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

large public and small local databases which would be accessible through the laboratory system itself. Some of the problems would be the development of the hardware and software needed to put together all the segments of such hierarchial systems, and the need for better "intelligent" terminals for the user. The papers in the seminar were divided into five broad categories: (i) laboratory hierarchical systems, (ii) real-time applications of minicomputers and microprocessors in experimental and instrumentation, (iii) application of pattern recognition techniques, (iv) uses of the computer in data processing, and (v) information retrieval.

With respect to the design of a hierarchical laboratory system, C. N. Reilley (University of North Carolina) discussed the up-to-date design of a computer system which serves approximately 30 varied instruments, types of experiments and users in the Chemistry Department at the University of North Carolina. The paper stressed the importance of making a detailed specification of the tasks to be performed and constraints imposed by the nature and types of experiments to be carried out before designing the computer system for real-time data acquisition and reduction and experimental control (automation). Another important point brought out was the necessity of taking into account in the design both the financial resources and technical expertise that would be available in the laboratory for support of the system. Another main point was the necessity to design a system with sufficient flexibility as needed to revamp the system as new developments in hardware and/or new experimental measurements arose. S. Fujiwara presented a discussion of the design of specific hierarchial systems used in fast Fourier transform nuclear magnetic resonance experiments. The importance of including ready access to a spectral data base in intepreting the experiment was stressed. J. W. Frazer (Lawrence Livermore Laboratories, University of California) also discussed philosophy of design of complex laboratory systems at the Lawrence Radiation Laboratories and other installations. Y. Fujiwara (University of Tsukuba) showed the use of "Data Structure", a concept which could be employed as a component of an integrated laboratory system. This was used for molecular design, and the routine would look up various known physical and chemical properties and predicted resulting spectral properties.

Discussing real-time applications of minicomputers and microprocessors in instrumentations for data acquisition and experimental control were C. N. Reilley, J. W. Frazer, T. R. Mueller (Oak Ridge National Laboratory), and R. E. Dessy (Virginia Polytechnic Institute and State University). Of particular interest was the fact that these speakers suggested that the future role of the microprocessor and the philosophy of its use in instrumentation would in each case be different. These papers described the varied use of the microprocessor in the construction of a complete low-cost "microcomputer" system of modular design that could readily be adapted to any instrumentation or experiment and act as a link in a central hierarchial system, in fixed operational mode to a specific instrument or experiment, and finally a small (almost pocket calculator size) self-contained unit that was programmable for different applications. Although it seemed that there was considerable disagreement in the future use of microprocessors, it turned out that these widely differing designs resulted mainly from the specific application needs and resources of the different laboratories. A comparative survey of the various microprocessors now available and anticipated in the near future was also presented in these papers. A large number of specific experimental instrument designs employing either microcomputers or microprocessor based systems were described. D. E. Smith (Northwestern University) discussed the construction of fast Fourier transform electrochemical instrumentation to determine heterogeneous electron transfer rate constants and mechanisms. K. Konishi (Kao Soap Co., Ltd., Wakayama City) described the design and application of automatic titration unit for the more accurate and precise analysis of industrial oils and derivatives. J. W. Frazer illustrated the tremendous potential of interactive on-line graphic techniques which were applied to an automated real-time potentiometric titrator using ion-selective electrodes. S. Ikeda (Osaka University) showed the application to ESCA and photoemission spectroscopy (PES) instrumentation. H. B. Mark, Jr. (University of Cincinnati), described the design of a rapid scanning UV–visible spectrophotometry, and R. E. Dessy demonstrated the applicability of computerization for high-performance liquid chromatography (HPLC) detectors.

The papers and discussions in the application of pattern recognition to problems in chemical analysis showed that tremendous progress had been made since the first seminar. S. P. Perone (Purdue University) described an operator-interactive routine featuring real-time feedback to generate the most appropriate feature set for a given classification problem. This routine was applied to the classification of electrochemical kinetic data. S. P. Perone also applied pattern recognition to the analysis of the composition of environmental particulates in the atomosphere. He illustrated how this could be used to identify the various sources of atmospheric pollutants. P. Jurs (Pennsylvania State University) used pattern recognition routines in structure activity studies. The ultimate goal of this work is the design of compounds of prescribed activity (in the case of drugs, for example). C. L. Wilkins (University of Nebraska) described the application of GC–mass spectral and infrared spectral analysis to correlate spectra with molecular structure. He compared several pattern recognition techniques. J. S. Mattson (National Oceanic and Atmospheric Administration) showed that pattern recognition techniques were very valuable in the analysis and determination of the origin of oil spills on the ocean.

In the area of data reduction problems, H. B. Mark, Jr., and R. E. Dessy described two different approaches for the display and analysis of time-resolved spectra obtained with the use of a rapid-scanning spectrophotometer as a detector for HPLC units. C. N. Reilley discussed a variety of programs used for the molecular structure predictions from C-13 NMR, a "Simplex" fitting routine for predicting and simulating response curves for different reaction mechanisms, shift reagent studies, ESCA spectra filtering and spectral convolution and deconvolution. I. Watanabe (Osaka University) presented a data treatment routine for PES spectra and showed how the computer was able to circumvent inherent stability problems in this instrument. A computer prediction technique of molecular structures from the analysis of NMR and IR spectral data was developed by S. Sasaki (Miyagi Kyoiku University). T. Sakai (Department of Information Science, Kyoto University) described a waveform and pictorial data-processing system for a computer network, and J. W. Frazer pointed out the advantages of the experimenter operating in an "interactive" mode during data reduction operations.

Discussing information storage-retrieval, H. Skolnik (Research Center, Hercules, Inc., Wilmington, Del.) pointed out and illustrated how it is concerned with obtaining and processing information and designing its input format so it may be evaluated, related, correlated, and extrapolated in a variety of output formats. He especially emphasized that keyword indices based on titles, or even on author abstracts, are poor substitutes for intellectually produced subject indexes which are conceptually oriented. T. Yamamoto (Computer Center, University of Tokyo) and H. Ishizuka (National Institute of Japanese Literature, Tokyo) described the evolution, structure,

example uses, and future course of development of the University of Tokyo Information Retrieval System. It was interesting to learn that the computer centers at the Japanese universities, in contrast to those in the United States, are principally devoted to the support of research and are largely staffed by scientists. S. Fujiwara described potential sources of errors in information retrieval systems. S. Heller (Environmental Protection Agency) and G. W. A. Milne (National Institutes of Health) described the NIH–EPA World-wide Chemical Information System which consists of 18 large databases. Y. Yoneda (Faculty of Engineering, University of Tokyo) described an information retrieval system for the retrieval and correlation of physical properties of organic molecules and the estimation of physical properties of other molecules from similar molecules, component functional group, and the like in the database.

The syllabus, approaches, and educational problems of a short course on computer-based instrumentation were discussed by R. E. Dessey.

Again, all the participants expressed the need for expert technical assistance in computer hardware and software in the chemical laboratory which is common in industrial and governmental laboratories. Several examples were given to illustrate how chemistry departments profit by the addition of full-time computer scientists. It was felt that this need will become more apparent in the future as it is anticipated that computer centers in universities will decentralize as size and costs of the large computers shrink. However, it was again apparent that in most university laboratories, the administration is not enthusiastic about creating such appointments in chemistry, and that the interaction and communication with personnel in other departments and especially computer service groups was difficult.

The seminar was sponsored by the National Science Foundation and the Japan Society for the Promotion of Science under the Joint United States–Japan Cooperative Science Program, and was organized by H. B. Mark, Jr., and S. Fujiwara.

The Role of Chemical Information Science in Computer-Assisted Chemical Research†

HERMAN SKOLNIK

Hercules Incorporated, Research Center, Wilmington, Delaware 19899

Received June 16, 1977

The essential role of chemical information science to the design of chemical research via computer is illustrated with subject indexes, notation systems, correlation methods, and on-line databases.

Chemical information science has become an integral component of computer-assisted chemical research planning and design.[8-16] Other components, in addition to mathematics and statistics, include computer science and the application of techniques such as artificial intelligence and pattern recognition.[1-6,18] Whereas computer science is concerned primarily with algorithms, i.e., the design of sequences to be processed in a computer to yield output information from input information, information science is concerned with the obtaining and processing of information and designing its input format so that it may be evaluated, related, correlated, and extrapolated in a variety of output formats.[16]

The above description of information science emphasizes the importance of data and information and, to a degree, deemphasizes the role of the computer. It is certainly true that the major contributions of information science by chemists have been in areas which are peripheral to what computer scientists do. Some of these contributions, which will be discussed in this paper, are indexing systems, notation systems, the correlation of data and information, and the design of on-line databases.

I do not mean to imply by the above that computer science is not important to the work of chemical information scientists. A knowledge of computer science is highly advantageous to the understanding of many disciplines of science in which computers are utilized. Interaction with a computer, such as in the construction, manipulation, or analysis of algorithms,[4] forces one to think through a problem, very much like a teacher does with students. Thus, when linguists attempted automatic translation via computer, they discovered that the process of translating from one language to another was considerably more complex than they had realized. When chemists attempted programmed organic synthesis, in which they were reasonably successful, they could not approach the achievement of the human intellect and intuition. They have gained, however, considerably more insight into how chemists conceive synthetic paths.

INDEXING

Indexing and classification systems are the basic concepts in information science for making data and information available and for arranging data and information in usable and useful formats. Classification and taxonomic concepts, such as those of Mendeleev, Dumas, and Linnaeus, have been of great value in the understanding and teaching of science and in the reduction of many facts to a matrix of knowledge.[16] Work in the area of artificial intelligence could benefit by paying more attention to classification concepts.

Of the variety of indexes in chemistry, such as subject, author, patent number, formula, and ring, the most difficult to achieve is the subject. Because a subject index is based on words, it is fraught with semantic, linguistic, and logic difficulties. To the user of an index, a word does not always mean what an author or indexer thinks it means. An index based on isolated words communicates the least amount of information and is characterized by the maximum amount of noise; it is like a dictionary without definitions, i.e., a simple word list.

Yet the isolated word index, called the uniterm index, was the one that designers of computerized information systems

† Hercules Research Center Contribution No. 1693. Presented in the Joint U.S.–Japan Seminar on "Computer Assisted Chemical Research Design", Aug 16–20, 1976, Washington, D.C.; sponsored by the National Science Foundation and the Japan Society for the Promotion of Science.