Implies attribute of connection partner.

those components are examined to determine whether they satisfy the signal–signal connectivity information. As an example, a certain NM submatrix corresponding to the set of secondary components is illustrated in Figure 3. The signal numbers corresponding to $(CH_3)_2CH$– group are 6, 9, and 10. Since the pairs of 6 and 9 as well as 6 and 10 exist, the examination is undertaken on the basis of the signal–signal connectivity information. Similarly, a pair of 5 and 8 is investigated for $CH_3CH$<. The NM submatrix shown in Figure 3 is said to pass the examination since all pairs of components do not contradict in view of signal–signal connectivity information.

In the case of L-menthol, six sets of secondary components are generated by the present CHEMICS. Four of the sets are eliminated by adding 2D-INADEQUATE information. The number of sets of tertiary components generated from the remaining two sets of secondary components is decreased from
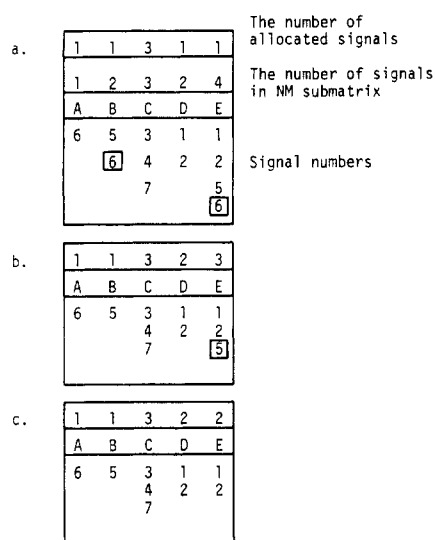
**Figure 4.** Process for obtaining preliminary assignment information (A, B, C, ... in this figure correspond to Table IV): (a) initial stage; (b) elimination of signal number 6 corresponding to B and E from initial stage a; (c) elimination of signal number 5 corresponding to E from second stage b.



**Figure 5.** Connectivity matrix divided into 11 parts by combination of the component pairs.



**Figure 6.** Connectivity matrices before (left) and after (right) use of the connectivity stack method: (-2) absolute nonlinkage; (-1) absolute linkage.

**Table V.** Component Connection 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 | + |

[a] (+) absolute linkage; (–) absolute nonlinkage.

four to one by a similar examination. The set of tertiary components thus obtained is shown in Table IV. Forty-six structures come out of the set if further examination is not carried out.

**Stage 2.** A unique correspondence obtained between the $^{13}$C NMR signal and the corresponding tertiary component in the NM submatrix for the set of tertiary components surviving from the above examination is called an unconditional correlation, and all of the possible assignments are derived from this correlation. It is not always necessary to obtain a complete assignment at this stage, since the assignment proceudre is undertaken again at stage 4. The correlation information of the assignment so far obtained is referred to as the preliminary assignment information. Next the pairs of tertiary components having no possibility to be connected to each other [(–)-connection] are obtained, and component pairs that need to be connected [(+)-connection] are examined on the basis of the preliminary assignment information and signal–signal connectivity information. This is called component connection information, and (–)-connection information is called a precondition since it puts a constraint on the examination that is undertaken prior to structure generation by the connectivity stack method. (+)-Connection information is called a postcondition since it is used after structure generation. The process up to the examination (check 2) of the connectivity matrix by precondition is called stage 2.

To obtain the pre- and postconditions necessary for this and the next stage's examination, preliminary assignment information is needed. First, the unconditional correlation between components and signals in the NM submatrix should be given sufficient attention. Figure 4 shows the process to obtain such information for the L-menthol sample. In this case, $(CH_3)_2$-CH– is regarded to be a component corresponding to the relation. In other words, only signal 6 is unconditionally assigned to A. Therefore, signal 6 in B and E in Figure 4a is obtained. Signal 5 in E is similarly eliminated because component B and signal 5 have a unique relationship as shown in Figure 4b. Finally, the preliminary assignment information is obtained as shown in Figure 4c. More detailed assignment will be carried out in stage 4.

Next the component connection information—pre- and postcondition—are searched for on the basis of the preliminary assignment information and signal–signal connectivity information.
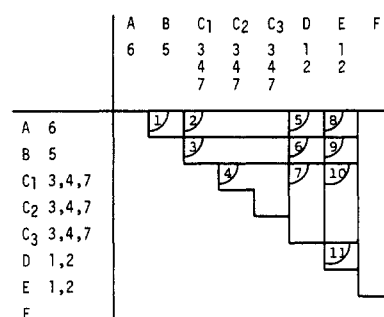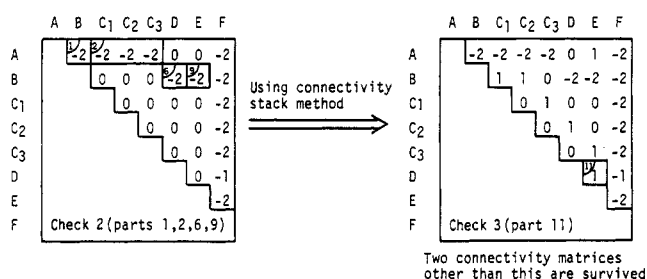
Precondition: When a connection between two components is judged to be not possible by the preliminary assignment information and signal–signal connectivity information, such connection is called a (–)-connection.

Postcondition: When two components are judged to be connectable, the connection is called a (+)-connection. For L-menthol, the elements of the connectivity matrix for a set of tertiary components now in consideration are divided into 11 parts by the combination of component pairs (Figure 5). Each part is examined to determine whether its connection is minus, plus, or uncertain. The component connection information for L–menthol obtained is shown in Table V. As can be seen from the table, 1, 2, 6, and 9 are minus and 11 is plus.

The connectivity matrix set up on the basis of the set of tertiary components that survived is called the initial connectivity matrix. Prior to the precondition examination, –2 (absolute nonlinkage) or –1 (absolute linkage) was established on the connectivity matrix through the attributes of the components.[3] In the case of L-menthol, the initial connectivity matrix prepared by a set of tertiary components that survived and passed the precondition examination (Figure 6, left). Successively, 1 or 0 was used as the undetermined element in the initial connectivity matrix using the connectivity stack

**Table VI.** Signal–Signal Connectivity Information for a Sesquiterpene[a]

| | | signal no. | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| connectivity with signal no. | 4 | 3 | 2 | 1 | 1 | 1 | 2 | 6 | 6 | 9 | 5 | 9 | 9 | 2 | 6 |
| | 5 | 7 | | | 7 | 8 | 5 | 7 | 10 | 11 | 10 | | | | |
| | 6 | 14 | | | 11 | 9 | 8 | | 12 | | | | | | |
| | | | | | | 15 | | | 13 | | | | | | |

[a] Benn, R.; Klein, J. *Magn. Reson. Chem.* **1986**, *24*, 638.

method. In the end, a full structure was constructed. When the precondition was introduced into CHEMICS, the number of candidate structures was dramatically decreased to five.

**Stage 3.** An examination of the connectivity matrix (check 3) was undertaken by using the postcondition, one element of the component connection information. At this time, each element in the connectivity matrix was already decided by the connectivity stack method. Such a connectivity matrix is called the established connectivity matrix. The five established connectivity matrices generated at the previous stage were examined through the postconditions. Two of them could not satisfy the examination. One of three is shown in Figure 6, right. Only the connectivity matrix giving (+)-connection toward part 11 passed the examination.

**Stage 4.** At the last step, the connectivity matrix was examined (check 4) on the basis of the information obtained during the assignment procedures. The assignment that resulted was not consistent with the signal–signal connectivity information; such an established connectivity matrix can be eliminated. The correlation information of the assignment obtained through these processes is called the complete assignment information.

The preliminary assignment information obtained at stage 2 is taken as the initial data, and an unconditional assignment was made for each element having 1 in the established connectivity matrix on the basis of the signal–signal connectivity information.

The fact that the element is 1 in the NM matrix indicates a meaningful correlation exists between the component and its corresponding signal. However, when the element already having 1 is changed to 0 by the signal–signal connectivity information, it means the negation of the above meaningful correlation. Suppose, for example, that one component corresponds to more than one signal. If it is possible to decrease, even by one, the number of corresponding signals through the signal–signal connectivity information and so forth, the assignment of the component to the signal becomes clearer.

There is a case in which the assignment relationship is decided before all of the elements having 1 in the established connectivity matrix are examined. In this case, the assignment correlation once decided is examined again for the remaining elements having 1 in the established connectivity matrix, and if a contradiction arises, such a connectivity matrix is discarded. As shown in Figure 7 for L-menthol, in the intermediate stage to obtain respective complete assignment information for three candidates, the two established connectivity matrices shown in Figure 7a,c were eliminated because they produced assignments that did not satisfy the signal–signal connectivity information. Finally, the established connectivity matrix shown in Figure 7b, which passed the examination, correctly gave the structure of the sample in a 2D form. For the sake of reference, all the candidates provided by the three connectivity matrices shown in Figure 7 are illustrated in Figure 8.
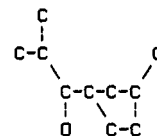
## RESULTS AND DISCUSSION

Another application of the above-mentioned method to an isomer of thujopsene (no. 3 in Figure 9) is described below. The signal–signal connectivity information for the sample's
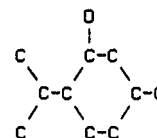
**a.**

| | A | B | C1 | C2 | C3 | D | E | F | Component (Assigned signal No.) | Ground for ←the assignment | Process of assignment check |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A | -2 | -2 | -2 | -2 | 1 | 0 | -2 | | A (6) | | 1  C2 - E  OK |
| B | | 1 | 1 | 0 | -2 | -2 | -2 | | B (5) | | 2  C3 - E  NO |
| C1 | | | 0 | 1 | 0 | 0 | -2 | | C1 (4) | ← C1 - C3 | 3  D - E  OK |
| C2 | | | | 0 | 0 | 1 | -2 | | C2 (3) | ← C1 - C3 | 4  B - C1  OK |
| C3 | | | | | 0 | 1 | -2 | | C3 (7) | ← C1 - C3 | 5  B - C2  OK |
| D | | | | | | 1 | -1 | | D (2) | ← A - D | |
| E | | | | | | | -2 | | E (1) | ← A - D | |
| F | | | | | | | | | | | *Out* |

**b.**

| | A | B | C1 | C2 | C3 | D | E | F | Component (Assigned signal No.) | Ground for ←the assignment | Process of assignment check |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A | -2 | -2 | -2 | -2 | 0 | 1 | -2 | | A (6) | | 1  C2 - D  OK |
| B | | 1 | 1 | 0 | -2 | -2 | -2 | | B (5) | | 2  C3 - E  OK |
| C1 | | | 0 | 1 | 0 | 0 | -2 | | C1 (4) | ← C1 - C3 | 3  D - E  OK |
| C2 | | | | 0 | 1 | 0 | -2 | | C2 (3) | ← C1 - C3 | 4  B - C1  OK |
| C3 | | | | | 0 | 1 | -2 | | C3 (7) | ← C1 - C3 | 5  B - C2  OK |
| D | | | | | | 1 | -1 | | D (1) | ← A - E | |
| E | | | | | | | -2 | | E (2) | ← A - E | |
| F | | | | | | | | | | | *Passed* |

**c.**

| | A | B | C1 | C2 | C3 | D | E | F | Component (Assigned signal No.) | Ground for ←the assignment | Process of assignment check |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A | -2 | -2 | -2 | -2 | 0 | 1 | -2 | | A (6) | | 1  C2 - E  NO |
| B | | 1 | 1 | 0 | -2 | -2 | -2 | | B (5) | | 2  C3 - D  NO |
| C1 | | | 0 | 1 | 0 | 0 | -2 | | C1 (4) | ← C1 - C3 | 3  D - E  OK |
| C2 | | | | 0 | 0 | 1 | -2 | | C2 (3) | ← C1 - C3 | 4  B - C1  OK |
| C3 | | | | | 1 | 0 | -2 | | C3 (7) | ← C1 - C3 | 5  B - C2  OK |
| D | | | | | | 1 | -1 | | D (1) | ← A - E | |
| E | | | | | | | -2 | | E (2) | ← A - E | |
| F | | | | | | | | | | | *Out* |

**Figure 7.** Examination of three established connectivity matrices in the process of obtaining complete assignment information for L-menthol.

STRUCTURE NO. = 1

```
        C
        |
    C-C        C
       \      /
        C-C-C-C
        |    \  |
        O     C-C
```

STRUCTURE NO. = 2

```
             O
             |
    C        C-C
     \      /    \
      C-C          C-C
     /     \      /
    C       C-C
```

STRUCTURE NO. = 3

```
             O
             |
    C        C-C
     \      /    \
      C-C          C
     /     \       |
    C       C-C    |
                \  |
                 C--C
                    \
                     C
```
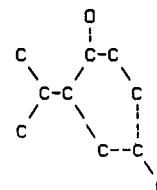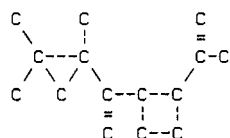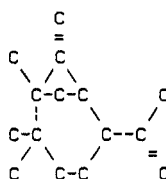
**Figure 8.** Candidate structures for L-menthol at stage 3 (the structure numbers correspond to the connectivity matrices a–c in Figure 7).

spectrum is listed as shown in Tables VI and VII. The number of candidates provided for the sample by standard CHEMICS using a molecular formula and $^{13}C$ and $^{1}H$ NMR data and without 2D-INADEQUATE information was 4450.

STRUCTURE NO.=    1          STRUCTURE NO.=    4          STRUCTURE NO.=    7

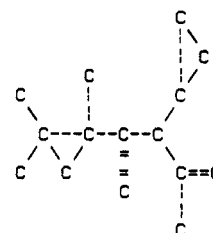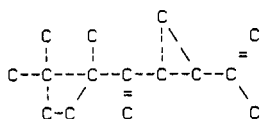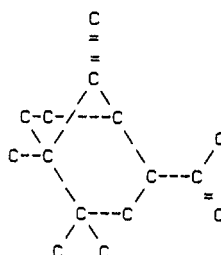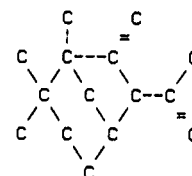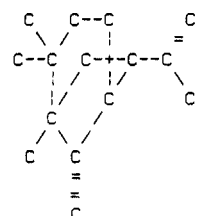STRUCTURE NO.=    2          STRUCTURE NO.=    5          STRUCTURE NO.=    8
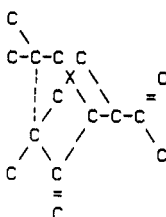
STRUCTURE NO.=    3                                       STRUCTURE NO.=    9
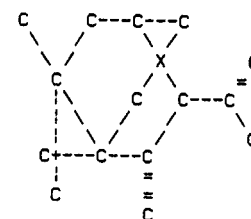
                             STRUCTURE NO.=    6

**Figure 9.** Candidate structures for an isomer of thujopsene at stage 3.

**Table VII.** Input Data for a Sesquiterpene[a] ($C_{15}H_{24}$)

| 200-MHz $^1$H NMR data, ppm (rel intensity) | $^{13}$C NMR data, ppm (rel intensity) (multiplicity) |
|---|---|
| 0.25 (24) | (1) 160.06 (8) (s) |
| 0.34 (12) | (2) 149.92 (8) (s) |
| 0.45 (4) | (3) 108.67 (23) (t) |
| 0.50 (4) | (4) 101.73 (20) (t) |
| 0.88 (8) | (5) 49.92 (22) (d) |
| 1.01 (12) | (6) 49.66 (9) (s) |
| 1.10 (4) | (7) 47.02 (18) (d) |
| 1.53 (4) | (8) 40.38 (22) (t) |
| 1.65 (4) | (9) 38.78 (12) (s) |
| 1.82 (4) | (10) 35.21 (19) (t) |
| 3.87 (4) | (11) 32.46 (23) (t) |
| 3.92 (4) | (12) 25.70 (23) (q) |
| 4.02 (4) | (13) 23.47 (18) (q) |
| 4.14 (4) | (14) 21.25 (16) (q) |
|  | (15) 18.13 (18) (q) |

[a] Benn, R.; Klein, J. Magn. Reson. Chem. **1986**, 24, 638.

**Table VIII.** Number of Candidate Structures at Respective Stages

|  | no. of candidates | | |
|---|---|---|---|
|  | menthol | sesquiterpene | linalool |
| $^1$H, $^{13}$C NMR (ordinal analysis) | 219 | 4450 | 48 |
| by check 1 (stage 1) | 46 | 920 | 3 |
| by check 1 + check 2 (stage 2) | 5 | 17 | 3 |
| by check 1 + check 2 + check 3 (stage 3) | 3 | 9 | 2 |
| by check 1 + check 2 + check 3 + check 4 (stage 4) | 1 | 1 | 1 |

Employment of the signal–signal connectivity information diminished 4450 to only one correct structure. Table VIII

$C_{15}H_{22}O_3$   α-Santonin

| $^{13}$C-NMR | | C-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 |
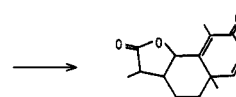
$C_{15}H_{22}O_3$   11,13-Dihydroparthenolide

| $^{13}$C-NMR | | C-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 |

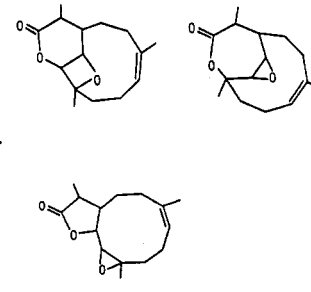(F.J. Parodi et al., SPECTROSCOPY LETTERS, 20(1987)445)

**Figure 10.** Examples for execution of the present CHEMICS.

expresses the number of candidate structures given for the three terpenoids at respective stages from 1 to 4.

Two more examples are illustrated in Figure 10. Using the molecular formulas, $^{13}$C NMR data, and $^{13}$C–$^{13}$C signal connectivity information. CHEMICS generated one and three candidates for the respective samples. It is interesting that in the second example in Figure 10 three candidates are

consistent with the signal connectivity information. In other words, this result indicates that the assignment possibility is not always one even when complete signal connectivity information is available.

As shown above, the introduction of the signal–signal connectivity information to CHEMICS turned out to be extremely effective. It must be emphasized that the information of the 2D NMR has been utilized by the simple treatment of the NM submatrix that has been used by CHEMICS so far.

Although the 2D-INADEQUATE method still has problems regarding sensitivity and measurement time, the information provided by the method is extremely useful for structure clarification. In the present study, the combination of the 2D-INADEQUATE technique and the fundamental idea of CHEMICS, where all possibilities without any deviation from the true structure should be listed, allowed structure elucidation that was more effective and correct for a variety of organic compounds.

The application of other 2D NMR techniques to CHEMICS is now under investigation in a similar manner to that of

the 2D-INADEQUATE technique. The procedure reported here is just the beginning of applying such a series of 2D NMR information to the CHEMICS system.

The present program was developed on super minicomputer MV-2000DC (Data General Corp.) and written in FORTRAN 77.

## ACKNOWLEDGMENT

We thank the computer center, Institute for Molecular Science, for allowing us the use of their facilities.

## REFERENCES AND NOTES

(1) Abe, H.; Fjuiwara, I.; Nishimura, T.; Okuyama, T.; Kida, T.; Sasaki, S. *Comput. Enhanced Spectrosc.* **1983**, *1*, 55.
(2) Funatsu, K.; Miyabayashi, N.; Sasaki, S. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 18.
(3) Christie, B. D.; Munk, M. E. *Anal. Chim. Acta* **1988**, *200*, 347.
(4) Bax, A.; Freeman, R.; Kempsell, S. P. *J. Am. Chem. Soc.* **1980**, *102*, 4849.
(5) Yamasaki, T.; Abe, H.; Kudo, Y.; Sasaki, S. *ACS Symp. Ser.* **1977**, *54*, 108.
(6) Kudo, Y.; Sasaki, S. *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 43.

# An Expert-Database System for Sample Preparation by Microwave Dissolution. 1. Selection of Analytical Descriptors[†]

F. A. SETTLE, JR.,*,‡ B. I. DIAMONDSTONE, and H. M. KINGSTON

Center for Analytical Chemistry, National Measurement Laboratory, National Bureau of Standards, Gaithersburg, Maryland 20899

M. A. PLEVA

Department of Chemistry, Washington and Lee University, Lexington, Virginia 24450

A hybrid expert-database system is being developed to provide advice on the preparation of samples for elemental analysis. This paper describes an expert system component that is designed to assist the analyst in the identification of four analytical descriptors necessary to develop procedures for sample preparation. When completed, the system will be able to furnish information on the dissolution of the sample. Future versions of the system will also provide advice on separations that may be required prior to the analytical measurement. A PC-AT microcomputer and commercially available software were used to develop the system. A compiled version of the system will run on PC-compatible computers.

## INTRODUCTION

Sample preparation remains an important component of chemical analysis despite the advances made in instrumentation. In inductively coupled plasma spectrometry (ICP), for example, approximately 90% of the analyses require preparation of a liquid sample.[1] Nevertheless, the instrumental measurement often receives much more attention than does the preparation of the sample. As a result, persons with experience in sample preparation have become "endangered species" in many laboratories. Thus, it becomes important to capture the expertise of these analysts and to make it available to the analytical community.

Transformation of the sample into a form that can be analyzed by a specific measurement technique often requires tedious, time-consuming procedures such as dissolution, dilu-

**Table I.** Relation between Analytical Descriptors and Conditions for Microwave Dissolution of a Wheat Flour Matrix

| analytical descriptor | utility |
|---|---|
| SRM matrix corresponding to sample | determines the acid(s) to be used for dissolution, temperature for dissolution (target temperature), and time to reach target temperature |
| analyte element | determines the acid(s) to be used for dissolution and in some cases the target temperature |
| level of analyte concentration | determines the amount of sample to be dissolved |
| instrumental technique | does not usually affect the conditions used for dissolution but is important in determining conditions for subsequent separations |

---

[†] Certain commercial equipment and software are identified in this paper to specify adequately the experimental procedure. Such identification does not imply recommendation or endorsement by the National Bureau of Standards, nor does it imply that the equipment or software is necessarily the best available for the purpose.
[‡] Professor of Chemistry, Virginia Military Institute, Lexington, VA 24450, visiting scientist in the Center for Analytical Chemistry 1987–1988.

tion, and separation. These procedures in turn depend upon four analytical descriptors: (1) the type of matrix represented by the sample; (2) the analyte(s); (3) the expected concentration level of the analyte; and (4) the instrumental technique to be used for the measurement. Once these four descriptors