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

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

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

The application of other 2D NMR techniques to CHEMICS is now under investigation in a similar manner to that of the 2D-INADEQUATE technique. The procedure reported here is just the beginning of applying such a series of 2D NMR information to the CHEMICS system.

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

ACKNOWLEDGMENT

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

REFERENCES AND NOTES

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

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

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

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

M. A. PLEVA

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

Received June 2, 1988

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

INTRODUCTION

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

Transformation of the sample into a form that can be analyzed by a specific measurement technique often requires tedious, time-consuming procedures such as dissolution, dilution, and separation. These procedures in turn depend upon four analytical descriptors: (1) the type of matrix represented by the sample; (2) the analyte(s); (3) the expected concentration level of the analyte; and (4) the instrumental technique to be used for the measurement. Once these four descriptors

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

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

† Certain commercial equipment and software are identified in this paper to specify adequately the experimental procedure. Such identification does not imply recommendation or endorsement by the National Bureau of Standards, nor does it imply that the equipment or software is necessarily the best available for the purpose.

‡ Professor of Chemistry, Virginia Military Institute, Lexington, VA 24450, visiting scientist in the Center for Analytical Chemistry 1987–1988.

0095-2338/89/1629-0011$01.50/0

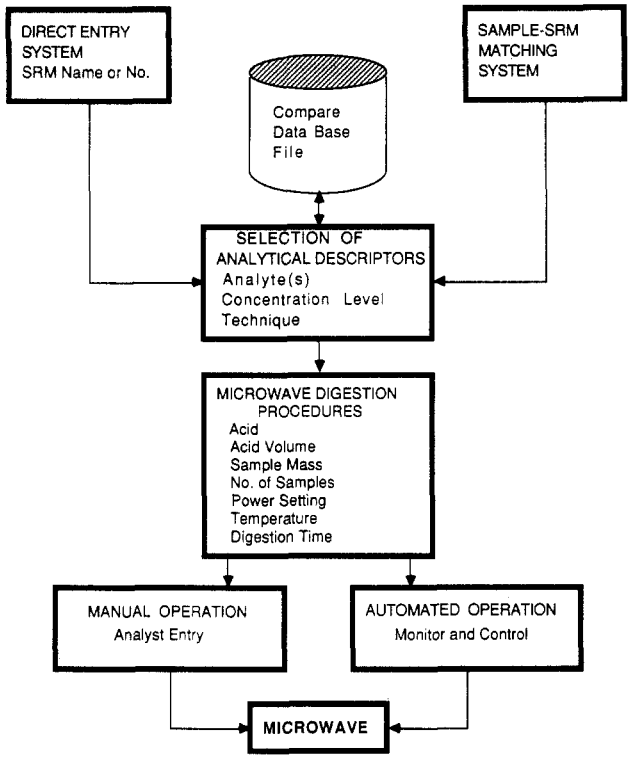


Figure 1. Expert system for sample preparation.

have been established for a given analysis, the procedures necessary for preparing the sample can be determined.

For dissolution, the first step in the preparation of many samples, closed vessel microwave acid digestion, offers advantages over conventional dissolution methods for a variety of sample types[2] and hence will be the first procedure treated by our expert system. This technique can reduce digestion times from hours to minutes for many samples and assures reproducible results by close control of digestion conditions. To function properly, a microwave procedure requires values for the sample mass, the identity and volume of the acid, and the temperature required for dissolution. These values are dependent on the analytical descriptors described above, with the matrix type being of primary importance (Table I). When these values are specified for a given dissolution, the power settings (in watts) for the microwave apparatus and times required to produce the temperatures for the dissolution can be calculated.

The proposed expert-database system will (1) assist the analyst in establishing the analytical descriptors, (2) provide information required for closed vessel microwave digestion of the sample, and, when necessary, (3) provide advice on methods for removal of interfacing substances. The modules associated with these tasks are being developed sequentially. This paper describes the initial module that guides the analyst in the assignment of values to the analytical descriptors (matrix, analyte, concentration level, and measurement technique). By use of these values and experience in the form of heuristic knowledge associated with microwave digestion, the second module will provide the analyst with information on microwave dissolution procedures where appropriate (Figure 1). While the initial version of the system furnishes information directly to the analyst, future versions will allow the system to interact directly with laboratory equipment, including robots, to produce automated dissolution procedures.

For example, the determination of iron at trace levels in wheat flour by ICP requires dissolution of the sample in concentrated nitric acid. With a knowledge of the analytical descriptors, the microwave dissolution module of the expert-database system can determine the dissolution procedure for

Table II. Conditions for the Microwave Dissolution of a Wheat Flour Matrix Predicted by Using the Analytical Descriptors

dissolution variable	reason for choice
acid, concd HNO_3	known to dissolve wheat flour SRM
sample mass, 0.25 g	trace amounts of iron expected and moisture content <5%
vol of acid per sample, 5 mL	based on sample mass
no. of samples, 6	determined by total mass of acid, target temp, time to target temp, initial power setting, and capacity of unit
total mass of acid (all samples), 43 g	only the acid absorbs microwave energy
digestion vessel, 120-mL Teflon–PFA	determines max temp
target temp, 170 °C	determined by matrix, acid, element, and vessel
time to target temp, 180 s	determined by energy vs mass equation
initial power setting, 67%	determined from identity of acid, total mass of acid, target temp, and power output of unit

the sample (Table II). In this example, iron can be determined directly in the resulting solution. In samples containing components known to interfere with the determination of a given analyte, the final module in the system will provide information on appropriate methods for separating or masking these interferences.

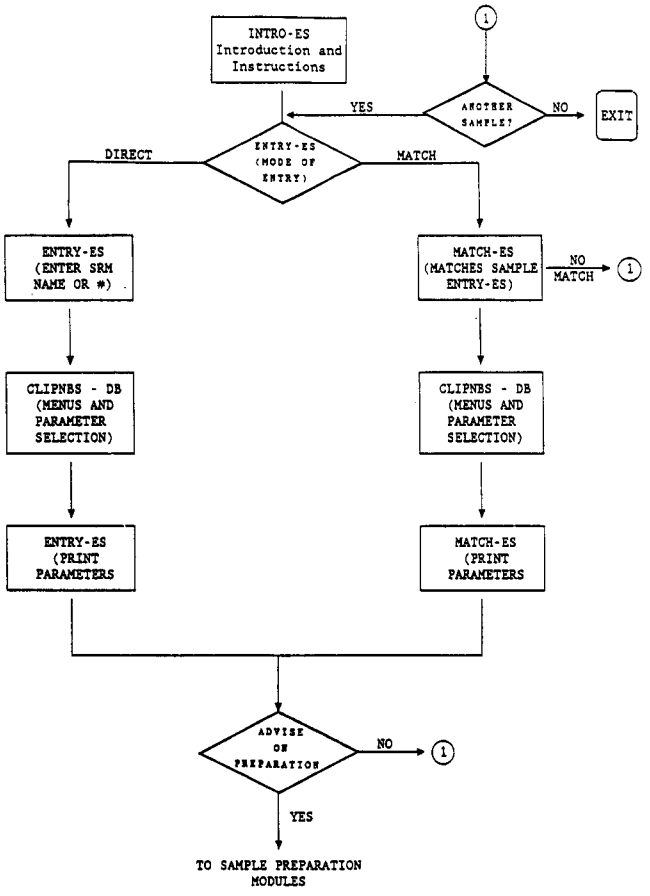
Expertise resident in the Inorganic Analytical Research Division of the Center for Analytical Chemistry at the National Bureau of Standards (NBS) forms the basis for both the rules of the initial module of the expert system and the contents of the database. This expertise has been developed over many years as a result of the analyses required for the certification of Standard Reference Materials (SRMs).[3] Many components of these materials have been determined with high precision and accuracy. Since the concentration of each element certified in an SRM is usually verified by using two independent measurement techniques, a large amount of data has been accumulated at NBS and has been stored in a database.

Because the SRMs represent a broad spectrum of common substances, ranging from environmental materials to alloys, they can form the basis for classifying the matrix of almost any material. For example, the bovine liver SRM, a biological matrix, contains proteins, carbohydrates, lipids, and fats in a ratio similar to that of many other types of mammalian tissue. The concentrations of alkali, alkaline earth, and transition elements, as well as those of common anions in bovine liver, are comparable to those found in other mammalian tissue samples. This classification of samples by comparison of the matrix and elemental composition to those of SRMs forms the basis for the initial module of the system. The better the match between the sample and an SRM, the more effective the procedures for sample preparation and analysis.[4]

THE EXPERT-DATABASE SYSTEM

The knowledge base of the expert system (ES) component of the analytical descriptors module contains rules for matching the matrix of the user's sample with that of an SRM. The database (DB) component of the module contains detailed information on the elemental composition of the SRMs. Thus, in a general sense, both the rules of the knowledge base and the contents of the DB represent knowledge available for reasoning by the inference engine of the ES.[5]

Initially the user interacts with the expert system to select the best match between the matrix of the user's sample and the matrix of an SRM (Figure 2). The ES assists the user



ES = Expert System Component
DB = Database Component

Figure 2. Analytical descriptors module of the ES–DB system.

in the selection process much as an experienced analyst provides assistance to a novice. This guidance takes the form of a series of menus generated by the rules of the ES. The user describes the sample by making selections from each menu. The user's responses, and the rules of the knowledge base, determine the sequence of menu presentation. When it acquires sufficient information from the user, the ES provides the name of the SRM most closely resembling the sample described by the user. The ES now utilizes the DB component to assist the user in locating information on elemental composition and analytical techniques. Using the database management (DBM) program, the user can select the set of analytical descriptors (analyte, concentration level, and measurement technique) for the analysis under consideration. These descriptors will also be used by the dissolution and separation modules of the system.

The Expert System Component. The expert system development tool selected for this project was PC-Plus (version 3.0) from Texas Instruments,[6] a rule-based system similar to EMYCIN.[7] EMYCIN originated with MYCIN, an expert system written in LISP at Stanford University in the mid-1970s for the diagnosis and treatment of blood diseases. Removal of the domain knowledge associated with blood diseases resulted in EMYCIN, the empty MYCIN shell. This shell has been used to develop diagnostic systems for use in engineering, agriculture, and geology.[8] Since the developement of procedures for sample preparation can be thought of in terms of a diagnosis—determining the four analytical descriptors—and a treatment—recommending procedures for dissolution and separation—the choice of PC-Plus is appropriate.

```
Rule:  35

       IF   :: ENTER = MATCH AND CAT-GEN = BIOLOGICAL-MATERIAL AND CAT-BIO =
               AGRICULTURAL AND CAT-AG = WHEAT-FLOUR
     THEN   :: MATRIX = "wheat flour" AND MATCH1
```

Figure 3. Typical rule.

The environment for developing the prototype system and the goals of the project dictated the use of (1) a microcomputer to produce the system and (2) a compiled version of the system for transferring the resulting technology to laboratories with PC-based microcomputers. The hybrid system was developed on a PC-AT computer with 2-MB extended RAM and a 40-MB hard disk. An IBM-XT-compatible system with a minimum of 640 kB of memory is required to run the compiled version of the prototype system. The current knowledge base of the analytical descriptors ES module contains approximately 100 rules.

The Database Component. The DB selected for the hybrid system was dBASE III PLUS from Ashton-Tate. The data for the prototype system was available in a dBASE file used by the Inorganic Analytical Research Division at NBS to compile data from various techniques used for the certification of SRMs. The use of these data saved time in the development of the system and illustrates an important advantage of hybrid expert-database systems—the sharing of a common database by independent systems.

The database management programs for the ES–DB system were first written in dBASE III PLUS source code and then compiled by using Clipper, a dBASE III PLUS compiler from Nantucket, Inc.[9] The compiled version offers advantages of increased speed, the security of object code, and independence from dBASE III.[10] The DB component of the system consists of dBASE files containing the data and the object code of the DBM program.

The DBF file used for the prototype system consists of 2300 records occupying 320 kB of memory. Each record contains results for the determination of a single element, in a given SRM, by a specific technique. Each record consists of 10 fields. The file, begun in 1982, contains data on only 50 SRMs and is therefore incomplete. As new analyses are carried out in the Division, the DBF file will be expanded.

Expert System–Database Interactions. Expert systems can obtain the data required to drive their inferencing mechanisms in three ways: (1) from the user in response to system prompts; (2) from real time sensors monitoring events and processes; and (3) from data residing in online secondary storage, i.e., databases. In the third method, of particular interest in the context of this paper, the DBM facilities of the DB, as well as the data, become available to the ES.

When an ES is combined with a DB, the location of information becomes an important issue.[11,12] What information belongs in the rules of the knowledge base, and what information belongs in the database? If fields must be added to the DB to accommodate special situations, then the addition of rules to the knowledge base of the ES should be contemplated. Conversely, if many rules are necessary to handle information occurring in a structured, repetitive format, the information should be placed in a DB. In general, DB systems are best at mindless sorts and searches of structured data, e.g., elemental analysis of SRMs, while the rules of ESs work best with the less structured information from the experts' experiences with the knowledge domain, e.g., microwave acid dissolution.

DESCRIPTION OF SYSTEM OPERATION

A set of rules illustrated by rule 35 (Figure 3) matches the sample with an SRM. These rules are composed of IF conditions and THEN action statements that efficiently prune unfruitful branches from the decision tree. The rules (1)

```
GENERAL
  BIOLOGICAL
  ENVIRONMENTAL
  FERROUS ALLOYS
  NONFERROUS-ALLOY
  GASES-IN-METALS
  PURE-METALS
  PURE-CHEMICALS
  FERTILIZER
  ORE
  ROCK
  NONE
```

```
BIOLOGICAL
 FOOD-&-BEVERAGE
 AGRICULTURAL
 BIO-FLUID
 CLINICAL
 NONE
```

```
AGRICULTURAL
 CITRUS-LEAVES
 TOMATO-LEAVES
 PINE-NEEDLES
 CORN-STALK
 CORN-KERNALS
 WHEAT-FLOUR
 RICE-FLOUR
 POWDERED-MILK
 WATER
 NONE
```

Figure 4. Menus used to match a sample with an SRM.

```
Rule:  141

    IF  :: MATCH1
  THEN  :: PRINT :ATTR '(REVERSE CYAN) "THE RESULTS OF THE SAMPLE-SRM MATCH"
           LINE 2  :ATTR '(CYAN) "The SRM whose matrix most closely
           resembles that of your sample is"  :ATTR '(WHITE) MATRIX
           :HOLD/CLEAR  :LINE 10  :TAB 15  :ATTR RED) "TRAVELING TO AND FROM
           THE DATABASE" AND MATCH2


Rule:  118

    IF  :: MATCH2
  THEN  :: STAT = "matx" AND EXPORT (QUOTE (DOS-FILE-OUT (STAT MATRIX)"
           linkfile.txt" CSV)) AND DOS-CALL "clipnbs.exe" "linkfile.txt"
           AND MATCH3
```

Figure 5. Rules used to inform user of a match and to access the database.

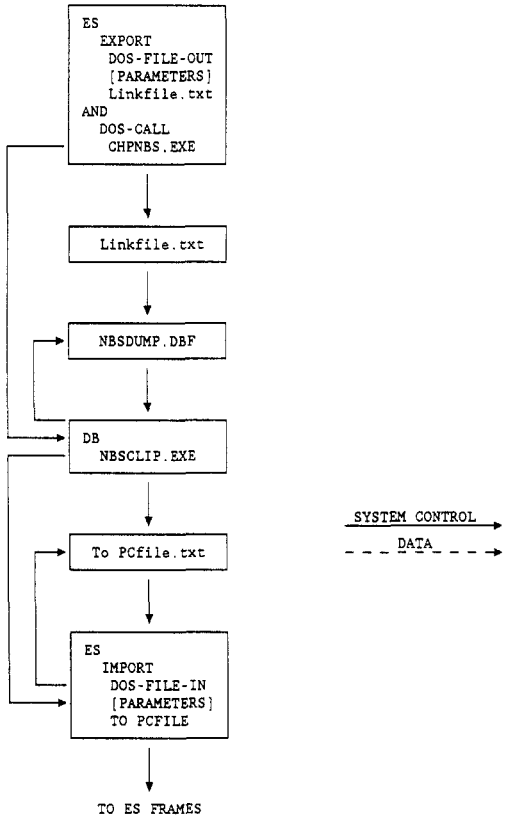


Figure 6. Transfer of data between PC-Plus and dBASE III PLUS.

express relationships between individual pieces of knowledge, known as parameters, and (2) produce conclusions based on these relationships. Each parameter can have one or more values. In the current knowledge base, the IF portion of the rules contains values for the parameters used to classify the sample and the THEN portion produces the name of the SRM. When a rule from this set is tested, the series of menus shown in Figure 4 is generated by the parameters of the rule. For example, if the user selects the value BIOLOGICAL-MATERIAL for the CAT-GEN parameter in response to the first menu, the second menu appears. If AGRICULTURAL is the value selected for the CAT-BIO parameter, the final menu of Figure 4 is displayed. If the user decides that the sample is best represented by wheat flour, then the value

```
          Your sample's matrix has been matched to that of
          wheat flour

This is the MAIN MENU.  Your options are:

     0 -- Quit dBase to expert system when you either change your matrix or
          continue the overall consultation

     1 -- Focus on element - This choice will present those elements which
          have been determined in wheat flour.  You may then select an element
          and receive information about the technique(s) of analyses used and
          the concentration levels found.

     2 -- Focus on technique - This choice will present those analytical
          techniques used in analyzing wheat flour.  You may then select a
          technique and receive information about the elements determined and
          the concentration levels measured.

     3 -- Overall scan - This choice will present all the analyses which have
          been performed on wheat flour.  You may sort these as you wish and
          receive information about the elements, techniques, and levels of
          concentration measured in these analyses.

                                                        Your choice?
```

Figure 7. Initial database menu.

COMPOSITION OF wheat flour

Number of records = 76

ID#	Element	Technique	Conc. Level*	SRM#
17	Cs	inaa	U	1567a
18	Cu	faas	T	1567a
19	Cu	faas	T	1567a
20	Cu	faas	T	1567a
21	Cu	icp	T	1567a
22	Cu	inaa	T	1567a
23	Cu	inaa	T	1567a
24	Cu	rnaa	T	1567a
25	Fe	faas	T	1567a
26	Fe	faas	T	1567a
27	Fe	icp	T	1567a
28	Fe	idms	T	1567a
29	Fe	inaa	T	1567a
30	Hg	cvaas	T	1567a
31	Hg	rnaa	U	1567a
32	I	inaa	T	1567a

Choose: F(wd), B(ack), H(elp), S(elect), or X(it)

*Definition of Concentration Levels:

M Major - analyte present at >1.0 wt%
m minor - analyte present at 0.1 - 1.0 wt%
T Trace - analyte present at 1 - 999 ug/g
U Ultratrace - analyte present at <1 ug/g

Figure 8. Data screen for a sort by element search of wheat flour SRM.

WHEAT-FLOUR is assigned to the parameter CAT-AG. At this point, all of the IF conditions in rule 35 are met, and the tasks specified in the THEN portion are activated by the inference engine of the ES; the value "wheat flour" is assigned to the parameter MATRIX, and the value of the parameter MATCH1 is set to YES. An SRM such as water appears as a selection in several menus, environmental and clinical, in addition to agricultural. This allows the user to arrive at an SRM from several starting points.

The inference engine now executes rules 141 and 118 (Figure 5). Rule 141 informs the user of the SRM selected. Rule 118 writes the values of the parameters STAT and MATRIX ("match" and "wheat flour" in the above example) to an ASCII text file, LINKFILE.TXT, and calls the executable DBM program, CLIPNBS.EXE (Figure 6). The DBM program first reads the name of the SRM from LINKFILE.TXT to a temporary DBF file, TEMP.DBF, and then creates an indexed file of records containing the name of the specified SRM (wheat flour). Using this indexed file, the user can examine the elemental composition of the SRM in several ways: by element, by instrumental technique, or by level of concentration (Figure 7). The sample matrix under consideration appears at the top of this display as a reminder for the user. The DBM program displays screens showing the data in the format selected by the user (Figure 8). Choices associated with the data screens include scrolling forward or backward, selecting the analytical descriptors on a given line

```
Your transfer went successfully.

If you Quit dBase now and return to the expert system the system will import
the following parameters:

        matrix -- wheat flour
        SRM Number -- 1567a
        element chosen -- Fe
        concentration level -- T
        Method of analysis using -- icp

Here are your options:

    0 - You may Quit dBase and return to the expert system.
        The parameters above go with you.  If you wish to
        change the matrix, or wish to continue with the
        overall consultation, this is the choice for you.

    1 - You may return to the Main Menu, where you can select
        (ultimately) a new set of parameters to be exported.

        Enter your choice:
```

Figure 9. Analytical descriptors transferred from the database to the expert system.

for transfer to the ES, or exiting to the original menu of the DBM program. Optional help screens provide definitions of concentration levels, explanations of acronyms associated with the measurement techniques, and general instructions for using the DBM program.

When the user locates and enters the identification number of the SRM data line that best represents the descriptors for the user's analysis, a screen containing only these values appears. For example, depressing the S(elect) key and entering ID 27 (Figure 8) causes the screen shown in Figure 9 to appear. The user now has the option of (1) transferring these descriptors along with the name of the SRM to the ES or (2) continuing to examine the composition of the SRM. If the transfer option is exercised, the DBM program writes the information associated with the four descriptors to an ASCII text file, TOPCFILE.TXT (Figure 6) and then returns to the ES. At this point the first module has completed its function. In situations where the user knows the name or number of the SRM most closely resembling the sample, the ES immediately calls the executable DBM program. This program asks the user to enter either the name or number of the SRM. A valid SRM name or number causes the menus associated with the DBM to appear, and the user interacts with the program as described above.

STRATEGY FOR CONTROLLING THE ANALYTICAL DESCRIPTORS MODULE

As we have seen, the values assigned to the parameters of the ES determine the overall performance of the system. In rule 35 (Figure 3), CAT-GEN, CAT-BIO, and CAT-AG are system parameters indicating the classifications for SRMs. In this ES, these parameters receive their values directly from the user, who is prompted by a series of menus. The order of menu presentation is determined (1) by the backward chaining logic of the knowledge base and (2) by the arrangement of the parameters in the rule (the leftmost parameter receives a value first).

In the backward chaining strategy used in this system, the ES starts with what it wants to prove, the value for the parameter designated as the goal parameter, and only executes rules relevant to establishing this goal. In the current example (rule 118, Figure 5), the parameter MATCH3 has been marked as a goal parameter, and the goal is reached when MATCH3 has a value of YES. Thus, the first rule to be tested is one whose THEN portion assigns a value of YES to MATCH3. If a parameter in the IF portion of this rule has not been assigned a value, the ES attempts to assign one by testing other rules that contain this parameter in their THEN portions. If the conditions of these rules cannot be met, the user is prompted for the value.

```
Parameter:CAT-GEN
   TRANSLATION  :: the menu for general classification of samples
        PROMPT  :: :attr (yellow) SELECTION OF THE GENERAL CATEGORY OF
                   THE SAMPLE MATRIX  :LINE 2 Into which of the following
                   general categories does the sample fit?  :LINE 2  :attr
                   (BLINK blue) PRESS F1 FOR HELP
          HELP  :: :ATTR (cyan) The system is attempting to match the
                   matrix of your sample with the matrix of an SRM.
                   Please provide thoughtful responses to the series of
                   menus which follows.  If the match provided does not
                   appear to be suitable for your sample matrix, you ...
        EXPECT  :: BIOLOGICAL-MATERIAL CLINICAL ENVIRONMENTAL
                   FERROUS-ALLOYS NONFERROUS-ALLOYS GASES-IN-METALS
                   HIGH-PURITY-METALS PURE-CHEMICAL FERTILIZER ORE
                   ROCK-MINERAL NONE
          TYPE  :: SINGLEVALUED
```

Figure 10. Properties of the parameter CAT-GEN.

In the above example, rule 118 (Figure 5) containing MATCH3 is tested first. The conditional IF portion requires a YES value for the parameter MATCH2, so the ES now traces this parameter to the THEN part of rule 141. The condition of rule 141, however, requires a value for MATCH1, so the ES traces MATCH1 to rule 35 (Figure 3). The ES system now attempts to find rules that give values to the parameters CAT-GEN, CAT-BIO, and CAT-AG and fails. At this point the ES prompts the user for the values. When the values entered meet the conditions of rule 35, the entire process is reversed; rule 35 sets the value for MATCH1 to YES that now meets the condition of rule 141 and sets the value of MATCH2 to YES, which finally causes the goal parameter MATCH3 to receive a YES value.

To avoid "dead-ending" the user, it is necessary to provide a smooth exit from the system for each of the following situations. (1) A category has been selected for which no rules have been implemented, e.g., FERROUS-METALS. (2) The user entry indicates that none of the menu selections resembles the current sample. In either case, the user is given the opportunity to enter another sample or exit the system. The inclusion of categories such as FERROUS-METALS on the menus indicates to the user areas to be addressed by the system in the near future.

The properties of a typical parameter, CAT-GEN, are shown in Figure 10. The PROMPT field indicates the title of the menu screen presented to the user, the EXPECT field generates the menu of values shown in the first column of Figure 5, and the help screen contains the contents of the optional explanation screen that can be obtained when the user depresses the F1 key. When the user selects one of the items from those contained in the EXPECT field, the selection becomes the current value for the CAT-GEN parameter and is automatically assigned a certainty factor[13] of +100, indicating that the user is certain that the selected value is true. The assignment of certainty factors in this manner does not allow for weighing of answers, known as "fuzzy logic"; i.e., the user cannot indicate that he is reasonably sure (confidence factor approximately 75%) a given choice characterizes the sample.

A total of approximately 75 SRMs will be available for matching in two general categories, biological and environmental. Some SRMs, such as water, fall into several groups—aqueous environmental, agricultural, and food and beverages. In some cases the sample may be best represented by more than one SRM. For example, when oil shale is the sample of interest, the ES will refer the user to both the clay and fuel oil SRMs to indicate the geological and organic nature of the sample.

DEVELOPMENT OF THE SYSTEM

A brief history of the system will show the evolution of our approach to the final implementation. One principle was most useful in the development of this module—optimization of the performance of each of the system components. The ES should not be searching or sorting large volumes of structured information, nor should the DB be struggling with situations

requiring the rules of thumb (heuristic knowledge) used by the expert.

Initially, the rules of the ES contained all the information necessary to obtain the four analytical descriptors. While this exclusive use of an ES was helpful in defining the organization of the knowledge domain, it soon became obvious that the use of a DB was essential for further development. At first the DB was relegated to the role of positioning data for access by the ES. A number of functions built into PC-Plus allowed direct interaction between the rules of the ES and dBASE III PLUS. Thus, values from the database were assigned to ES parameters, and these parameters were then used to generate prompts for the user. In this way, unique prompts were generated from each interaction between the ES and DB. The multiple interactions between the ES and DB resulted in a prototype system in which the PC-Plus and dBASE III PLUS were intimately intermingled.

The resulting hybrid system was operational but slow and clumsy to use. Each interaction of the ES with the DB resulted in a delay due to the dBASE III PLUS introduction displayed on the screen and the DBM program and files having to be reloaded from the disk. In addition to this pause, the appearance of the dBASE screen was disconcerting to the users. (It is impossible to disable this screen.)

At this point, a dBASE program was written to perform the operations for selecting the analytical descriptors. This program could be called once by the ES, allowing the user to select the descriptors and automatically returning these values of the ES. This approach minimized the interactions between the ES and DB and increased the overall efficiency of the hybrid system. The DBM program sorted the information for a given SRM by element, technique, or concentration level and presented it to the user who could then quickly scan this information and select a set of analytical descriptors for the sample under consideration. This version of the system reduced the time necessary to obtain a set of analytical descriptors by a factor of 8 over the previous version.

The system was optimized further by compiling the dBASE program. The Clipper program was used to produce an executable program in object code from the source code of the dBASE programs described above. Several macro commands used in the dBASE source code had to be broken down into elementary commands for the Clipper compiler. The compiled DBM program ran approximately twice as fast as the programs written in dBASE source code. This difference in speed between object and source code programs should increase as the size of the DB increases.

A final consideration in the development of the system was that of making the interaction between the user and the computer as uncomplicated and informative as possible. The contents of each screen of the ES were color-coded—yellow for input instructions, white for results, cyan for information required to understand instruction or results, and blue for special messages such as the availability of help screens or graphical displays. Each ES screen contained a title on the first line to inform the user of the progress of the consultation. The DB screens were also color-coded to improve the displays. Use of the ES also permitted rapid, convenient implementation of help screens, both text and graphics.

Since the method used for the matching is transparent to the user, this task could have been implemented by using the dBASE programming language. The use of the PC-Plus development software allowed greater flexibility in designing an effective user interface. Another factor favoring the use of PC-Plus was the use of this ES shell in the second module of the system to predict conditions for the sample digestion.

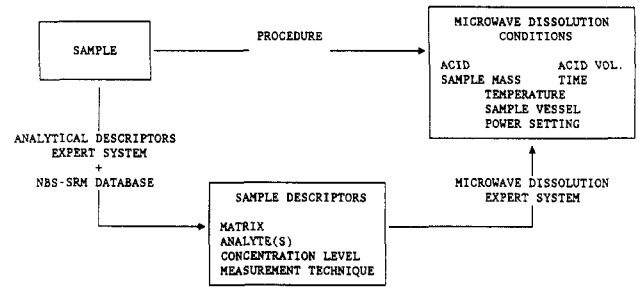


Figure 11. Microwave information system.

Limited response from users to date indicates a positive reaction to the color-coded menus and help screens generated by the ES. They also reacted favorably to the menus of dBASE. However, they found the selection of analytical parameters from the SRM database to be restrictive for many analyses. This indicates that the SRM database should be expanded to provide a broader choice of analytical parameters.

USE OF ANALYTICAL DESCRIPTORS FOR MICROWAVE DISSOLUTION OF SAMPLES

The ES component for microwave dissolution uses the four descriptors to define a procedure for dissolution of the sample.[14] If, for example, the analyst needs to determine iron at trace levels by ICP in a matrix resembling wheat flour, the expert system asks the user for the following additional information: number of samples to be digested simultaneously, estimated moisture content of the sample material, digestion vessel (size and material), and the maximum power delivered to the sample cavity by the microwave unit being used. The system then provides the information shown in Table II, which agrees with the actual conditions for the dissolution of a dry wheat flour SRM for the analysis specified. The system also provides a list of all elements in the SRM matched with current sample that have been determined by the technique selected.

The analyst now has the information required to prepare samples and to set the microwave power output to obtain the target temperature in the optimum time for the dissolution. The resulting ES–DB system combines the classical methods of acid decomposition with the new concepts of microwave technology to provide expertise in a form that can be transferred efficiently to a broad audience. The procedures produced from this expertise can be carried out under controlled conditions, thus assuring both laboratory safety and quality control of the dissolution. This component is currently being expanded to handle a variety of sample types with corresponding sample preparation procedures.

Once a procedure has proven successful in the dissolution of a given sample type, it may be stored by the system and accessed directly (Figure 11). Thus, the system has the capability of "learning" new procedures.

SUMMARY

The prototype component of a hybrid ES–DB system that assists analysts to determine four analytical descriptors critical for selecting methods of sample preparation has been implemented. The system is currently being evaluated in several laboratories. Preliminary results indicate that the combination of the heuristic ability of an ES with the properties of a complied relational DB provides an efficient means of transferring current procedures for sample preparation to laboratory personnel.

ACKNOWLEDGMENT

This work was funded in part by the sabbatical programs of Virginia Military Institute and Washington and Lee

University. NSF Grant CHE-8517147 provided partial support for Dr. Pleva.

REFERENCES AND NOTES

(1) Barnes, R. M. *Spectroscopy* **1987**, *1*(5), 24.
(2) Kingston, H. M., Jassie, L. B., Eds. *Introduction to Microwave Sample Preparation: Theory and Practice*; American Chemical Society: Washington, DC, 1988.
(3) NBS Standard Reference Materials Catalog 1986-87. *NBS Spec. Publ.*, **1986**, *260*.
(4) Taylor, J. K. "Handbook for SRM Users"; *NBS Spec. Publ.* **1983**, *260-100*.
(5) Brachman, R. J.; Levesque, H. J. In *Expert Database Systems, Proceedings of the First International Conference*; Kerschberg, L., Ed.; Benjamin/Cummings Publishing: Menlo Park, CA, 1986; Part I, p 6.
(6) PC-Plus (version 3.01); Texas Instruments Inc.: Austin, TX, 1987.
(7) Waterman, D. A. *A Guide to Expert Systems*; Addison-Wesley: Reading, MA, 1986; pp 112-116.
(8) *Ibid.*, Chapter 25.
(9) *Clipper, Applications Manual*; Nantucket Corp.: Culver City, CA 90230.
(10) Robbins, J.; Brady, K. *Expert dBASE III PLUS*; Sybex Inc.: San Francisco, 1987; Chapter 12.
(11) Al-Zobaidie, A.; Grison, J. Expert Systems and Database Systems: How Can They Serve Each Other? *Expert Syst.* **1987**, *4*(1), 30.
(12) Settle, F.; Pleva, M.; Booker, C.; Iams, H.; McClintock, D.; Moore, T. An Information System for Selecting Methods of Chemical Analysis. *Am. Lab.* **1987**, *64*(4), 17.
(13) Schutzer, D. *Artificial Intelligence—An Applications-Oriented Approach*; Van Nostrand Reinhold: New York, 1987; pp 59-67.
(14) Kingston, H. M.; Jassie, L. B. *Anal. Chem.* **1986**, *58*, 2534-2541.

Automation of Numerical Data Compilations

RANDOLPH C. WILHOIT* and KENNETH N. MARSH

Thermodynamics Research Center, Texas A&M University, College Station, Texas 77843

Received August 19, 1988

The preparation of large numerical data compilations of interest to chemists includes searching the literature, identification of pertinent data, interpretation, selection and smoothing of data, creation of suitable printed tables and/or computer-readable files, publication, and distribution. As scientific information accumulates, the magnitude of all these tasks increases. Many of them can be, and are being, automated. The development of computer hardware and software is changing the way in which such compilations are produced and used.

THERMODYNAMICS RESEARCH CENTER PUBLICATIONS

The Thermodynamics Research Center operates four continuing projects for the collection and compilation of thermodynamic properties of organic and nonmetallic inorganic compounds.

TRC Thermodynamic Tables—Hydrocarbons[1] are a continuation of the American Petroleum Institute Research Project 44 started under the direction of F. D. Rossini in 1942. They give the selected best values of various properties, based on measurements reported in the scientific literature, when available, or on theoretical calculations or correlations when not available. The tables are issued in the form of two supplements of loose-leaf data sheets per year. The current set consists of about 3000 sheets on some 35 properties of around 6000 hydrocarbons and sulfur compounds of importance to fuel technology. At the top level, the tables are organized by types of properties identified by alphabetic codes. Thus, for example, a-tables contain density and refractive index at 20 and 25 °C, normal boiling point, dt_b/dP, and melting point. ka-tables contain values of parameters of the Antoine equation for vapor pressure and boiling points at a series of standard pressures from 0.05 to 1500 Pa, k-tables from 0.02 to 2.0 bar, and kb-tables from the boiling point at 2 bar to the critical temperature. h-tables contain values of the second virial coefficients. p-tables contain thermochemical data at 25 °C. Altogether there are 30 types of property groups. Within each property group the tables are organized by classes of compounds, identified by a finding number that identifies the elements and a table number that identifies a particular homologous series or group of isomers. A Specific Reference sheet listing references to sources of data accompanies each table of numerical values. Until 1984 the Specific Reference sheets listed only the authors' names. The complete references are contained in a General List of References at the end of the set. Since 1984 the complete references are given directly on the Specific Reference sheets.

TRC Thermodynamic Tables—Non-Hydrocarbons[2] contain selected values of properties of other types of organic compounds and nonmetallic inorganic compounds in a manner similar to the Hydrocarbon tables. They began in 1957 under the sponsorship of the Manufacturing Chemists Association. The current set contains about 2440 sheets on some 2800 compounds.

To keep the sets current and useable, subscribers must file the supplements of new and revised tables in the appropriate place as they are received. Because of the open-ended nature of these sets and the complexity of the organization, this requirement places a heavy burden on the recipients. A recent reissue of both complete current sets included page numbers, which simplifies the filing operation somewhat but does not eliminate it.

In 1986 a formula index listing the compound name, Chemical Abstracts Service Registry Number, Wiswesser line notation, table numbers, and page numbers was included with each set. It is revised biennially.

The NBS Project on Properties of Chemical Compounds has published reviews and data compilations in the *Journal of Physical and Chemical Reference Data* during the past 24 years. It has been supported by the Office of Standard Reference Data of the National Bureau of Standards. Major groups of compounds reviewed so far include aliphatic alcohols, halogenated ethanes, organic oxygen compounds containing one to four carbon atoms, and, most recently, organic amines.[3] The organic section of the Bulletin of Chemical Thermodynamics,[4] an annual bibliography on thermodynamic data, is part of the NBS project at the Thermodynamics Research Center.

The *Internationl Data Series*[5] is a collection of tables of thermodynamic properties of binary nonelectrolyte mixtures. The publication is similar in some respects to a research journal. Data are contributed by authors and evaluated by

0095-2338/89/1629-0017$01.50/0