nications subsystem is being converted from BTAM to TCAM (*B*asic *T*elecommunications *A*ccess *M*ethod; *T*elecommunications *A*ccess *M*ethod). OLSIS is being extended to handle a wider variety of character fonts and to provide a facility for handling circles and arcs. Sometime in 1978 OLSIS will be used to process the chemical structure diagrams that appear in several of the primary journals published by the American Chemical Society. A program (similar to OLSIS) is under development which will provide the capability to computer-compose the mathematical equations that appear in the primary journals. A capability for handling tables will be added sometime in 1978.

ACKNOWLEDGMENT

CAS is pleased to acknowledge the funding of this work by the National Science Foundation (Contract C656).

LITERATURE CITED

(1) E. J. Corey and W. T. Wipke, "Computer-Assisted Design of Complex Organic Syntheses", *Science*, **166**, 178–192 (1969).
(2) P. A. Castleman et al., "The Implementation of the PROPHET System", *Proc. Nat. Comput. Conf., 1974*, **43**, 457–468 (1974).
(3) R. J. Feldmann and S. R. Heller, "An Application of Interactive Graphics—The Nested Retrieval of Chemical Structures", *J. Chem. Doc.*, **12**, 48 (1972).
(4) B. E. H. Saxberg, D. S. Blom, and B. R. Kowalski, "A Flexible Interactive Graphics System for Searching Atom Connectivity Matrices", *J. Chem. Inf. Comput. Sci.*, **16**, 234–238 (1976).
(5) PDP-15 GRAPHIC-15 Reference Manual, Digital Equipment Corp., 1971.
(6) D. E. Knuth, "The Art of Computer Programming", Vol. 1, "Fundamental Algorithms", Addison-Wesley, Reading, Mass., 1968, p 306.
(7) N. A. Farmer, F. A. Tate, C. E. Watson, and G. A. Wilson, "Extension and Use of the CAS Chemical Registry System", CAS Report No. 2, April 1973, pp 3–10.
(8) P. G. Dittmar, R. E. Stobaugh, and C. E. Watson, "The Chemical Abstracts Service Chemical Registry System. I. General Design", *J. Chem. Inf. Comput. Sci.*, **16**, 111 (1976).
(9) S. J. Tauber and K. Rankin, "Valid Structure Diagrams and Chemical Gibberish", *J. Chem. Doc.*, **12**, 30 (1972).
(10) P. G. Dittmar, J. Mockus, and K. M. Couvreur, "An Algorithmic Computer Graphics Program for Generating Chemical Structure Diagrams", presented at the 169th National Meeting of the American Chemical Society, Philadelphia, Pa., April 1975; *J. Chem. Inf. Comput. Sci.*, **17**, 186 (1977).
(11) N. A. Farmer and J. C. Schehr, "A Computer-Based System for Input, Storage, and Photocomposition of Graphical Data", *Proc. ACM*, 1974, 563–570 (1974).
(12) C. Christensen and E. N. Pinson, "Multifunction Graphics for a Large Computer System", Proceedings of the 1967 FJCC, Vol. 31, AFIPS Press, Montvale, N.J., pp 697–711.

Computer-Assisted Analysis of Infrared Spectra of Nitrogen-Containing Organic Compounds

Y. MIYASHITA, S. OCHIAI, and S. SASAKI*
Miyagi University of Education, Aoba, Sendai, Japan

Received June 15, 1977

Set theory has been applied to the computer-assisted analysis of infrared spectra of the compounds $C_nH_{2n+3}N$, $C_nH_{2n+1}N$, and $C_nH_{2n+3}NO$.

A computer system to enumerate all the possible structural isomers of an unknown compound based on its molecular formula and partial structures provided by the analysis of its spectral data, NMR, IR, and so forth (system CHEMICS), has been reported.[1] However, the system works to elucidate only the structures of organic compounds with C, H, and O. In other words, the computer-assisted analysis of spectra in system CHEMICS has been designed for only compounds with C, H, and O.

To widen the applicability of the system, IR spectral analysis by computer for compounds containing N in addition to C, H, and O has been investigated by using set theory. In the present paper, the program system to analyze IR spectra of $C_nH_{2n+3}N$ (A), $C_nH_{2n+1}N$ (B), and $C_nH_{2n+3}NO$ (C) will be described.

EXPERIMENTAL

A small computer (FACOM 230-15) was used and the program was written in Fortran IV. IR spectra were measured with IR-27G (Shimadzu) and all spectra have been recorded using liquid film for pure liquid. IR spectra were also collected from IRDC cards,[2] Standard Sadtler-Infrared Prism,[3] the Aldrich Library of Infrared Spectra,[4] and some other sources.

SYSTEM

Analysis of IR Spectra. An infrared spectrum includes information on positions and intensities of bands. Characteristic group frequencies have been widely used for qualitative structural analysis of compounds because the approximate constancy of positions of group frequencies has been well established. On the other hand, the intensities of the characteristic bands have been used for the quantitative analysis of functional groups.[5] However, the intensity is rather difficult to express as a constant of the spectrum because the value varies according to the differences in experimental conditions (instrument, slit width, and others), although methods for converting apparent intensities into true molar absorption coefficients and integrated areas have been reported.[6]

To establish the effective analysis of IR spectrum for CHEMICS, we tried to use the intensities of the characteristic bands as well as the characteristic frequencies of a sample compound. The former (intensity of band) is especially useful in determining the presence of a particular partial structure. The molar extinction coefficient (ϵ) is adopted as the information of band intensity. Some of the IR spectral data of A-, B-, and C-type compounds were measured in our laboratory and others were collected from the literature.

Not all the vibrations are good group frequencies because some group frequencies overlap each other in certain compounds. Spectral regions of group frequencies which seem to be effective for the analysis are decided as described below from the IR spectra collected. Nineteen kinds of substructures, named "components" (1–19), previously defined for the analytical work and spectral region(s) corresponding to some of them (6–8, 10–19), are shown in Figure 1 (all of the components will be shown later). Three regions are defined as follows: region 1 (3500–3150 cm^{-1}), region 2 (1690–1630 cm^{-1}), and region 3 (1629–1560 cm^{-1}).

Needless to say, one is always given the apparent peak intensity which never exceeds the true peak intensity in IR spectral charts. It is well known that when the ratio of the spectral slit width of a spectrometer and the half band width ($\Delta\nu_{1/2}$) is less than 1/3, the error of the difference in intensity between the apparent peak and the true peak is not beyond 10%.

Thus $\Delta\nu_{1/2}$ of several typical compounds were measured to examine the influence of the effect of finite slit width. The observed values of $\Delta\nu_{1/2}$ of group frequencies which seem to be effective for the analysis are shown in Table I. Here, $\nu^{a}_{NH_2}$, $\nu^{s}_{NH_2}$, and $\delta^{s}_{NH_2}$ stand for the frequencies of asymmetric, symmetric stretching, and scissors modes of an amino group,

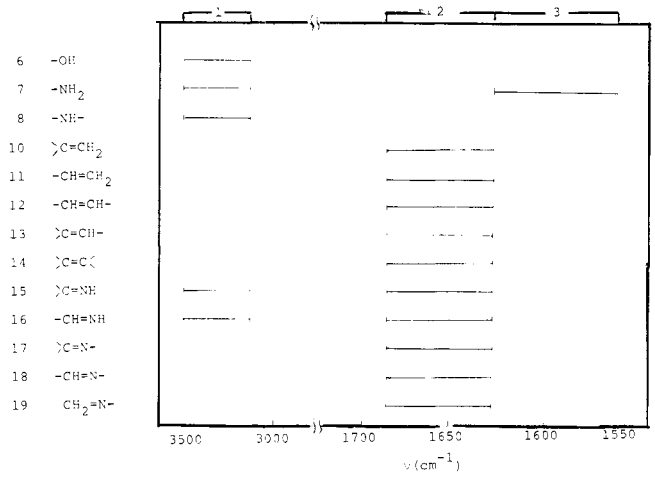


Figure 1. Spectral region for each component.

Table I. Half-Band Width of IR Spectra (cm^{-1})

Mode \ Type	A	B	C
$\nu^{a}_{NH_2}$	100 110	75 90	30 65
$\nu^{s}_{NH_2}$	130 200	95 120	105 133
ν_{NH}	65 100 140	106	150
ν_{OH}	- - -	- - -	396 403 643
$\nu_{C=C}$	-	9	-
$\delta^{s}_{NH_2}$	95 105	60 80	68 80

respectively. Further, ν_{NH}, $\nu_{C=C}$, and ν_{OH} express the frequencies of stretching modes of the NH, C=C, and OH groups, respectively. The values of Table I show that the characteristic bands which are useful and effective for the analysis appear with relatively broad half band widths, excepting the $\nu_{C=C}$ band. Namely, the influence of the effect of slit width is generally not so large in comparison with the resolution of spectrometers used. Anyhow it is never easy to get the truly accurate value of ϵ, so the system in which the analysis is carried out by the use of approximate value of ϵ was studied.

It has been known that the absorption coefficient for each functional group is approximately constant for certain compounds in a homologous series.[7] However, in general the intensity of a functional group in the molecule varies with the rest of the molecule.

Molar extinction coefficients of the above-described groups are useful for obtaining information concerning the nature of O and N, and the index of hydrogen deficiency for A-, B-, and C-type compouds is shown in Table II. Peaks corresponding to ν_{NH_2} and $\delta^{s}_{NH_2}$ are always found while a couple of bands (being worthy) of $\nu^{a}_{NH_2}$ and $\nu^{s}_{NH_2}$ are not always detectable at the same time in primary amines. It is also seen that the band of ν_{NH} is only occasionally detectable in secondary amines because of its weak intensity. The band of $\nu_{C=C}$ is frequently undetectable because the intensity of the band is a function of structural groups attached to the C=C unit. As shown in Table II(c), the bands of ν_{OH} and ν_{NH_2} cannot always be distinguished from each other and the overlap of ν_{OH} and ν_{NH_2} brings on considerably larger values of ϵ. As a threshold value to determine the presence of the above-mentioned partial

Table II. Number of Occurences of Molar Extinction Coefficients for Each Type of Compounds

(a) A-type compounds

ϵ ($M^{-1}\cdot cm^{-1}$)	$n(\nu^{a}_{NH_2})$	$n(\nu^{s}_{NH_2})$	$n(\nu_{NH})$	$n(\delta^{s}_{NH_2})$
0	0	0	2	0
1 - 9	1	0	5	1
10 - 19	8	8	3	4
20 - 29	2	1	0	4
30 - 70	3	4	1	6

(b) B-type compounds

ϵ	$n(\delta^{a}_{NH_2})$	$n(\nu^{s}_{NH_2})$	$n(\nu_{NH})$	$n(\nu_{C=C})$	$n(\delta^{s}_{NH_2})$
0	0	0	1	5	0
1 - 9	1	1	3	3	0
10 - 19	1	0	1	0	0
20 - 29	1	1	0	1	1
30 - 70	0	0	7	1	2

(c) C-type compounds

ϵ	$n(\nu_{OH}+\nu_{NH_2})$	$n(\nu_{OH})$	$n(\nu_{NH_2})$	$n(\delta^{s}_{NH_2})$
0	0	0	0	0
1 - 9	0	0	0	0
10 - 29	0	0	1	1
30 - 99	2	4	3	4
100 - 150	2	0	0	0

structures, $\epsilon = 10\ M^{-1}\ cm^{-1}$ may be safely set up on the basis of the data of Table II.

Judgment. According to set theory, a collection of all the possible structural isomers consistent with the given information, such as a molecular formula and spectral and nonspectral data, is represented by a set of "informational homologues". The universal sets A, B, and C include all the possible structural isomers belonging to A-, B-, and C-type compounds which are consistent with given molecular formulas. The presence of partial structures, for instance, OH (H), NH_2 (I), NH (II), and double bonds C=C and C=N (D), is elucidated by the spectral information of IR. Therefore the subset is restricted to some fixed class of structural isomers, such as compounds which contain a partial structure X_i or do not contain X_i whose sets are denoted by $E(X_i)$ or $\overline{E(X_i)}$, respectively (the set $\overline{E(X_i)}$ denotes the complement of the set $E(X_i)$).

A judgment table was prepared on the basis of the data of Table II to deliver a limitation to the universal set by the IR spectral data (Table III). By using the judgment table with a dead zone, the restricted subset R is expressed as an intersection of the sets for each region R_i.

$$R = \bigcap_i R_i \tag{1}$$

For instance, the following data were supposedly given for B-type compounds:

Region 1 $\epsilon \geq 10$ a single band
Region 2 $\epsilon \geq 10$ a single band
Region 3 $\epsilon = 0$ no band

a judgment is carried out by the use of Table III(b) as follows:

$$R = R_1 \cap R_2 \cap R_3 = (E(\mathrm{I}) \cap E(\mathrm{II})) \cap E(\mathrm{D}) \cap \overline{E(\mathrm{I})}$$
$$= \overline{E(\mathrm{I})} \cap E(\mathrm{II}) \cap E(\mathrm{D}) \tag{2}$$

That is, the subset includes compounds which contain partial

Table III. Judgment Tables for Each Type of Compound

(a) A-type compounds

Region	ε	mode	R
1	ε ≥ 10 (a pair of bands)	ν_{NH_2} ↓	E(I)
	ε ≥ 10 (a single band)	ν_{NH} ↓	E(I) ∪ E(II)
	0 < ε < 10		A
	ε = 0		$\overline{E(I)}$
2	ε ≥ 10		ϕ
	0 < ε < 10		A
	ε = 0		A
3	ε ≥ 10		E(I)
	0 < ε < 10	$\delta^s_{NH_2}$ ↓	A
	ε = 0		$\overline{E(I)}$

(b) B-type compounds

Region	ε	mode	R
1	ε ≥ 10 (a pair of bands)	ν_{NH_2} ↓	E(I)
	ε ≥ 10 (a single band)	ν_{NH} ↓	E(I) ∪ E(II)
	0 < ε < 10		B
	ε = 0		$\overline{E(I)}$
2	ε ≥ 10		E(D)
	0 < ε < 10	$\nu_{C=C}$ ↓	B
	ε = 0	$\nu_{C=N}$ ↓	B
3	ε ≥ 10		E(I)
	0 < ε < 10	$\delta^s_{NH_2}$ ↓	B
	ε = 0		$\overline{E(I)}$

(c) C-type compounds

Region	ε	mode	R
1	ε ≥ 10	ν_{OH} ↓ ν_{NH_2} ↓ ν_{NH} ↓	E(I) ∪ E(II) ∪ E(H)
	0 < ε < 10		C
	ε = 0		$\overline{E(I)} \cap \overline{E(H)}$
2	ε ≥ 10		ϕ
	0 < ε < 10		C
	ε = 0		C
3	ε ≥ 10		E(I)
	0 < ε < 10	$\delta^s_{NH_2}$ ↓	C
	ε = 0		$\overline{E(I)}$

structures D and II and do not contain I. If erroneous IR spectral data are given for A-type compounds:

Region 1 $\epsilon \geq 10$ a pair of bands
Region 2 $\epsilon = 0$ no band
Region 3 $\epsilon = 0$ no band

it then follows from Table III(a)

$$R = E(\mathrm{I}) \cap A \cap \overline{E(\mathrm{I})} = \phi$$

The restricted subset becomes the null set ϕ, which means no informational homologues are included. Such structural information afforded by the judgment is utilized to construct the molecular structure in the next step of the system.

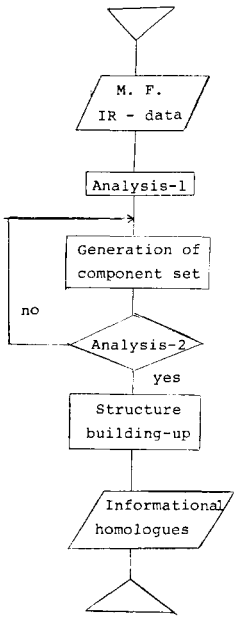


Figure 2. Flowchart of structure construction.

Structure Construction. Structural information obtained by the method described above is sent to the routine of structure construction as shown in Figure 2. As the subsets afforded by each region are represented in two types of forms

$$\bigcup_i \overline{E(\mathrm{X}_i)} \text{ and } \bigcup_j E(\mathrm{X}_j)$$

the following is given for eq 1 (general form)

$$R = AN_1 \cap AN_2$$

where

$$AN_1 = \bigcap_i \overline{E(\mathrm{X}_i)},\ AN_2 = \bigcap_j \left(\bigcup_{k_j} E(\mathrm{X}_{k_j})\right) \tag{3}$$

The information of the subset AN_1 and AN_2 are used at the stages of analysis 1 and analysis 2, respectively. All of the components (1–19) are listed in Table IV, with which one can set up all component sets consistent with a molecular formula belonging to A-, B-, or C-type compounds. By feeding a molecular formula and a result (AN_1) judged by the mode described above for the IR spectrum of an unknown sample, a maximum and a minimum of each component are computed. That is, every number of maximum components including partial structures X_i given by the intersection of complements AN_1 is equal to zero.

For example, the number of maximum and minimum components is decided for compounds containing n carbon atoms through the result obtained by eq 2 as shown in Table IV; then a component set consistent with molecular formula is generated. Next, whether the component set is appropriate is examined using the resulting AN_2 at the stage of analysis

Table IV. Component List for Generating Component Sets

No.	1	2	3	4	5	6	7	8	9	10
component	CH_3-	$-CH_2-$	-CH<	>C<	-O-	-OH	$-NH_2$	-NH-	>N-	CH_2=C<
Max	n	n	n	n	0	0	0	1	1	1
Min	0	0	0	0	0	0	0	0	0	0

No.	11	12	13	14	15	16	17	18	19
component	CH_2=CH-	-CH=CH-	-CH=C<	>C=C<	>C=NH	-CH=NH	>C=N-	-CH=N-	CH_2=N-
Max	1	1	1	1	1	1	1	1	1
Min	0	0	0	0	0	0	0	0	0

Table V. New Component List for Structure Build-up

No.	component	No.	component
1	CH_3-	13	-CH=
2	$-CH_2-$	14	=C<
3	-CH-	15	>C=
4	-C-	16	>C=NH
5	-O-	17	-CH=NH
6	-OH	18	-N=
7	$-NH_2$	19	=C<
8	-NH-	20	-CH=
9	-N-	21	=N-
10	CH_2=C<	22	CH_2=N-
11	CH_2=CH-	23	< = >
12	-CH=CH-		

2. Let Z_m (m = 1, 2, . . ., 19) be the number of components of the component set and suppose that AN_2 contains only one union $\bigcup_{k_1} E(X_{k_1})$; then the following inequality is examined:

$$\sum_{k_1} \sum_{m \in X_{k_1}} Z_m \geq 1 \tag{4}$$

We denote that a component m includes a partial structure X_{k_1} by writing

$$m \in X_{k_1}$$

It is clear that eq 4 says that the component set should contain at least one component corresponding to an absorption peak. Therefore, in general, the following inequality is examined according to eq 3

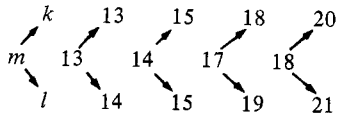
$$\prod_j \sum_{k_j} \sum_{m \in X_{k_j}} Z_m \geq 1 \tag{5}$$

For example, eq 2 gives the following inequality according to eq 5

$$(Z_8 + Z_{15} + Z_{16}) \cdot \sum_{m=10}^{19} z_m \geq 1 \tag{6}$$

When eq 6 is satisfied, the component set is sent to the routine of structure build-up.

To build up the full structure from components in the component set, it is necessary to convert components having nonequivalent bonds, e.g., no. 13, 14, 17, and 18, into new components by elimination of the double bond of C=C or C=N. In other words, when component "m", used in the component set, has nonequivalent bonds, they should be separated into "k" and "l" to be used for structure build-up. Thus, components 13, 14, 17, and 18 must be modified as follows:



New components are listed as shown in Table V. Component 23 in Table V plays the role of connecting two double bond-removed components. With the component sets modified in this manner, all possible structures corresponding to structural information are constructed and enumerated without duplication.

RESULTS AND DISCUSSION

As a link in the study of CHEMICS, enumeration of all the informational homologues consistent with IR spectral data of N-containing compounds has been investigated. Since there are finite cases of IR spectra given by the intensities of the bands in the three regions in Figure 1, as input for the judgment, we are able to show all cases of IR data analysis. Actually, 5, 10, and 6 restricted sets appear for A-, B-, and C-type compounds, respectively, as shown in Table VI.

Table VI. Number of Informational Homologues Corresponding to the Result of Judgment for Each Type of Compound

(a) A-type compounds

No. \ n	R	6	7	8	9	10
1	E(II)	15	33	82	194	482
2	E(I)	17	39	89	211	507
3	$\overline{E(I)}$	22	50	122	296	731
4	E(I) ∪ E(II)	32	72	171	405	989
5	A	39	89	211	507	1238

(b) B-type compounds

No. \ n	R	6	7	8	9	10
1	$\overline{E(I)} \cap E(II) \cap E(D)$	53	145	392	1068	2903
2	$E(I) \cap E(D)$	56	149	398	1068	2876
3	$\overline{E(I)} \cap E(D)$	93	249	670	1808	4893
4	E(I)	100	284	801	2258	6355
5	$E(D) \cap (E(I) \cup E(II))$	109	294	790	2136	5779
6	E(II)	111	312	883	2487	7017
7	E(D)	149	398	1068	2876	7769
8	$\overline{E(I)}$	184	517	1457	4097	11529
9	E(I) ∪ E(II)	211	596	1684	4745	13372
10	B	284	801	2258	6355	17884

(c) C-type compounds

No. \ n	R	6	7	8	9	10
1	$\overline{E(I)} \cap \overline{E(H)}$	102	286	795	2202	6080
2	E(I)	166	437	1157	3087	8276
3	$\overline{E(I)} \cap (E(II) \cup E(H))$	202	543	1468	3969	10770
4	$\overline{E(I)}$	232	631	1719	4682	12773
5	E(I) ∪ E(II) ∪ E(H)	368	980	2625	7056	19046
6	C	398	1068	2876	7769	21049

It is interesting and worthwhile to compare the numbers of informational homologues computed by us (Table VI) with those obtained by Henze and Blair,[8,9] who calculated the number of structural isomers of acyclic compounds by using a recursion formula. It is known that the number of primary, secondary, and tertiary amines of n carbon atoms is equal to that of primary, secondary, and tertiary alcohols of $n + 1$ content, respectively, and the number of primary amines of n carbon atoms ($A_p(n)$) is equal to the total number of amines of all types containing $n - 1$ carbon atoms ($A(n - 1)$). According to the former relation, the number of amines should agree with that of alcohols. The latter relation is represented by the following recursion formula

$$A_p(n) = A(n - 1) \tag{7}$$

It is observed that the values in Table VI(a) and (b) satisfy eq 7: in Table VI(b) the values for the amines containing a double bond also satisfy eq 7.

Further, the number of primary amines of C-type compounds of n carbon atoms ($C_p(n)$) is equal to the sum total of isomers of C-type compounds containing $n - 1$ carbon atoms ($C(n - 1)$) and that of primary amines of A-type compounds of n carbon atoms ($A_p(n)$); that is,

$$C_p(n) = C(n - 1) + A_p(n) \tag{8}$$

The recursion formula 8 is satisfied in Table IV(a) and (c). Therefore, the above-mentioned fact supports the validity of our method for the enumeration of informational homologues.

When the number of carbon atoms varies from 6 through 10, the number of informational homologues in the first case of Table VI(a) amounts to 15, 33, 82, 194, and 482, re-

Table VII. Partitioning of the Set for C-Type Compounds ($n = 6$)

O \ N	E(I)	E(II)	E(III)	Total
E(H)	80	80	50	210
E(O)	86	72	30	188
Total	166	152	80	398

spectively. The numerals at the bottom of Table VI(a–c) express all the possible isomers consistent with only information of a molecular formula. It is seen that the number of informational homologues rapidly decreases (sometimes in half) in accordance with the addition of IR spectral information. Therefore, the analytical method in which both information of molecular absorption coefficient (ϵ) and that of band position are used functions very effectively for the automated structure elucidation. Finally the number of elements of each subset set up for analysis of C-type compounds containing 6 carbon atoms is shown in Table VII. The disjoint subsets refer to two different atoms, O and N, which partition the universal set C: E(O) and E(III) express the sets of ethers and tertiary amines. This table makes the correlation between the restricted sets and their number of informational homologues (Table VI(c)) clear-cut and shows that the analytical method takes advantage of information for the presence of a subset.

NMR, one of the strongest weapons for treating compounds with C, H, and O in CHEMICS, will be combined with IR to cover the analysis of not only the compounds with C, H, and O but also with C, H, O, and N in the near future.

REFERENCES AND NOTES

(1) S. Sasaki, H. Abe, Y. Kudo, and T. Yamasaki, "CHEMICS: A Computer Program System for Structure Elucidation of Organic Compounds", 1977 ACS Symposium Series on Computer-Assisted Structure Elucidation of Organic Compounds, in press; Y. Kudo and S. Sasaki, "Principle for Exhaustive Enumeration of Unique Structures Consistent with Structural Information", *J. Chem. Inf. Comput. Sci.*, **16**, 43–49 (1976).
(2) IRDC, Nankodo, Japan.
(3) "Standard Spectra-Infrared Prism", Sadtler Research Laboratories.
(4) C. J. Pouchert, "The Aldrich Library of Infrared Spectra", Aldrich Chemical Co., Milwaukee, Wis., 1972.
(5) R. N. Jones and C. Sandorfy, "Technique of Organic Chemistry", Vol. 9, A. Weissberger, Ed., Interscience, New York, N.Y., 1956, p 247.
(6) K. S. Seshadri and R. N. Jones, "The Shapes and Intensities of Infrared Absorption Bands", *Spectrochim. Acta*, **19**, 1013–1085 (1963).
(7) S. H. Hastings, A. T. Watson, R. B. Williams, and A. Anderson, Jr., "Determination of Hydrocarbon Functional Groups by Infrared Spectroscopy", *Anal. Chem.*, **24**, 612–618 (1952).
(8) H. R. Henze and C. M. Blair, "The Number of Structurally Isomeric Alcohols of the Methanol Series", *J. Am. Chem. Soc.*, **53**, 3042–3046 (1931).
(9) H. R. Henze and C. M. Blair, "The Number of Structural Isomers of the More Important Types of Aliphatic Compounds", *J. Am. Chem. Soc.*, **56**, 157 (1934).

Computer-Assisted Chemical Research Design. Second Joint Japan–United States Seminar

HARRY B. MARK, JR.*

Department of Chemistry, University of Cincinnati, Cincinnati, Ohio 45221

SHIZUO FUJIWARA

Department of Chemistry, University of Tokyo, Tokyo, Japan

Received September 12, 1977

The first joint Japan–United States seminar on Computer-Assisted Chemical Research Design was held in Honolulu, July 1973. The second joint seminar, held in Washington, D. C., August 1976, was concerned with how individual aspects of computer applications fit together into the design of total systems. Papers in the second seminar are reviewed and discussed; three papers from the seminar follow this review.

In the past 15 years, computer technology and design, plus the development of first the low-cost "minicomputers" and recently, the "microcomputer" and other integrated circuitry, have advanced at an almost unbelievable rate, and the future promises fantastic new hardware and software. Thus, computer systems and computer-based instrumentation are now readily available and are revolutionizing research in chemistry. The applications of computers in the areas of computation, information storage and retrieval, data acquisition and reduction, simulation, and instrumentation (on-line data processing and experimental control or optimization, or both, in real time) have, of course, opened up a new dimension in research design.

Although chemists have been working with these new techniques and tools, and also have been developing new methods and instrumentation over the past few years, we have found, in general, that on an individual basis, we have only limited or narrow awareness of the new developments in the area of computer uses in chemistry and allied fields. Thus, the need for greater communication led to the organization of the first joint Japan–United States seminar on this topic. This was held in Honolulu in July 1973. The object of this seminar was to bring together representatives working in many different aspects of computer uses so that we could find out what new developments could be used in our own specific areas of research, and to consider areas of future development and needs. It was also felt, at that time, that it would be desirable to meet again in the future to discuss the results of the applications of these techniques to our research and to examine how further experience in their use pointed out new needs, etc. Thus, the second joint Japan–United States seminar was held in Washington, D.C., in August 1976.

An overview of the various applications of computers in chemical research and the basic aims of the seminar were given by S. Fujiwara (University of Tokyo). In the first seminar, the participants, for the most part, discussed only individual aspects of computer applications. It was decided that the major aim of this second seminar would be to consider how to fit these various segments together into the design of *total systems*. This would not only include complex hierarchial systems for laboratory instrumentation and data reduction, but also hierarchial information retrieval systems to merge