ASTM Data Banks and Chemical Information Sources

ALBERT L. BATIK* and ELEANOR HALE
American Society for Testing and Materials
1916 Race St., Philadelphia, Pa. 19103

Received May 18, 1972

A comprehensive presentation of the data banks, computer tapes, punched cards and disks, microfilm, and books available from ASTM are discussed. Contents, utilization, and applications are described. Plans for the immediate future are outlined. Among the data described are infrared indexes, mass spectral data, chromatographic data, X-ray emission data, odor and taste threshold data, and thermodynamics data. In addition, information on recommended practices, monographs, and standards are presented. This paper provides the chemical documentarian a complete reference source to a wide variety of analytical data.

Although the primary product of ASTM is the world's largest collection of voluntary, consensus, standard methods of test, definitions of terms, specifications of materials, and so on, ASTM maintains four large important data banks that supplement its standardization projects. In addition ASTM has numerous other chemical information sources.

The data banks are: infrared spectroscopy, gas chromatography, mass spectrometry, and X-ray emission spectrography. Presently, the largest of these data banks is the Infrared Spectral Index, designated AMD 33A, which indexes the significant absorption bands of some 102,000 organic and inorganic compounds. This collection is truly an index to the world's published collections of infrared spectra. Every major collection in the world is indexed: Infrared Data Committee of Japan, Sadtler Catalogue, Documentation of Molecular Spectroscopy, Coblentz Society, American Petroleum Institute Research Project 44, Manufacturing Chemists Association, Aldrich Chemical Company, and abstracts from related literature. Twenty-five thousand additional spectra have been processed and were recently added to the file.

The ASTM infrared spectral index provides the data necessary for the user to create a retrieval program for his particular computer hardware. If a card file is desired, an object deck of the tape-to-card program is included.

The magnetic tape is available at 556 or 800 bits per inch, BCD, with a blocking factor of 10. The tape can be used on an IBM 1401 computer; and just recently, the tape was made available on 9 track, EBCD tape for the IBM 360 computer.

There are several types of efficient references to this master file of infrared spectra. One of these references is the "Infrared Name Formula Index" or AMD 36 on magnetic tape with its updated supplement AMD 36-S14 also on magnetic tape. Using the same 102,000 spectra of the master file these two tapes together provide the molecular formula, compound name, and the publisher's serial number of the spectra. These data are in either molecular formula sequence or serial number sequence whichever is preferred.

Another type of reference to the master name formula file is a series of three books with their updated supplements. One of these reference books "The Molecular Formula List" or AMD 31 and its supplement AMD 31-S14, offer a printed index of the 102,000 spectra of the master name formula file in molecular formula order. It also includes the compound names and the serial numbers.

The "Infrared Serial Number List" or AMD 32 and its updated supplement AMD 32-S14 are reference books to the master name formula file. These two books are listed in serial number order; they also include the compound names.

The third type of reference book to the master name formula index is the "Alphabetical List," designated AMD 34. Its updated supplement AMD 34-S14 is now available. These books contain the compound names of 102,000 spectra in alphabetical order; the molecular formulas and the serial numbers are also included.

In all three of these reference books and their respective supplements, the infrared and far-infrared spectra abstracted from the general literature and books are listed by ASTM assigned serial numbers with journal references. A list of CODEN assignments is also included for easy title conversion. Work is now on-going for a 15th supplement.

A vital addition to the master spectral index are ASTM's infrared file searching systems SIRCH III and SIRCH 360. These infrared retrieval programs have been designed to make the interaction between the spectroscopist and the computer as convenient as possible. The file has been generated from the punched data cards of ASTM's entire master spectral index file, but only those data necessary to the identification of an unknown spectrum have been included. These are the actual absorption bands, elemental and selected group data, and the ASTM serial number and file designation which refer to the original source.

The SIRCH program can be used with the IBM 360 and the 1130 computer systems. Moreover, the systems will operate on either single or multi-disk machines. This versatility makes the infrared retrieval systems available to a wider range of companies and universities. Both the SIRCH 1130 and the SIRCH 360 systems offer the same data with only slight modifications in the actual output.

In general, the way the SIRCH programs operate is this: data from the unidentified spectrum (which would be the

*To whom correspondence should be addressed.

groups and bands present, groups absent, and the no-band regions) are entered via keypunched cards or directly through the typewriter keyboard. These data may be in the form of wave number or wavelength or both, and in random order. The computer retypes these data into the masks.

After each search, data can be selectively added to, or deleted from the search "masks." No re-entry of data is necessary to search additional data files. The spectroscopist may continually modify the data until he has identified his spectrum. The output consists of the group data and assigned serial numbers indicating the source of the published spectra.

The SIRCH procedure has many features which add to the efficiency of the spectroscopist. One principal feature is speed: the data file, and that means 102,000 spectra, are searched at a rate of 1000 spectra per second. Another feature is selectivity: far-infrared standards (15.1 to 34.0 microns) are searched independently of those coded in the 5.5 to 15.0-micron range. Another feature is adaptability: the users' own data files may be added to and searched in addition to the ASTM data. Furthermore the spectroscopist does not need to learn complex codes, nor worry about keeping entry data in a single format or order. If a mistake is made, the computer alerts him.

In addition to the straightforward search capability for organic qualitative analysis, these data can be further manipulated for valuable educational or theoretical studies. Such possibilities as semiquantitative studies of the effects of electron-donor groups, electron-withdrawing groups, steric hindrance, or a study of active infrared sites are feasible because of the large reservoir of data available. Organic structural manipulation for undergraduate study could include such factors as position unsaturation in chains and rings, presence of unsaturation, substitutions in rings and chains, cis-trans isomerism, directions of rotation, and many other structural features. The student could also construct variations of compounds from structural information and search the data bank for confirmation. These many possibilities for the ASTM infrared data bank would be of immediate value to courses in structural theory, spectroscopy, and chemical literature.

Another of ASTM's data banks is one on gas chromatography. The "Gas Chromatographic Data Compilation" is in book form designated AMD 25A and AMD 25AS1 and provides the chromatographer with the conditions used for the separation of a wide variety of compounds facilitating the tentative identification of materials producing peaks on chromatograms. These books contain tabulations of gas chromatographic data printed from punched cards in a format which provides convenient means for rapid retrieval of data. Most of the data were obtained from the literature, although some are unpublished data contributed by the laboratories of the Eastman Kodak Co., and the Lockheed-California Co., These books provide data on the retention indices, column temperature, reference material, liquid phase, and solid support for 3800 compounds in the first volume and 4000 in the second. The compounds have been arranged in molecular formula sequence and assigned serial numbers accordingly. All data have been subject to computer analysis to ensure high quality.

The supplement to the "Gas Chromatographic Data Compilation" adds nearly 4000 new compounds to those listed in the first edition of the compilation doubling the data. This book includes for the first time in alphabetical order, liquid phase/active solid codes, solid support codes, and reference material codes. This supplement is expected to be used closely with the first edition which is the original compilation. All data which have been corrected or changed from the first edition are relisted in this supplement. The changes in the system have been made to supply more information to the user and reduce the number of codes assigned to liquid phases, solid supports, and active solids. A second supplement is now in preparation and will add an additional 10,000 compounds to the data bank.

In addition to the book form, the "Gas Chromatographic Data Compilation" is available on magnetic tape and microfilm. The 9-track magnetic tape contains all of the data in the first two books and provides the user with an even faster, more efficient reference to the retention data. The microfilm also contains all of the data in the first two books, but it can be purchased only in combination with the purchase of the Supplement. The same arrangement will probably be followed with the second supplement.

A third section of ASTM's Data Series provides Mass Spectral information. ASTM's data publication on the Index of Mass Spectral Data or AMD 11 is a comprehensive index of the approximately 8000 mass spectra contained currently in the informal spectra program of ASTM Committee E-14 on Mass Spectrometry. The spectral data consist of the six strongest peaks in each spectrum and their relative intensities, along with a coded number and name of the compound. There are seven different indices. The first is based on molecular weight, the second is based on the most intense peak, and the rest are based on the second through the sixth peak.

The index was printed directly from computer printout pages prepared from the Göteborg University Mass Spectral Data Collection compiled at the Institute of Medical Chemistry under the direction of Einar Stenhagen.

ASTM has also produced the Mass Spectral Index in punched card form to be used for searching mass spectral data. These cards contain the six strongest peaks, molecular weight, serial number, source code and chemical structures code so that it is possible to create a computer program to search the mass spectral data. These punched cards are available in three different forms; there is a set grouped according to Data-Structure, a set of Data-Name cards, and a set of Name-Formula cards. All of these sets of punched cards have updated supplements to them.

Work is now going forward on expanding the mass spectral data file. However, it is expected that several years will elapse prior to a major addition to the file published by ASTM.

Another of the ASTM Data Series is the section on X-ray emission. X-ray emission spectrography (X-ray fluorescence) and the electron microprobe are two basic techniques for elemental chemical analysis. Fundamental to the successful use of these techniques is the proper application of X-ray emission line and absorption edge wavelength data, together with calculated values for the Bragg diffraction angles of the analyzing crystals. These developments have created a need for reference tables in convenient form and ASTM has met this need with an edition titled "X-Ray Emission and Absorption Wavelengths and Two Theta Tables" or DS 37A that provides a listing of all X-ray emission lines (160 angstroms and shorter), incorporating some 3400 first order lines, absorption edges, and the calculated two-theta values for 23 commonly used analyzing crystals. Since new analyzing crystals are being developed continually, ASTM has published a Second Edition to the original 1965 edition.

Adequate information for the projected Second Edition was not available. Therefore, it was necessary for ASTM to underwrite a research project that was carried out by

G. G. Johnson, Jr., and E. W. White at Pennsylvania State University. This experimental work was necessary to refine existing data, develop data for the seven additional analyzing crystals coming into common use, add pertinent satellite lines, and increase the range of the wavelength tables to 250 angstroms to accomodate new instruments reaching the market. Most important, this experimental work was conducted in close cooperation with the precise wavelength standards of A. J. Bearden at the Johns Hopkins University.

For the Second Edition, a complete change in format was designed to provide easy access to the information for the user. The X-ray line designation and column headings are very readable, thanks to the use of a special IBM print train that carried most of the Greek letters as well as large and small type, upper and lower case, and prime characters. As a result, the table has the appearance of being printed from set type. The wavelength table has been compiled with the use of high resolution spectrometers in mind, and because of this the edition integrates the absorption edges in a manner that makes them easily retrieved.

In the first section of the book, data are ordered primarily on the basis of increasing atomic number and present all wavelength lines and X-ray absorption edges shorter than 160 angstroms. However, there are drastic changes such as spectra complexity with increasing atomic numbers, and they have been taken into account here. Relative intensities have been determined for the M-series lines. No attempt has been made to assign N-series lines intensities as these are seldom used analytical lines, and we were unable to measure N spectral series experimentally.

Data in the second portion of this book give all the lines shown in the first section and are ordered on the basis of increasing wavelength. This arrangement makes possible the phasing in and out of crystals during just their useful lambda range. This allows compilation of two theta value for the total of 23 analyzing crystals.

As an extra feature, there is a separate periodic chart of the elements which gives the wavelengths of related X-ray emission lines of the elements. This chart is suitable for hanging or taking to the spectroscope.

Recently developed, high-resolution X-ray detectors have resulted in widespread application of nondiffractive analysis also called nondispersive analysis, as an alternative or complementary technique to X-ray emission spectrography or X-ray fluorescence analysis. The same basic technique is used, but signal detection is different. Because of this ASTM has published a table for use in nondiffractive analyses, the basic data for which was underwritten by ASTM. All of the X-ray emission lines shorter than 50 angstroms have been tabulated in two basic arrangements. The first section of the table is a compilation of each line for each element arranged on the basis of increasing atomic number and increasing wavelength (decreasing energy) regardless of element.

All lines shorter than 50 angstroms that were used in the preparation of the second edition of the DS 37 mentioned earlier have also been included in this table. Some of the very weak lines may not be observed in nondiffractive spectra, but most of the lines should be resolved in favorable cases. This material is found in "X-Ray Emission Wavelengths and KEV Tables for Nondiffractive Analysis," DS 46.

Of particular interest to those in the X-ray analysis field is a comprehensive series of papers published in book form by ASTM. The book is entitled "Energy Dispersion X-ray analysis: X-ray and Electron Probe Analysis" (STP 485) and includes thirteen papers covering such topics as X-ray spectrometry, detectors, electron probes, and emission sources to form an extremely comprehensive volume. A supplement to this book is entitled "The Theoretical Characteristics of X-Ray Spectra" (STP 485X) and contains a single sheet of K, L, M, spectra.

The foregoing is one of a series of Special Technical Publications (STP) that ASTM has published that is of particular usefulness to chemical information specialists. A few of the others are: "Laboratory Handling and Storage of Peroxy Compounds" (STP 471); "Effect of Automotive Emission Requirements on Gasoline Characteristics" (STP 487); "The Manual on Hydrocarbon Analysis" (STP 332A); "Atomic Absorption Spectroscopy" (STP 443); "Microorganic Matter in Water" (STP 448) and the "Manual on Water" (STP 442).

In addition, several data publications are available that are extremely useful. As examples, "Physical Constants of Hydrocarbons C_1 to C_{10}" (DS 4a) and "Available Standard Samples, Reference Samples, and High-Purity Materials for Spectrochemical Analysis" (DS 2a).

The "Annual Book of ASTM Standards" itself is the source of much useful information and standard methods of tests of particular interest to chemists. As examples, there are Part 18 on Petroleum Testing; Part 21 on Paint Testing; Part 23 on Water and Air Analysis; Parts 26 and 27 on Plastics Testing; Part 30 which includes Emission, Molecular and Mass Spectral Methods, Chromatographic and Resinographic Methods; and Part 32 on Emission Spectrochemical Analysis, Chemical Analysis of Metals and Metal Bearing Ores.

Subsidiary to those are two specially prepared manuals, "Methods for Emission Spectrochemical Analysis" (E 2) and "Manual on Recommended Practices in Spectrophotometry" (E 13).

Some very interesting projects are currently underway which will develop future banks of data and methods of tests. As examples, we have just recently published "Compilation of Odor and Taste Threshold Values Data." Data or threshold values are widely scattered in the literature. Much is not published. In this volume, the following data are assembled: name of compounds, ACS-IUPAC nomenclature, synonyms, physical data, purity, media, temperature, threshold volume, methodology, and references. This is a continuing project and will be followed up with supplementary data.

Another project with experimental work done at Rutgers University deals with Energy Hazard Evaluation of Chemical Thermodynamics, code named Project CHETAH. When the project is completed there will be a package consisting of several thousand punched cards and an operating manual. By use of the manipulation of the punched cards, it will be possible to evaluate maximum heats of reaction and energy hazards. It is hoped that this data bank will become available this year.

Project Threshold is an on-going project which should be completed within two years. Its purpose is to determine the repeatability, reproductibility and accuracy of 35 ASTM standard methods for sampling and analysis of ambient air and emission from stationary and other sources. Upon completion this assembled data bank will form the foundation of a *validated* series of standard methods for air analysis, the first of its kind in the world.

These data banks, books, microfilms, computer tapes, and disks that have been discussed have been designed to provide useful information in the most efficient way for each scientist and engineer's individual needs. It is hoped that this brief review will provide sufficient information to estimate its value. All of this material is described in the ASTM List of Publications available free from the Society.