Application of a Digital ^{1}H-NMR Spectrum to the Survival Test of Substructures and the Assignment Test

Kimito Funatsu,* Binod P. Acharya, and Shin-ichi Sasaki

Department of Knowledge-Based Information Engineering, Toyohashi University of Technology, Tempaku, Toyohashi 441, Japan

Received September 1, 1993®

A computer program for testing the survival of a substructure in an unknown by calculating theoretical ^{1}H-NMR subspectra has been developed. In addition, a program for assisting ^{1}H-NMR signal assignment on the basis of the above program has been developed. The usefulness of these programs is presented using several examples.

I. INTRODUCTION

The application of ^{13}C and ^{1}H nuclear magnetic resonance spectroscopy to the determination of organic structures is a well developed branch of chemistry. Moreover, without going into the complex theory in detail one could apply the spectrum–structure relationship property to the structure determination. Thus, many researchers are tempted to develop a system which elucidates the structure of an unknown organic compound from its spectroscopic information.

In structure elucidation processes, in order to judge effectively whether the substructure of interest exists positively, the analyses by some parameters of a ^{1}H-NMR spectrum become indispensable. In the present study, the survival test program has been developed to judge whether the observed ^{1}H-NMR spectrum of an unknown supports the existence of the focused substructure, on the basis of the mean ^{1}H-NMR chemical shifts predicted for each atomic group (analytical units in this program, Table 1) in the substructure. Furthermore, in the latter half of this paper the program for assisting the assignment of ^{1}H-NMR spectra, which is developed by the modification of the above survival test program, will be described.

II. SURVIVAL TEST OF A SUBSTRUCTURE BY ^{1}H-NMR SPECTRUM

The newly developed program tests the survival of a substructure not with the predicted mean ^{1}H-NMR chemical shifts but with the parameters of the predicted ^{1}H-NMR subspectra (mean ^{1}H-NMR chemical shifts, multiplicity of the signals, integration of the signals, and resolution among the signals). In this paper the program performance will be described.

The program computes a theoretical subspectra, from the mean ^{1}H-NMR chemical shifts predicted by a previous method[1] and the topological information on a substructure, for each individual atomic group of the substructure in question and calculates parameters to describe the subspectra. The parameters are then matched with the similar subspectral parameters derived from the input observed ^{1}H-NMR spectra. The survival of an individual atomic group of the substructure is decided by its performance in the matching, and then the survival of the substructure is decided by an average of the performance of all its atomic groups. The flow chart of the survival test of a substructure is shown in Figure 1.

Table 1. Atom and Atomic Group Units[a]

code no.		code no.		code no.	
1	>C<	17	=N–	33	–NH_2
2	>CH–	18	=NO–	34	–NO
3	>C=	19	–ACH–	35	–NO_2
4	>N–	20	–AO–	36	–NS
5	>NO–	21	–AS–	37	–NSO
6	>AC–	22	–AN	38	–N_3
7	–CH–	23	–ANH–	39	=CH_2
8	–C≡	24	–ANO–	40	=C=O
9	–CH_2–	25	–ASO–	41	=NH
10	>C=O	26	–ASO_2–	42	=N=N
11	–O–	27	CH_3–	43	=S
12	–S–	28	–CH=O	44	≡CH
13	–NH–	29	–C≡N	45	–F
14	–SO–	30	–N≡C	46	–Cl
15	–SO_2–	31	–OH	47	–Br
16	=C=	32	–SH	48	–I

[a] The symbol "A" in the codes stands for aromatic character.

(A) Program Constraints. (1) Simplification of the Digital ^{1}H-NMR Spectrum Input. Instead of mean chemical shifts and proton counts of a ^{1}H-NMR spectrum, the digital ^{1}H-NMR spectrum is input to the system. The digital spectrum is the simplified one. In the simplification of a digital ^{1}H-NMR spectrum, the undesired signals (TMS, lower noises, etc.) are removed, signal intensities are standardized (one proton equivalent signals are made, 10 000), the resolution of the spectrum is lowered to 4 Hz, signals due to methyl protons are identified, and the signals are grouped according to their relative vicinity (Table 2). Then, the simple parameters (geometric mean chemical shift, total integration, number of signals in the group, and the mean resolution among the signals in hertz) are calculated to each grouped band. This simplification procedure has been optimized with several digital spectra so that any important information is not lost in the simplification.

(2) Calculation of the Theoretical Subspectrum. The theoretical ^{1}H-NMR subspectrum is calculated by a semiempirical quantum chemical method. This calculation program has been derived from LAOCOON3[2] and modified for our purposes. The prerequisites of this method, the coupling constant value among ^{1}H nuclei from different atomic groups, are computed, from correlation tables of coupling values and the substructures stored in the program, by recognizing the substructural patterns from its connectivity matrix, whereas the mean ^{1}H chemical shifts of different atomic groups are predicted by the previous method.[1] A part of the correlation table is shown in the Table 3. A detailed description of the

® Abstract published in *Advance ACS Abstracts*, June 1, 1994.

0095-2338/94/1634-0735$04.50/0

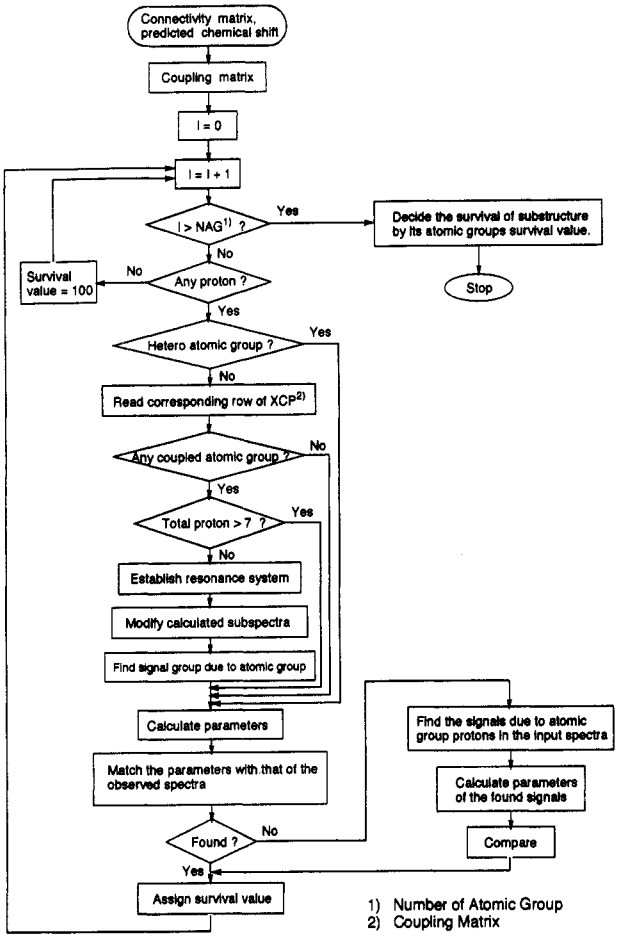


Figure 1. Flow chart of the survival test algorithm.

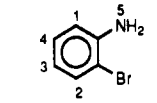
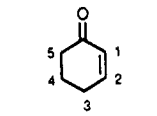
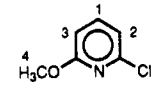
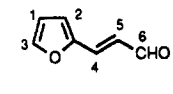
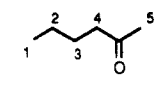
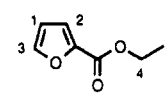
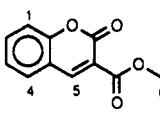
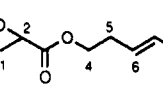
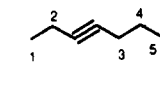
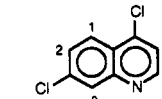
Table 2. Vicinity Range for the Grouping Signals in the ^{1}H-NMR Spectra Simplification

chemical shift range (ppm)	vicinity (Hz)	chemical shift range (ppm)	vicinity (Hz)
–4	24	7–8	12
4–6	18	8–9	10
6–7	16	9–	8

assignment of coupling constants is given in Appendices A1 and A2. A symmetrical matrix of the coupling constants among atomic groups in the substructure is then deduced. Reading along a row of this "coupling matrix" tells the coupling relation between the corresponding atomic group and the coupled atomic group(s).

Limitations in the quantum chemical calculation method do not allow us to calculate the total subspectrum of the substructure which may have more than seven ^{1}H nuclei. Hence, the subspectra for the environment of each individual atomic group are calculated. A resonating ^{1}H nuclei system, for the environment of a target atomic group, is established where the protons of the target atomic group reside with the protons from other atomic groups which have the coupling relation with the target atomic group. Another matrix of coupling constant values among the protons of the ^{1}H nuclei resonance system is derived. This new matrix can be called a "resonance matrix". A vector of chemical shifts in hertz for the established ^{1}H nuclei resonance system is also derived. In fact the chemical shift vector and the resonance matrix are prerequisites of the quantum chemical calculation method. A theoretical subspectrum is calculated from the vector and the matrix for the protonic environment of the target atomic groups. This subspectrum (in frequency and integration) is then modified by using the routines which are applied in the simplification of the input digital spectrum to make it

Table 3. Assignment Results of the Simulation

	true assignment[7] (ppm)		error in the assignment for given induced error (ppm) in the subspectra calculation (ppm)				
			–0.2 ppm	–0.1 ppm	0.0 ppm	0.1 ppm	0.2 ppm
1	6.75		0.08	0.08	0.08	–0.05	–0.10
2	7.41		0.15	0.06	0.02	0.02	0.02
3	6.62		–0.05	–0.05	–0.05	–0.08	–0.23
4	7.11		0.14	0.03	0.03	–0.05	–0.24
5	4.05		0.01	0.01	0.01	0.01	0.01
1	6.02		–0.01	–0.01	–0.01	–0.01	–0.06
2	7.01		0.01	0.01	0.01	0.01	–0.07
3	2.38		0.11	0.05	0.02	–0.01	–0.03
4	2.00		–0.02	–0.05	–0.12	–0.23	–0.27
5	2.38		0.10	0.05	–0.01	–0.01	–0.03
1	7.50		0.0	0.0	0.0	0.0	–0.03
2	6.89		0.12	0.12	0.07	0.01	0.01
3	6.65		0.0	0.0	–0.06	–0.12	–0.17
4	3.95		0.01	0.01	0.01	0.01	0.01
1	6.54		0.01	0.0	–0.03	–0.08	–0.12
2	6.78		0.16	0.16	0.12	0.03	–0.17
3	7.59		0.12	0.12	0.02	0.02	0.02
4	7.25		0.04	0.04	0.04	–0.22	–0.22
5	6.58		0.04	–0.04	–0.04	–0.04	–0.15
6	9.64		0.02	0.02	0.02	0.02	0.02
1	0.91		0.01	0.01	0.0	–0.08	–0.17
2	1.32		0.33	0.08	–0.10	–0.10	–0.11
3	1.59		0.17	0.16	0.09	0.05	–0.44
4	2.43		0.23	0.23	0.01	0.01	–0.03
5	2.15		0.01	0.01	–0.01	–0.04	–0.05
1	6.52		0.02	0.02	0.02	0.02	0.0
2	7.19		0.02	0.02	0.02	0.02	–0.26
3	7.59		0.14	0.02	0.02	0.02	0.02
4	4.38		0.01	0.01	0.01	0.01	0.0
5	1.39		0.01	0.01	0.01	0.01	0.01
1	7.38		0.04	0.04	–0.05	–0.08	–0.12
2	7.69		0.22	0.19	0.13	0.07	0.01
3	7.34		0.0	0.0	–0.09	–0.12	–0.16
4	7.64		0.21	0.16	0.08	0.02	–0.04
5	8.55		0.02	0.02	0.02	0.02	0.02
6	4.44		0.02	0.02	0.02	0.02	–0.04
7	1.43		0.01	0.01	0.01	0.01	–0.01
1	1.44		0.08	0.03	0.03	0.03	–0.01
2	4.28		0.08	0.08	0.08	0.05	0.0
3	2.82		0.30	0.19	0.0	0.0	0.0
4	4.20		0.02	0.0	0.0	0.0	–0.03
5	2.43		0.18	0.15	0.07	0.01	–0.03
6	5.32		–0.07	–0.07	–0.07	–0.07	–0.07
7	5.52		0.13	0.13	0.13	0.13	0.13
8	2.07		0.02	–0.02	–0.08	–0.15	–0.21
9	0.98		0.01	0.01	0.01	0.01	–0.30
1	1.12		0.06	0.06	0.03	–0.08	–0.10
2	2.13		0.03	0.01	0.0	0.0	–0.03
3	2.13		0.01	0.0	0.0	0.0	–0.03
4	1.37		0.29	0.15	–0.01	–0.07	–0.09
5	1.65		0.24	0.19	0.10	0.01	–0.24
6	0.90		0.01	0.0	–0.05	–0.06	–0.09
7	0.90		0.01	0.0	–0.05	–0.06	–0.09
1	8.21		0.07	0.07	0.07	0.02	–0.04
2	7.64		0.05	0.05	0.05	0.02	–0.08
3	8.11		0.02	0.0	–0.03	–0.03	–0.08
4	8.82		0.02	0.02	0.02	0.01	0.01
5	7.58		0.02	–0.01	–0.01	–0.01	–0.09

consistent. The calculation of a theoretical subspectrum of a substructure is shown in Figure 2.

(B) Algorithm. An atomic group without proton(s) does not produce ^{1}H-NMR subspectra. In such cases the atomic group is allowed to survive.

It is assumed proton(s) bonded to an heteroatom will have no coupling relation with other kind of protons. Hence an atomic group subspectrum of a singlet with $n \times 10000$ intensity can be expected, where n is the number of protons. Moreover

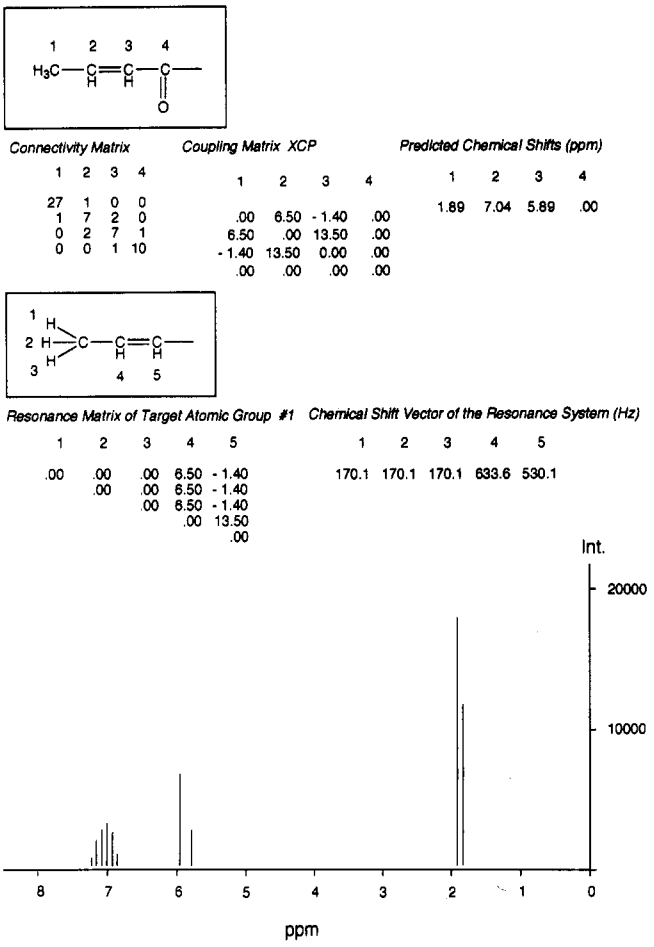


Figure 2. Calculation of the theoretical subspectrum of crotonic acid.

in such cases no subspectrum calculation is essential. Therefore an atomic group subspectrum of a singlet signal and the corresponding parameters (mean chemical shift and integration value) are given for the survival test.

If no nonzero element is found while reading along the row of the corresponding atomic group in the coupling matrix, then the atomic group has no coupling relationship with others. In such cases the atomic group subspectrum will have only one signal and therefore no calculation of the subspectrum is done, but a subspectrum and the corresponding parameters similar to those of an heteroatomic group are deduced directly.

If the so-called resonating ^1H nuclei system has more than seven protons, no subspectrum calculation is done (limitation). However in such cases (a few), the subspectrum and the corresponding parameters similar to those of an atomic group without a coupling relation are deduced for the survival test.

A theoretical subspectrum is calculated for the established resonance system if the system has seven or less proton nuclei. This calculated subspectrum belongs to the environment of the target atomic group.

Thus the calculated subspectrum is then simplified and the signals due to the protons of the target atomic group only (i.e. atomic group subspectrum) is then identified by finding a signal band (group) whose geometric mean chemical shift is closest to the chemical shift of the proton of the target atomic group. All signals except those from the identified signal group are erased. For the survival test of the target atomic group, the parameters of the atomic group subspectrum are then calculated.

(C) Survival Test of the Target Atomic Group. The survival test of the target atomic group is performed by matching its subspectral parameters with that of one signal group from the input spectrum whose geometric mean chemical shift is closest to that of the calculated atomic group subspectrum. The evaluation procedure of the matching is as follows:

step 1	Before matching, the survival value of the atomic group is −100.
step 2	If the geometric mean chemical shift of the atomic group subspectrum is within the range (+0.3 ppm) of those from the input spectrum, then the survival value is 20. If it is not within the range, then go to step 6.
step 3a	If the integration value is less than that of the input spectrum, then the survival value is 40.
step 3b	If the number of signals in the atomic group spectrum is within the range (+1), then the survival value is 50.
step 4a	If the integration value is within the range (+4000), then the survival value is 60.
step 4b	If the number of signals in the atomic group spectrum is within the range (+1), then the survival value is 80; if it is not, then go to step 5.
step 4c	If the mean resolution is within the range (+3 Hz), then the survival value is 100.
step 5	End of survival test.
step 6	Search the signals for the atomic group subspectrum in the simplified input spectrum. Calculate the parameters of the thus-found subspectrum. Match the atomic group subspectral parameters with these new parameters. Evaluate the matching by steps 7 and 8.
step 7	If the geometric mean chemical shift of the atomic group subspectrum is within the range (+0.3 ppm) of those from the newly found subspectrum, then the survival value is 20. If it is not, then go to step 5.
step 8	Go to steps 3a–5.

(D) Survival Test of the Substructure. An average of the survival values of atomic groups for the substructure is calculated, and the existence of the substructure is decided by the following:

(1) If the number of no-matchings of the atomic groups (i.e. the number of the negative survival value) is two or more, then the substructure is denied for its existence in the compound.

(2) If the average is below 20, then also the substructure is denied for its existence.

(3) If the average is below 100 but above 20, then the substructure is not denied for its existence.

(4) If the average is 100, then the substructure exists in the compound.

These survival integrals are arbitrary but have been optimized by several test runs with various types of structures and their corresponding spectra. The result of the survival test for a substructure is shown in Figure 3.

Evaluation of the chemical shift and the coupling constant can be carried out without a problem for the whole structure. However, when substructures bearing protons on carbons with hanging valencies are examined, the evaluation of the chemical shift and the coupling constant becomes ambiguous because the environment of any proton is not precisely defined.

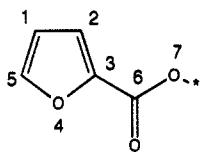
Digital Spectra Input

Spectrometer Type : 90 MHz
Solvent : $CDCl_3$
Molecular Formula : C_7 H_8 O_3

PPM × 100	Int.
759	1994
758	2557
757	2643
756	2639
727	824
720	2258
719	2200
716	2156
715	2489
653	2013
652	2207
649	2233
647	2131
449	1394
441	4922
433	5050
425	1718
146	5213
138	11150
130	4703
0	2566

Report

Atomic Group	Existence Value
1	100
2	80
3	100
4	100
5	100
6	100
7	100



Perhaps the substructure exists.

Figure 3. Output of the program showing the existence of a substructure with the corresponding digital spectra of 2-furancarboxylic acid ethyl ester (input).

However, this is a natural aspect for evaluation of the incomplete structure because there are many possibilities of substituents for a terminal with hanging valencies. The program for survival test of substructures evaluates the possibility of the substructure by an average of the survival values of each atomic group of the substructure. If it is needed to evaluate only the substructure of which the environment is precisely defined, it is easy to ignore the survival value for the atomic node with hanging valencies.

(E) Discussion for the Survival Test Program. The approximations made in the determination of the coupling constant and in the calculation of the theoretical subspectrum, perhaps, could not be denied. The number of nuclei in the resonance system can be increased, but the CPU time will be higher and the computer memory required, for a fast calculation, has to be increased considerably. Therefore the limitations of seven ^{1}H nuclei in the resonance system are not denied. However the program has been tested with various partial and full structures with their corresponding spectra. The program never contradicts the spectra with the corresponding structures (partial or full).

In the test runs it was found that the survival values of atomic groups of a substructure are higher when the number of groupings in the input spectrum is increased by narrowing the vicinity range. But since it is not advisable in the automatic procedure, the narrower groupings can be achieved by the user's interference with displaying the spectrum in a graphic terminal for the grouping. These groupings will not be consistent with those of the calculated subspectrum where the user's interference is not practicable.

Moreover the test performance is more efficient if the input observed digital ^{1}H-NMR spectrum has a little nuclear Overhauser effect and the signals are not concentrated in a few bands or groups. However in either case the survival test is more discriminating and reliable than that of the consistency test or simple matching of mean chemical shifts only.

The program does not attempt to consider stereochemistry. Perhaps the program is not so effective in such isomeric consideration.

The above computer program for the observed ^{1}H-NMR spectra considers various kinds of ^{1}H-NMR parameters. In this sense, it may be expected that the application of this program to automatic structure elucidation processes leads to more substantial results.

III. COMPUTER ASSISTED ASSIGNMENT OF ^{1}H-NMR SPECTRA

Among many structure elucidation strategies, one strategy widely used is the development of an efficient spectroscopic database. Of course, a comparison of the spectrum for an unknown source with those in a simple collection of digital spectra could identify the compound if its spectrum is within the database itself. But, the importance of the strategy is to exploit the spectrum–structure relationship in the structure elucidation of a totally new compound. Therefore in such a spectroscopic database the structural information are stored with the corresponding spectra using any of the standard techniques which can handle the structural information.[3]

For the establishment of the spectrum–structure relationship there are two widely used techniques:

For the establishment of the spectrum–structure relationship there are two widely used techniques: (1) construction of linear models relating structural features;[4] (2) building of another database (index file) of structure and subspectra from the NMR spectroscopic database.[5]

However, either technique requires assigned NMR spectra. But, it is not always possible to get many correctly assigned NMR spectra. The published assigned NMR spectra (presumably correctly assigned) are scattered around many books and journals. There are only a few publications of the collected assigned spectra, which could amount to not more than a few thousand, whereas the number of organic compounds is in seven digit figures. Therefore the assignments of the recorded NMR spectra are essential in the development of an automatic structure elucidation system.

Unlike ^{13}C-NMR spectral assignments, the ^{1}H-NMR spectral assignments are not straightforward and simple. There are two main reasons for it. First the chemical shift range is narrow, and second the spectra of the most interesting compounds are non-first order. Hence the interpretation and the assignment of the ^{1}H-NMR spectrum requires experience and skill. We have developed a computer program which assists in the assignments of a recorded ^{1}H-NMR spectra.

(A) Algorithm. The user inputs temporary assignments of mean ^{1}H-NMR chemical shifts for each atomic group of an organic structure. The program also computes coupling constant values among the atomic groups of the structure which have a proton(s). The computed coupling values are handled by a symmetrical matrix which is also called a coupling matrix. From the predicted chemical shifts (which is the so-called temporary assignment of ^{1}H-NMR chemical shifts by the user) and the coupling matrix, the program computes individual theoretical subspectra of the environment of each atomic group. The subspectra for the atomic groups (signals due to protons of the atomic group only) are then differentiated from the theoretical subspectra. These subspectra for the atomic groups are then searched in the whole ^{1}H-NMR spectrum to be assigned. Thus-found subspectra in the input

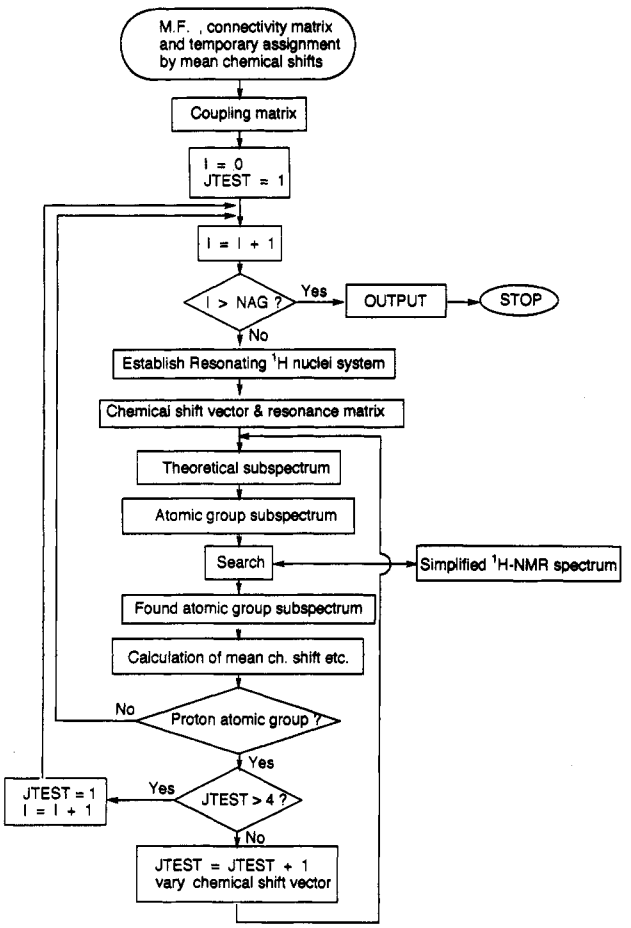
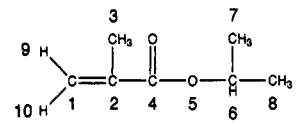


Figure 4. Flow chart of the assignment test algorithm.

spectrum are assigned to those protons. The flow chart of the algorithm is shown in Figure 4. The simplification of the input and the theoretical subspectra are prerequisites of the algorithm. The prerequisites of the program and other consideration are explained in the following sections.

(B) Simplification of the Input Digital ^{1}H-NMR Spectrum. Nowadays the recording of a digital NMR spectra by the spectrometer is common, and the program is suitably developed for the digital spectra. The basic principles of the simplification of a digital ^{1}H-NMR spectrum in this program is similar to those in the program for the survival test of the generated substructure by a ^{1}H-NMR spectrum, but is consists of the following steps only: (1) removal of undesired signals (e.g. TMS, solvent, and smaller noises, (2) standardization of the signals intensities (i.e. one proton equivalent signal intensity), made equal to 10000; (3) lowering of the signal resolution to 4 Hz.

(C) Calculation of a Theoretical Subspectrum. A quantum chemical method is applied to calculate the theoretical subspectrum of an atomic group environment. The program for calculating the subspectrum has been derived from the program LAOCOON3[2] and has been modified for our purposes. The prerequisites and limitation of the program, and the procedure for modifying the calculated subspectrum and for deriving the signals due to the protons of an atomic group are similar to those applied in the survival test of the substructures by the digital ^{1}H-NMR spectra. However, for derivation of a subspectrum due to the protons of an atomic group only from the theoretical subspectrum of the environment of the atomic group, the grouping of the signals of the spectrum is carried out with the vicinity value (0.5 Hz higher than the largest coupling constant value of the established resonance system).



Input NMR Assignment

	Parameters of calculated spectrum					Parameters of observed spectrum					
AN	PR.CS.	CAL.CS.	INTG	NOP	M.RES	AS.CS.	INTG	NOP	M.RES	FROM	TO
3	1.93	1.92	26609	1	0.0	1.93	33844	3	7.2	1.77	1.93
6	5.08	5.08	10001	7	6.3	5.06	6809	5	6.3	4.92	5.20
7	1.28	1.29	28188	2	5.4	1.28	60265	2	5.3	1.24	1.31
8	1.28	1.29	28188	2	5.4	1.28	60265	2	5.3	1.24	1.31
9	5.51	5.50	8757	1	0.0	5.52	8894	1	0.0	5.52	5.52
10	6.11	6.11	11671	1	0.0	6.07	10196	1	0.0	6.07	6.07

Please Check the Result with the Followings

P.A.No.	PR.CS.1	AS.CS.	FROM	TO	PR.CS.2	AS.CS.	FROM	TO
1	5.81		1.24		5.81		1.24	
1	5.71	5.52	5.52	5.52	5.91	6.07	6.07	6.07
1	5.61	5.52	5.52	5.52	6.01	6.07	6.07	6.07
1	5.51	5.52	5.52	5.52	6.11	6.07	6.07	6.07

where,

AN	=	Node number.
PR.CS (ppm)	=	Mean chemical shift given by predictor program.[1]
CAL.CS (ppm)	=	Mean chemical shift calculated on the basis of subspectrum predicted by the above PR.CS.
INTG	=	Total integration.
NOP	=	Number of peaks in calculated or observed subspectra.
M.RES (Hz)	=	Mean resolution among signal peaks.
AS.CS. (ppm)	=	Mean chemical shift of observed subspectrum corresponded to the calculated subspectrum.
P.A.NO	=	Parent atom number.
PR.CS.1 (ppm)	=	Temporary mean chemical shift of geminal proton 1 given by the system.
PR.CS.2 (ppm)	=	Temporary mean chemical shift of geminal proton 2 given by the system.
FROM..TO (ppm)	=	Range of assigned chemical shifts.

Figure 5. Output of the assignment test program.

(D) Conformational and Stereochemical Isomer Consideration. The geminal protons of a methylene group bonded to a chiral center and a vinyl group which has two different substituents are considered magnetically different protons implicitly. The program analyzes the connectivity matrix to find such methylene and vinyl groups. The protons of these groups are then made into individual atomic groups (symbol 99). Hence, the connectivity matrix and the corresponding coupling matrix are then modified accordingly. The program, however, does not make any effort to differentiate between axial and equatorial geminal protons of monocyclic or bicyclic compounds.

(E) Assignment of Unequivalent Geminal Protons. It has been seen from the published assigned ^{1}H-NMR spectra that the maximum difference of chemical shifts between magnetically unequivalent protons is about 0.8 ppm (in the case of a 90-MHz spectrometer). From our experiment it has also been seen that the error of +0.2 ppm in a chemical shift prediction has less effect in the correct assignment. Therefore for a pair of such protons the program calculates a theoretical subspectrum from the predicted mean chemical shift value for a first time. Then, the program corrects the chemical shifts of each proton with an increment of +0.1 ppm and calculates the theoretical subspectra three times. Thus, altogether four subspectra are calculated. From each subspectrum the assignment is done for the pair of protons. Here, the user has to apply her/his judgment in the output to identify the correct assignment. For an example an output is shown in Figure 5 for the assignment of geminal protons of the vinyl group in isopropyl methacrylate. The observed spectrum is shown in Figure 6. The user will have no difficulty in identifying the right chemical shifts (5.52 and 6.07 ppm).

(F) Experiment. Before annexing this program with the ^{1}H-NMR chemical shift predictor program,[1] a simulation experiment was carried out from the published assigned spectra. Instead of the chemical shifts from the predictor program, the correct chemical shift values with error (0.2, 0.1, 0.0, –0.1, and 0.2 ppm) were used in the simulation, and

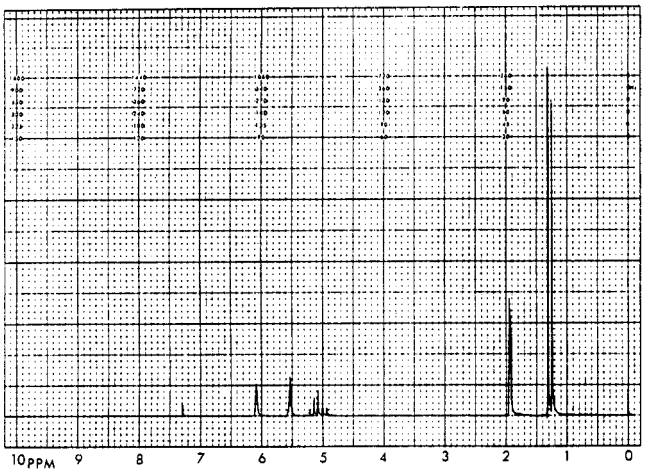


Figure 6. Observed spectrum of isopropyl methacrylate (90 MHz). Reprinted with permission from ref 6. Copyright 1985 Academic Press.

Table 4. Statistical Evaluation of the Assignment

induced error (ppm)	n	X	variance of the error	av error range by t-test
–0.2	58	0.07	0.008	0.03
–0.1	58	0.05	0.005	0.02
0.0	58	0.01	0.003	0.02
0.1	58	–0.02	0.004	0.02
0.2	58	–0.08	0.010	0.04
overall	290	0.01	0.009	0.02

range = $t_\alpha[\sigma_n^2/n]^{1/2}$ $\phi = 0.01$ and $t_\alpha = 2.75$

[a] n = no. of the sample; X = mean of the error.

the results of the assignment for ten compounds were then analyzed statistically. The results of the assignment are shown in Table 3 and its statistical evaluation is shown in Table 4. According to these results, it can be understood that this program gives the correct assignment if the error from the correct chemical shift value is not so large.

(G) Discussion for the Assignment Program. The usefulness of this assignment program can be evaluated from the simulation of experimental results, which has shown that the possible small errors in the predicted chemical shifts have less effect in the actual assignments. The assignment of complex and overlapping signals in the spectra of an organic compound of moderate size is not so easy. Perhaps, assignment of very complex spectra could be doubtful. In such extreme cases one has to be satisfied with the chemical shift range which is implicitly given in each assignment.

APPENDIX A.1. ASSIGNMENT OF COUPLING CONSTANTS

There is a direct relationship between the coupling constant value between two resonating nuclei and the chemical structure to which these nuclei belong. The coupling constant value is affected by the number of bonds between them and their type, the special distance between them, and the α-environment to each nuclei. We can recognize the pattern of the relationship between the coupling constant value and the substructure by any standard pattern recognition technique easily. Unfortunately only a few published coupling data are available to use any standard technique of pattern recognition efficiently.

Table 5. Vicinal Coupling $J_{H-C-C-H}$

classification	$J_{H-C-C-H}$ (Hz)	remark
(a) In CH_3CH_2– (6.9–7.7 Hz)		
general	7.1	
CH_3CH_2C=	7.0	–CH=: –CH=, >CH=, >C=O, –CH=O
$CH_3CH_2C\equiv$	7.4	–C≡: –C≡ or –C≡N
$CH_3CH_2A<$	7.6	any kind of aromatic group
CH_3CH_2S–	7.3	any kind of S-atomic group
CH_3CH_2F	6.9	
CH_3CH_2Cl	7.2	
CH_3CH_2Br	7.3	
(b) In $CH_3CH<$ (5.7–6.7 Hz)		
general	6.4	
	6.7	any kind of aromatic group
$H_3CCHA<$		
	6.2	any kind of O-atomic group
H_3CCHO–		
	6.8	X: halogen
H_3CCHX		
	5.7	
$H_3CCH(CH_2-)(CH_2-)$		
(c) In $-CH_2CH_2-$ (5.2–6.7 Hz)		
general	6.3	
$-CH_2CH_2C<$	6.7	
$-CH_2CH_2CH<$	6.7	
$-CH_2CH_2X$	6.3	X: halogen
$-CH_2CH_2C$=	6.7	–CH=: –CH=, >CH=, >C=O, –CH=O
$-CH_2CH_2C\equiv$	6.7	–C≡: –C≡ or –C≡N
$-CH_2CH_2A<$	6.6	any kind of aromatic group
(d) In $-CH_2CH<$ (5.9–8.2 Hz)		
general	6.7	insufficient data
	8.2	
$-CH_2CHCl$		
(e) In –CH–CH (3.6–6.0 Hz)		
insufficient data, therefore only general classification is possible, value 4.5 Hz		

Table 6. Coupling in Olefine

(a) $J_{HC=CH}$ (4.7–21.0 Hz)

classification	$J_{HC=CH}$ (Hz) cis	trans	mean	remark
general	9.5	17.0	13.3	
CH_2=CH_2	11.6	19.1	15.4	
–CH=CHC=	9.8	17.2	13.5	–C=: –CH=, >CH=, >C=O, –CH=O
–CH=CHO–	6.4	14.1	10.3	–O–: –O–, –OH
–CH=CHF	4.7	12.8	8.8	
–CH=CHX	7.0	13.4	10.2	X: Cl or Br
–CH=CHC≡	11.7	17.9	14.8	–C≡: –C≡ or –C≡N
–CH=CHA<	11.3	15.5	13.4	any kind of aromatic group

(b) J_{HCCH}

classification	J_{CCH} (Hz)	remark
general	11.45	insufficient data
5-membered ring	2.6	for any kind of atomic group
6-membered ring	6.0	for any kind of atomic group
HC(=O)–CH	7.85	for any kind of atomic group

However, from the available published data[7-9] a visual pattern recognition was made. The relationship between the substructure and the proton coupling constant is shown in Tables 5–8.

However, these tables do not contain data of coupling relations between topologically same but chemically or

Table 7.

classification	range (Hz)	coupling (Hz)	remark
	(a) Miscellaneous Vicinal Coupling		
	5.3–7.4	6.5	insufficient data
HC=CH		9.8	
HC≡CH			
	(b) Four Bonds Coupling Series		
HCC=CH	0–1.81	−1.4	relative sign
HCC≡CH		−2.7	relative sign
HC=CCH		−1.2	relative sign
HC≡CCH		−1.2	relative sign

magnetically different protons. Moreover, in a theoretical subspectra calculation, a pair of coupled resonating nuclei with the same chemical shift value provides a singlet resonance signal.

Coupling via heteroatoms, coupling with heteroatoms, and coupling between protons in more than four bonds away (except benzene and its derivatives, five bonds) are not considered. Since isomers due to stereochemistry and conformation analysis are not distinguishable topologically, the coupling between protons from such isomers is not considered as well. In such cases the assigned value may not be correct. However in cycloalkene the coupling value due to the cis isomer is assigned. Though equatorial and axial protons of a cyclic compound are not considered, its ring size is considered in the substructures shown in Table 6b.

APPENDIX A.2. PROGRAM COUPLING

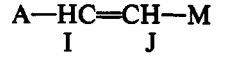
This program has been developed to assign the coupling constant value among atomic groups (with protons) of a substructure from its connectivity matrix by recognizing the substructural pattern shown in Tables 5–8. The assigned coupling constant values are stored in a symmetrical matrix XCP. This matrix is suitable for the theoretical subspectra calculation, and it can be called a "coupling matrix."

A.2.1. Analysis of the Connectivity Matrix To Find Substructural Pattern. The substructural environment of the atomic group is found by analyzing the connectivity matrix (MT). Each row (or column) in the connectivity matrix contains information on the direct connectivity of an atom. Thus reading along the row allows the atoms directly bonded to atom number *n* to be recognized; each nonzero off-diagonal element contains a bond order, while the column index of such an element indicates several numbers of the directly bonded atoms. This is the basis of analysis; at each such encounter, a branch is made to search for the connectivity of the dependent atoms thus found. The process is repeated for any atom found at this depth. This essentially recursive process is carried out to an arbitrary depth (level), depending on the nature of the substructure analyzed in each particular case. Moreover, backtracking to the immediately preceding node is not allowed.

Since the details of the analysis process are different at each depth, the program makes no attempt to use recursive routines. Thus each atomic node (diagonal element) is tested for the groups mentioned in Table 9 for a start point of the analysis to be done by different subroutine(s).

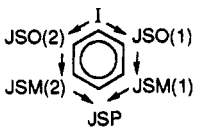
Perhaps it is not necessary to describe the subroutines in detail. However, some complex routines are described in the followed sections for references only.

A.2.1.1. Subroutine CH12CP. This subroutine is basically called for the substructure given as follows:



where A and B are substituents. This subroutine assigns a coupling value not only between protons I and J but also between proton I and a proton(s) from substituent B and between proton J and a proton(s) from substituent A. The coupling between I and J is affected by the substituents in A and B, but the effect is not additive. From the available data, it has been seen that only A or B has the dominating effect in the coupling value. For an example, if A is Cl and B is phenol, then the dominating group is phenol. Therefore a priority order of the substituent has been determined from the examination of available data. Another subroutine OLESUBST assigns an integer value to each substituent according to the priority. The top priority substituent group has the highest integer value. The substituents and their priority integer values are shown in Table 10. The main subroutine CH12CP finds the adjacent substituents A and B which are then evaluated for the order by the subroutine OLESUBST. To have the priority effect, the assignments of coupling constant values are done in the ascending order of the priority integer.

A.2.1.2. Subroutine AROM1CP. This subroutine recognizes a benzene substructure (partial or full) and also distinguishes it from polyaromatic and polyheteroaromatic, pyridine, etc., compounds. Actually this subroutine turns back if –AN–, –ANO–, or –A< with three adjacent aromatic bonds is encountered while traversing along the connectivity matrix. Such substructures are recognized by other subroutines (see Table 9). The traversing is done in both directions from a starting point and along the aromatic bonds only. The accounting of atomic node numbers encountered in the traversing is kept in local dimensions JSO(2), JSM(2), and JSP if the encountered is –AH–. Hence the dimension number will be 1 in one direction and 2 in the other direction, shown as follows. However, these dimension numbers are relative.

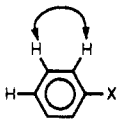
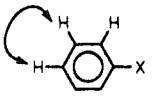
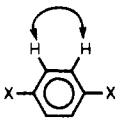
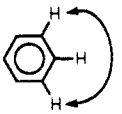
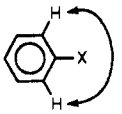
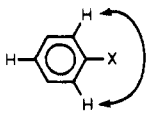
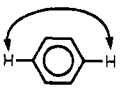
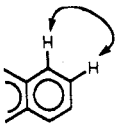
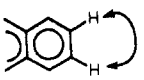
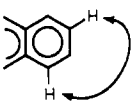


From the survey of the available coupling data it was found that the substituents have sufficient effect in the coupling value between protons of a benzene and its derivatives, but the kind of substituent has little effect in the coupling value. This is better illustrated in the correlation Table 8a. Hence a finding of an –A< atomic node is sufficient to know that there is a substitution in that position.

Since all relative node numbers are derived from starting point I, coupling values among atomic nodes of all possible combinations are assigned except for those nodes which have a substituent. For an example, assignment between two nodes JSO(1) and JSM(2) will be assigned only when both nodes have no substituent. However, the coupling value depends on the substituent in other positions. All possible combinations are defined.

Since computation is done for all possible combinations, it is not necessary to call this subroutine when another –AH– atomic node is found in traversing along the diagonal elements of the connectivity matrix. Therefore the atomic node number is made negative after the assignment so that the same atomic node is not found in the loop of the traverse. Moreover this operation is essential for differentiating the –AH– group of benzene from the –AH– group of a polyaromatic or hetero-

Table 8. Various Compounds Studied and Their Derivatives

classification	coupling (Hz)	remark
	(a) Benzene and Its Derivatives	
	8.0	any substituent
	7.5	any substituent
	8.6	any substituent
	−1.5	relative sign
	−1.8	any substituent, relative sign
	−2.4	any substituent, relative sign
	0.4	any substituent
	(b) Polyaromatic Compounds and Their Derivatives	
	8.2	derivatives include heteropolyaromatic compounds also
	6.9	
	−1.3	relative sign

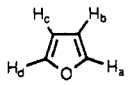
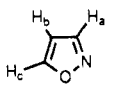
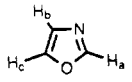
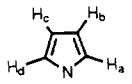
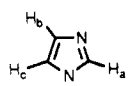
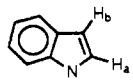
	coupling	range (Hz)	value (Hz)	remark
	(c) Furan and Its Derivatives			
	$J_{a,b}$	1.6–2.2	1.9	
	$J_{b,c}$	3.2–3.7	3.6	
	$J_{a,c}$	0.5–1.2	−0.8	relative sign
	$J_{a,d}$	1.3–1.7	−1.5	relative sign
	$J_{a,b}$		1.78	insufficient data available
	$J_{b,c}$		1.68	
	$J_{a,c}$		−0.27	
	$J_{b,c}$		0.9	insufficient data available
	$J_{a,c}$		−0.09	
	(d) Pyrimidine and Its Derivatives			
	$J_{a,b}$	2.3–3.1	2.6	
	$J_{b,c}$	2.9–4.5	3.5	
	$J_{a,c}$	1.0–2.4	−1.1	relative sign
	$J_{a,d}$	2.0–2.2	−2.1	relative sign
	$J_{b,c}$		1.6	insufficient data available
	$J_{a,b}$		3.1	other coupling similar to benzene derivatives

Table 8 (Continued)

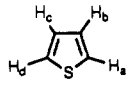
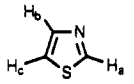
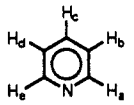
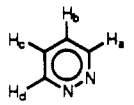
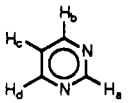
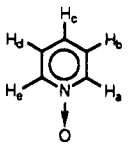
	coupling	range (Hz)	value (Hz)	remark
		(e) Thiophene and Its Derivatives		
	$J_{a,b}$	4.0–5.7	4.8	
	$J_{b,c}$	2.9–4.5	3.9	
	$J_{a,c}$	1.0–2.4	−1.5	relative sign
	$J_{a,d}$	2.5–3.9	−3.2	relative sign
	$J_{b,c}$		3.0	insufficient data available
		(f) Pyridine and Its Derivatives		
	$J_{a,b}$	4.0–5.7	4.8	
	$J_{b,c}$	6.8–9.1	8.0	
	$J_{a,c}$	0.0–2.5	−1.25	relative sign
	$J_{a,d}$	0.0–2.3	1.2	
	$J_{a,e}$	0.0–0.6	−0.3	relative sign
	$J_{b,d}$	0.5–1.8	−1.2	relative sign
	$J_{a,b}$		4.9	insufficient data available
	$J_{b,c}$		5.0	insufficient data available
	$J_{a,b}$		6.47	insufficient data available
	$J_{b,c}$		7.65	
	$J_{a,c}$		−1.88	

Table 9. List of Subroutines for Various Substructural Patterns

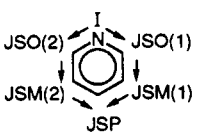
atomic group	subroutine	substructural patterns
CH_3–	CH3CP	CH_3CH_2–, $CH_3CH<$, $CH_3CH=$, $CH_3CH=O$, $CH_3C=CH$–, $CH_3C=CH_2$, and $CH_3C\equiv CH$
–CH_2–	CH2CP	–CH_2CH_2–, –$CH_2CH<$, –$CH_2CH=$, –$CH_2CH=O$, –$CH_2C=CH$–, –$CH_2C=CH_2$, and –$CH_2C\equiv CH$
–CH<	CH1CP	>CHCH<, >CHCH=, >CHCH=O, >CHC=CH–, >CHC=CH_2, and >CHC≡CH
CH_2=	CH22CP	$CH_2=CH_2$, $CH_2=CH$–, and $CH_2=CCH=$
–CH=	CH12CP	–CH=CH–, –CH=CCH=, =CHCH=, and =CHCH=O
HC≡	CH13CP	CH≡CH, CH≡CCH=, and CH≡CCH=O
–AN–	AROMN1CP	pyridine and its derivatives
–ANO–	AROMN1CP	pyridine oxide and its derivatives
–AH–	AROM1CP	benzene and its derivatives
–N< and –NH–	NCYCLIC5CP	pyrimidine fused to an aromatic ring only (this subroutine is called only if the input substructure has a ring size of 5)
–N< and –NH–	HCYCLIC5CP	pyrimidine and its derivatives
–O–	HCYCLIC5CP	furan and its derivatives
–S–	HCYCLIC5CP	thiophene and its derivatives (this subroutine is called only if the input substructure has ring size of 5)
–A<	POLYAROMCP	polyaromatic and polyheteroaromatic derivatives

Table 10. Priority Integers of the Substituents

substituents	priority integer	substituents	priority integer
>CH=, –CH=, >C=O, and –CH=O	1	–Cl and –Br	4
		–C≡ and –C≡N	5
–O– and –OH	2	–A<	6
–F	3		

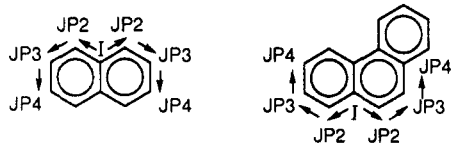
aromatic substructure. This essential procedure prevents the possible mixup of different aromatic substructures. Already assigned nodes with a coupling value by other similar subroutines (AROMN1CP and POLYAROMCP) will be ineffective when this called subroutine becomes effective, and the new coupling value according to this subroutine is assigned. The negative nodes are made negative at the end of this subroutine. The negative nodes are made positive in the MAIN of this program when all assignments are finished.

A.2.1.3. Subroutine AROMN1CP. This subroutine is similar to AROM1CP except that this subroutine turns back when the atomic group –A< with three adjacent aromatic bonds is encountered in the traverse of the connectivity matrix. Naturally the –AN– or –ANO– node will be starting point I, as shown in the following.



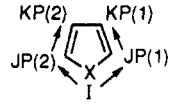
Due to a lack of sufficient data, the substituent effect in the coupling value is ignored. However, this subroutine can be modified and easily made similar to AROM1CP if the substituent effect in the coupling value assignment is required.

A.2.1.4. Subroutine POLYAROMCP. This subroutine is called when the atomic node –A< with three adjacent aromatic bonds is found. Traversing through a node –A< with three adjacent aromatic bonds is not allowed. For example the directions of the required traverse are shown as follows, where JP2, JP3, and JP4 are local variables for the accounting of atomic nodes.



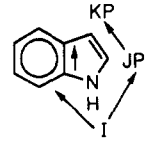
The coupling value depends on the relative bond distance from starting point I, as shown in the correlation Table 8b. Therefore the coupling value is assigned only when the traversing of the connectivity matrix is done up to the JP4 position in both directions. The assignment is done for all combinations of atomic nodes provided each node of one combination has no substituent.

A.2.1.5. Subroutine HCYCLIC5CP. This subroutine is called when an atomic node –N<, –NH–, –O–, or –S– is found in the traverse of the connectivity matrix and the substructure has a smallest ring of size 5. The traverse is done up to 2 bonds from a starting point, which is a heteroatom itself. In a traverse the first bond (α-bond) must be a single bond and the second bond (β-bond) must be a double bond and the substructure should satisfy the ring closure test. The atomic node numbers relative to the starting point are stored in the local dimensions, JP and KP.



The coupling value assignment is done separately for different heterocyclic compounds (furan, thiophene, etc.). The assignment is done in all possible combinations provided the combinations of atomic nodes have no substituent. A heterocyclic compound with two or more heteroatoms are treated individually.

A.2.1.6. Subroutine NCYCLIC5CP. This subroutine is similar to HCYCLIC5CP but is called when an atomic node –N< or –NH– is found with conditions similar to those described in HCYCLIC5CP. Actually this subroutine is called to find pyrimidine and its derivatives fused to an aromatic ring as shown in the following.



In such cases a β-bond could be an aromatic bond. For the assignment of the coupling value of protons from an aromatic ring of such substructures the subroutine AROM1CP is called. In such cases an aromatic ring is taken as a benzene with an ortho-substituent. Moreover, so as not to mix up the assignment of such heterocyclic substructures with those in pyridine and its derivatives, the subroutine HCYCLIC5CP is called immediately after this subroutine. If the conditions required for HCYCLIC5CP are fulfilled, then the new coupling assignment is carried out.

A.2.2. Completion of Coupling Matrix XCP. Since the coupling always occurs between two resonating nuclei, the assignment of the coupling value is done between two atomic groups except for geminal couplings. The coupling value between protons A and B is the same as the coupling value between protons B and A. Therefore the assignment of the coupling value is done in both directions and stored in a two-dimensional array XCP, which is now a symmetrical matrix and also called a coupling matrix. Hence, a traverse of the coupling matrix can tell a coupling relation between two atomic groups of the substructure. In AROM1CP the reverse assignment is done at the end part of the subroutine, and the reverse assignment for the heterocyclic compound is done by calling another subroutine, whereas it is done individually for the remaining substructures.

APPENDIX A.3. OPERATION OF COUPLING

Operation of this program requires an input of the connectivity matrix of the substructure. At the moment, the connectivity matrix is read from a disk file. This program as a subroutine can be called from any host program without any serious modification. The program requires no special numbering scheme. When this program is called from a host program as a subroutine the following information should be given: (1) number of atomic groups; (2) connectivity matrix; (3) set of smallest rings in the substructure. The coupling matrix XCP will be then forwarded to the host program.

REFERENCES AND NOTES

(1) The system for predicting the mean ^{1}H-NMR chemical shift of an organic structure has been developed, which is assisted by retrieval of chemical shift-substructure index files derived automatically from a large database of 8000 assigned ^{1}H-NMR spectra and by statistical analysis of the result of the retrieval. The brief algorithm is described in the following papers: Funatsu, K.; Eguchi, K.; Sasaki, S. Development of a System for ^{13}C-NMR Chemical Shift Prediction with Aid of Index-file and for Ranking of Candidate Structures. In *Proceedings of the 16th Symposium of Chemical Information and Computer Sciences/21th Symposium on Structure-Activity Relationships*; Tsukihara, T., Ed.; Tokushima University: Tokushima, Japan, 1993; pp 81–84. Funatsu, K.; Del Carpio, C. A.; Sasaki, S. Automated Structure Elucidation *System* CHEMICS. *Fr. Z. Anal. Chem.* **1986**, *324*, 750.

(2) Castellano, S.; Bothner-By, A. A. Analysis of NMR Spectra by Least Squares. *J. Chem. Phys.* **1964**, *41*, 3863. Castellano, S. In *Computer Programs for Chemistry*; Deter, D. F., Ed.; Benjamin: New York, 1968; Vol. 1.

(3) Lynch, M. F.; Harrison, J. M.; Town, W. G.; Ash, J. E. *Computer Handling of Chemical Structural Information*; Elsevier: London, 1971.

(4) Small, G. W.; Jurs, P. C. Determination of Topological Similarity of Carbon Atoms in the Simulation of Carbon-13 Nuclear Magnetic Resonance Spectra. *Anal. Chem.* **1984**, *56*, 1314.

(5) Bremser, W. *Chemical Shift Ranges in Correlation Tables for C-13 NMR Spectroscopy*; Verlag Chemie: Weinbeim, 1981. Jankowski, W. C.; Johnson, L. F. *Carbon-13 NMR Spectra*; Wiley-Interscience: New York, 1972. Gray, N. A. B.; Crandell, C. W.; Nourse, J. G.; Smith, D. H.; Dageforde, M. L.; Djerassi, C. Computer-Assisted Structure Interpretation of Carbon-13 Spectral Data. *J. Org. Chem.*, **1981**, *46*, 703.

(6) Sasaki, S. *Handbook of Proton-NMR Spectra and Data*; Asahi Research Center Co. Ltd., Academic Press: Tokyo, Japan, 1985; Vols. I–V.

(7) Chamberlain, N. F. *The Practice of NMR Spectroscopy with spectra-structure correlations for Hydrogen-1*; Plenum *Press*: New York, 1974.

(8) Clerc, P.; Simon, S. *Tabellen zur Strukturauf Klarung Organischer verbindungen mit spektroskopischen methoden*; Springer-Verlag, Berlin, 1981; pp H5–H35.

(9) Simons, W. W.; Zonger, M. *The Sadtler guide to NMR spectra*; Sadtler Research Laboratories, Inc.: Philadelphia, 1972.