

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—Hydrocarbons¹ 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-Hydrocarbons² 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 *International 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

an international panel of editors. The editor-in-chief is Dr. Henry Kehiaian, located in Paris, France. Each sheet presents values of a particular property on a particular system in a predefined format in SI units. Both original experimental data and extracts of previously published data are included along with information on measurement technique, purity of materials, and measurement accuracy. The series began in 1973 and appears as four supplements per year. It now contains over 3000 sheets.

During past years the Thermodynamics Research Center has also prepared a number of special compilations for organizations such as the American Petroleum Institute, The Gas Processors Association, and the Department of Energy.⁶ Most recently, a project on thermodynamic and transport properties of certain mixtures was started for the Design Institute for Physical Property Data of the American Institute of Chemical Engineers.

STEPS IN THE PREPARATION OF NUMERICAL DATA COMPILATIONS

The first step in the compilation of experimental data is the location of sources of such data in the primary literature and the extraction of the pertinent information. One or more of the following additional steps is needed to produce the final product: (2) normalization; (3) assignment of uncertainties; (4) smoothing; (5) adjustment for internal consistency; (6) selection of recommended values; (7) design of a format suitable for dissemination.

Normalization refers to the conversion of the collected data to some common basis. It usually involves conversion to a consistent set of units and reference states and may include recalculation with a consistent set of auxiliary data such as atomic weights and fundamental constants. It may also include corrections for variations in instrument calibrations or the presence of impurities. Normalization is required for intercomparison of related data from different investigators.

The assignment of uncertainties is a critical step that is difficult to describe rigorously. A statistical analysis may play a role, but more often a subjective judgment, taking into account the state-of-art at the time of publication, the limitations of the experimental technique, the information provided by the author, and the reputation of the investigators, is the basis of the final selection.⁷

Smoothing converts sets of experimental data to some regular function of the appropriate independent variables. This may be done graphically by reading values from a curve drawn through or near the plotted data points or analytically by fitting data to a smoothing function according to some criterion. The analytic procedure usually consists of a least-squares fit to the parameters in a smoothing function. It is common practice to derive statistical weighting factors for the fit from the assigned uncertainties. The function may be entirely empirical or based to some extent on theoretical considerations. The independent variables may be state variables such as temperature, pressure, and composition or may include parameters describing molecular structure or other properties. The purpose of smoothing is to reduce the effect of random experimental errors, to reduce the size of the original data set or represent it by a mathematical function, to generate tables at regular increments of the independent variables, and to help in the intercomparison of data from different sources. IUPAC has published a guide for smoothing properties of fluids.⁸

Data values may be adjusted to satisfy some known relationships among various properties. This kind of internal consistency is especially important for thermodynamic data because of the many rigorous relationships implied by thermodynamic principles. For example, vapor pressure is related, through the Clapeyron equation, to the enthalpy of vapori-

zation and the volumetric properties of coexisting condensed and gas phases. The enthalpy of vaporization is in turn related to the entropy and heat capacity of these phases. Users of thermodynamic compilations often derive properties other than those directly tabulated. Even if each tabulated value is within the assigned uncertainty tolerance, the derived values may be unreliable if the tabulated set is not consistent.

In principle such relationships can be incorporated into the smoothing process by treating them as mathematical constraints on the fitting criterion. They can also be accommodated by simultaneously fitting sets of related properties to the same set of parameters by the use of an appropriate objective function. Thus, an equation of state might be fit simultaneously to P - V - T data, to vapor pressure, to enthalpy of vaporization, and to changes in heat capacity with pressure. Thermochemical properties for a set of chemical reactions involving common reactants and/or products are related by a set of linear algebraic equations. An entire set might be simultaneously fit by incorporating such constraints.

If the objective of the compilation is to present the single most accurate value for each property and system covered, the choice must be made at some point. It may be based on data from a single source or on some kind of average of data from several sources or from the smoothing process. The final selection may result from an iterative application of several of the preceding steps.

Examples of compilations incorporating all conceivable combinations of these steps can be found. They range from the direct listing of raw data from the literature to the presentation of just the parameters in the recommended smoothing or correlating equations.

Most familiar compilations of physical and thermodynamic properties have been issued as printed tables. Some are in the form of bound volumes, possibly with subsequent revisions or supplements to maintain currency. Others, such as the TRC Thermodynamic Tables, are issued as a continuing series of loose-leaf data sheets. We now see a strong movement toward distribution of compilations in computer-readable form. This trend will undoubtedly become stronger in the future.

AUTOMATION OF DATA COMPILATIONS

Most of the data compilation steps identified above, from the literature search to the composition and publication of tables, are amenable to computer automation. The exceptions are the identification and interpretation of pertinent data from the literature, the assignment of uncertainties, and the final selection of the best values. The evolution of practices at the Thermodynamics Research Center over the years reflects this situation. For a number of years all the steps were carried out manually. Literature searching was done by direct scanning of journals, reports, and *Chemical Abstracts*. Our database was in the form of handwritten 3-in. \times 5-in. index cards, each containing the value of one property of a particular compound along with the reference, sample description, and comments. Around a half million of these were collected during the first 45 years of the project. Tables were composed by manual entry on various typesetting devices.

A major problem has been keeping appropriate records on all these activities. To generate a new table, a compiler will accumulate all of the available information related to that table and then carry out the evaluations, smoothing, and selections. To revise a previously issued table, it is necessary to gather both the data available for the previous version and the newer data accumulated since that time. The whole process of evaluation and selection must be repeated for all the values in the table, even though some of them may not change. It is important that the person making such revisions understand the procedures and reasoning of the previous compiler even

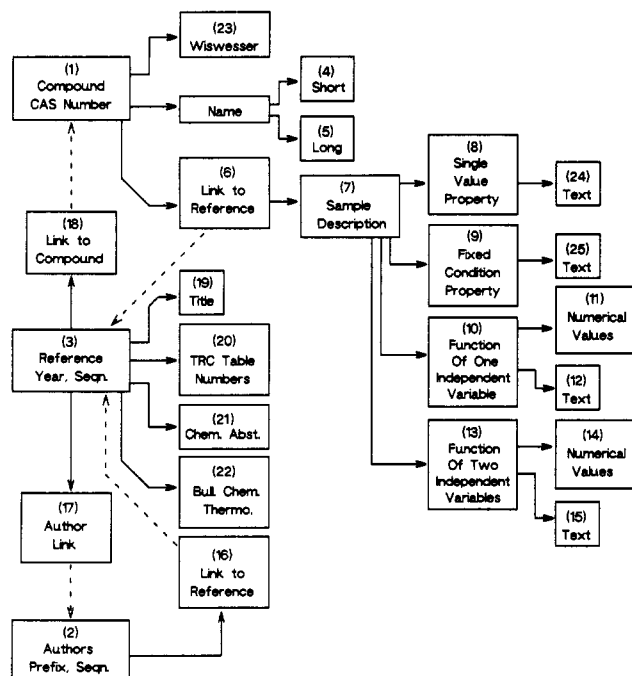


Figure 1. Schema TRC source file.

when 10–30 years has elapsed. With the passing of time, the record-keeping chores become rapidly more burdensome because both the number of extant tables and the size of the pool of data in the literature increase.

In recent years we have moved toward automating as much of this process as possible. The purpose is to allow the compilers to concentrate on the critical tasks of interpretation, evaluation, and selection of data. Although we make use of on-line literature searching services such as those offered by Chemical Abstracts Service, we cannot rely on them exclusively. The index terms available in such searches do not efficiently identify experimental property data. In principle, direct text searches of articles, such as ACS JOURNALS ONLINE, could be useful. Unfortunately, tables of numerical data, which are our main input, are not yet included. The automated searches do furnish an initial screening, but much manual searching is still required.

The other end of the process is now running smoothly. Compilers enter their selections for new and revised tables into a computer file. Software has been developed to convert these files into files that drive a laser printer. Master copies of the final tables and reference sheets are then produced in-house.

COMPUTER RESIDENT DATABASES

To help automate the intermediate compilation steps, the Thermodynamics Research Center has implemented two large computer-readable databases during the past three years. The TRC Source File contains numerical values of physical and thermodynamic properties extracted from the chemical literature. The TRC Selected Data File contains the selected values that have been published in the printed versions of the two TRC Thermodynamic Tables—Hydrocarbons and Non-Hydrocarbons. Both are organized by a commercial database management system that can retrieve data in either a hierarchical or a relational mode.⁹ The results can be displayed directly on the console screen or sent to external files for printing or further processing.

The schema of the TRC Source File is illustrated in Figure 1. Each box in the figure represents a type of record or, in relational database terminology, a relation. In any case the records are indexed by certain sets of key values, or attributes in relational terminology. The solid lines show the hierarchical

relationships among the key values. The set of key values at any level includes all those at the next higher level. Dashed lines represent links in which the data fields in one set of records are key values for another set of records. Records of type 1 are indexed by the Chemical Abstracts Service Registry Numbers. They also contain the molecular formula and a systematic name. Additional names, as many as needed, are kept in records 4 and 5. Provision is made for Wiswesser line notations, but not many are included at this time. References are identified in records of type 3. They are keyed by the year of publication and sequence number within the year. Each of these records contains the reference citation, a comment field, and key values for up to three authors. Additional authors are identified in records of type 17. Titles of articles and references to *Chemical Abstracts* and the *Bulletin of Chemical Thermodynamics* are stored in records 19, 21, and 22. The finding numbers of TRC tables that contain data related to data in the reference are stored in records of type 20. Names of authors, keyed by the first six letters of the last name and the initials, are stored in records of type 2. Key values of references written by a given author are stored in records of type 16. Registry numbers of compounds reported in a particular reference are kept in records of type 18.

The remaining records are concerned with the numerical data and their descriptions. Record 6 identifies the references that contain data on a particular compound. Following this hierarchy, record 7 describes the sample (preparation, purification, and purity) for a particular compound in the reference identified in record 6. Records 8–11, 13, and 14 contain numerical values of measured properties and various attributes of pure compounds. The various properties are identified by a list of 35 basic symbols. Record 8 stores properties, such as triple point, a normal boiling point, or a critical property, represented by a single number. Properties that are functions of some variable, as temperature or pressure, are stored in records 9–11. These correspond to measurements along a two-phase coexistence line or a single phase at constant pressure or temperature. If only a few values are given, record 9 is used; if more extensive tables are given, records 10 and 11 are used. Properties that are functions of two independent variables are stored in records 13 and 14. Text comments are stored in records 12, 15, 24, and 25.

The data in the Source File have been normalized. They are converted to SI units as entered. Each one is also accompanied by a preliminary estimate of uncertainty. The records identify the phase or phases present, constraints if any, conditions of measurement, and other information useful for evaluation. At present the database contains 25 000 registry numbers, 29 000 references, and around 115 000 numerical values. The total number of records of all types is 385 000.

The schema for the TRC Selected Data File is illustrated in Figure 2. It is designed primarily to regenerate the tables as printed in the TRC Thermodynamic Tables—Hydrocarbons and Non-Hydrocarbons. Record type 1 contains the Chemical Abstracts Service Registry Number, formula, and the single name used in the printed tables. Data for the various tables are stored in records of type 6. They are keyed by the table number that appears in the title of each printed table. Data for the specific reference sheets are stored in records of type 9. The references are identified by the same key used in the TRC Source File. Records of type 2 identify the tables in which data for a particular compound may be found. The other records contain the information needed to regenerate complete printed versions of the tables. At the present time we have accumulated about half the data in the extant sets. The software that drives the printer to produce new tables also merges the new data into this database.

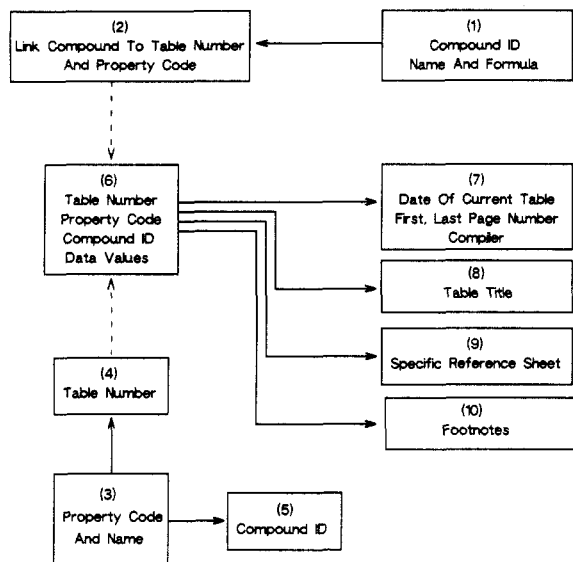


Figure 2. Schema TRC selected data file.

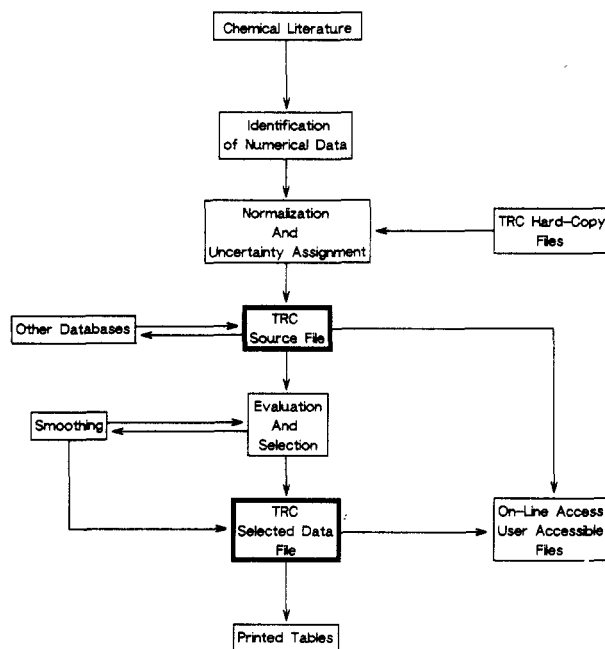


Figure 3. Data flow for TRC compilations.

For the past 10 years the *International Data Series* tables have been produced from computer files assembled in France. All of these data, along with other mixture data, are kept in a database maintained by Dr. Kehiaian.

GOALS FOR FURTHER AUTOMATION

Our immediate goal is to link the various steps more tightly together. This is illustrated in Figure 3. Experimental data, from both the chemical literature and existing TRC hard-copy files, are entered into the TRC Source File. We are also arranging for exchange of information with databases being maintained at other locations. Whenever an appropriate set of related data is complete, the evaluation and selection process can begin. The box labeled "smoothing" represents a group of linear and nonlinear least-squares fitting programs. The data to be examined are extracted from the Source File and displayed as text or graphs or sent to the smoothing software. The selections, along with the references to sources, are then placed in the Selected Data File. Data sheets for the two TRC thermodynamic tables projects can be generated automatically at this point. This will also give us the practical option of

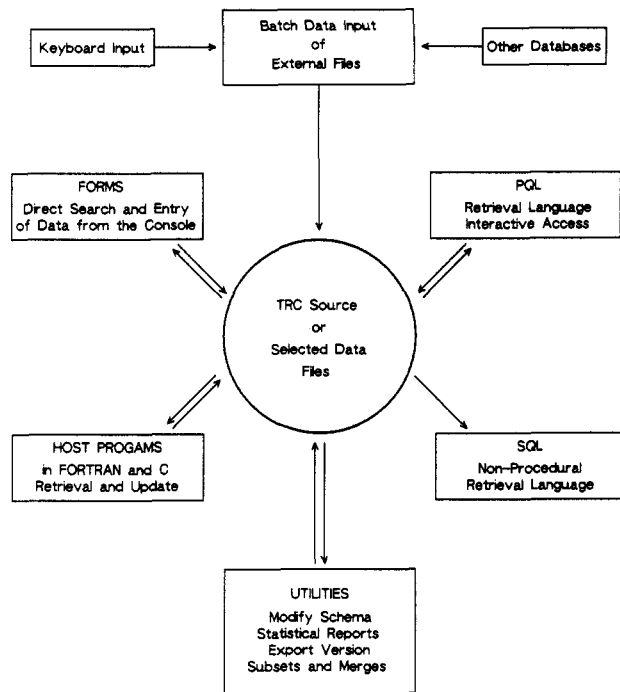


Figure 4. Interaction with TRC databases through use of SIR DBMS.

changing the format and organization of the tables for special purposes.

Of greater importance, this scheme will make possible the distribution of machine-readable files from both the Source File and the Selected Data File. We have conducted an experiment along this line for the past 2 years. Vapor pressure data extracted from our files have been sent to Numerica, an on-line service offered by Technical Database Services, Inc.¹⁰ It is in the form of two files. One from the TRC Source File contains the direct experimental data, estimated uncertainties, and references. The current collection contains about 16 000 compound-reference combinations. The other contains parameters in a smoothing equation for vapor pressure from the TRC Selected Data File and from other compilations. From this the Numerica system generates tables of vapor pressure at designated temperatures (or boiling points at designated pressures) in various units. A rating of accuracy is included. Parameters for nearly 6000 compounds are included here. Additional properties are being planned. We are also experimenting with a floppy disk version of this data set.

The picture that is emerging is shown in Figure 4. Data from the literature and from exchange with other databases flows to the TRC Source File. The residue after evaluation, selection, and smoothing goes to the TRC Selected Data File. Information from the two computer-readable files can be readily manipulated to produce a variety of outputs.

Although both experimental and selected data are represented by numerical values with associated descriptions, they differ in some fundamental ways. The Source File is an archive of published information. It must be able to accommodate a wide, and sometimes unpredictable, variety of properties, systems, and experimental conditions. Each number is associated with only a single source, but there may be, and usually are, duplications of a particular compound-property combination. The Source File grows as new data are entered, but each record is static, except for correction of errors. A much smaller set of properties and conditions is required for the selected values. Many of them refer to idealized standard states or take the form of parameters in smoothing functions. Each compound-phase-property combination has only a single value (or a single-valued function of state variables), but it may be derived from many sources. The collection of selected

values is dynamic in the sense that it changes as new data become available and are evaluated. For these reasons we make a definite distinction between experimental and selected data and store them in different databases that are organized differently.

Normally, a user of data would prefer to have the selected values. But, since the selection process will always lag behind the available published data, access to the direct experimental data is needed as well.

The identification, interpretation, evaluation, and transcription of numerical data from the literature is a slow step that cannot be easily automated. However, duplication of this work among the several large databases being maintained around the world can be reduced by mutual exchanges of data in machine-readable form. TRC has been involved in the development of two means of ready transfer of data. One is the establishment of a universal standard for coding and formatting of numerical thermodynamic property data for exchange among databases. A proposal, termed COSTAT, is being considered by several CODATA task groups for a possible official CODATA recommendation.¹¹ If this, or a similar, standard is generally accepted, managers of individual databases would develop software for converting their local format and codes to and from the COSTAT standard. Thus, COSTAT messages could be easily exchanged among the particular data collections. The other procedure is the creation of interactive software packages that convert keyboard entry of data into a form that can be readily merged with the TRC Source File. We have also conducted some experiments in such off-site entry of data.

The current versions of our two databases accept data only for pure compounds. We have started the expansion of the TRC Source File to accommodate data on mixtures as well. This requires a more complex indexing scheme capable not only of locating systems of any particular set of components but also of locating systems formed from any subset of components. Additional record types that include composition variables, a larger number of coexisting phases, and an expanded set of property codes including partial molal and excess quantities are needed. Altogether these will double the number of record types in the schema.

APPENDIX

The operation of a large complicated database such as the TRC source file requires the establishment of some general policies as well as many procedural details. These are described in an internal TRC document. The principal objective is to record the numerical values of directly measured thermodynamic properties along with sufficient descriptive information, called metadata, to give them meaning and help with the comparison and selection of the best values. The metadata are also used by the retrieval software to identify particular properties and systems. The data and metadata are recorded in a manner such that revisions are not needed, except to correct errors. It is not practical to record all pertinent metadata for all kinds of data in the literature. Our goal is to make the primary numerical values available to the compiler and to guide him in a preliminary selection of the best values. It will sometimes be necessary to consult the original document to resolve close calls.

There is no clear or simple distinction among data in the literature between directly measured properties and derived properties. Thus, many arbitrary decisions must be made about what data to include in the database. For example, standard enthalpies of formation of compounds are frequently reported. These are usually derived from various kinds of primary data. If they are based on heats of combustion in a bomb calorimeter, we prefer to store values such as $\Delta_c H^\circ / M$

and $\Delta_c H$ rather than, or in addition to, $\Delta_c H^\circ$. If a reported enthalpy of reaction or phase change was obtained by a second-law calculation, we would record the observed equilibrium constants or pressure-temperature data, along with any auxiliary data used in the calculation. If a reported adiabatic compressibility was calculated from the observed speed of sound, we would include the speed of sound data as well.

Normalization of data to SI units simplifies the preliminary comparison of related properties from different sources. This also creates some potential problems. Changes in definitions of units and in fundamental physical constants over the years cause changes in factors for conversions of units. We can expect such changes to occur in the future also. Our practice is to convert the units used for the original data to SI units by a standard set of conversion factors to produce data in "nominal" SI units. Thus, the original values can be easily retrieved if needed. For example, we use the factor 4.184 to convert all data in calories (except Steam Table calories) to joules. However, in past years, especially before 1940, many kinds of calories, and several kinds of joules, were used. When the accuracy is sufficient to justify these distinctions, the exact unit or calibration procedure is described in the comment record associated with the data value.

The propagation of changes in atomic weights and in the fixed points and interpolation equations that define the temperature scale among thermodynamic data are especially complicated. Except for density and heats of combustion, we use the mole as the amount of substance. When the original values are reported on the mole basis, we retain the value but include the original atomic or molecular weight on the comment record, if justified by the accuracy of the measurement. When the original values are reported on the mass basis, we convert them to the mole by using the current official table of atomic weights and record these values if justified. Original temperature scales are retained, except for the conversion of Celsius to Kelvin by addition of 273.15 and the usual formulas for conversion of Fahrenheit or Rankine. If the measurement is of sufficient accuracy, a more specific definition of the scale is given in the comment record.

These practices produce a set of nominally normalized data that can be quickly compared and used to narrow the range of choices. They also avoid the necessity of revising the stored data every time fundamental physical constants or auxiliary data are changed. The exact unit conversions and corrections are then applied only to the remaining candidates at the time a selection is made.

REFERENCES AND NOTES

- (1) TRC Thermodynamic Tables—Hydrocarbons, Thermodynamics Research Center, Texas A&M University, College Station, TX 77843 (formerly API Project 44 tables).
- (2) TRC Thermodynamic Tables—Non-Hydrocarbons, Thermodynamics Research Center, Texas A&M University, College Station, TX 77843 (formerly TRC Data Project).
- (3) Examples: Wilhoit, R. C.; Zwolinski, B. J. "Physical and Thermodynamic Properties of Aliphatic Alcohols". *J. Phys. Chem. Ref. Data* 1973, 2, Supplement No. 1. Chen, S. S.; Wilhoit, R. C.; Zwolinski, B. J. "Ideal Gas Thermodynamic Properties and Isomerization of *n*-Butane and Isobutane". *J. Phys. Chem. Ref. Data* 1975, 4, 859. Chen, S. S.; Wilhoit, R. C.; Zwolinski, B. J. "Thermodynamic Properties of Normal and Deuterated Methanols". *J. Phys. Chem. Ref. Data* 1977, 6, 105. Wilhoit, R. C.; Chao, J.; Hall, K. R. "Thermodynamic Properties of Key Organic Oxygen Compounds in the Range of C1 to C4. Part 1. Properties of Condensed Phases". *J. Phys. Chem. Ref. Data* 1985, 14, 1. Chao, J.; Hall, K. R.; Marsh, K. N.; Wilhoit, R. C. *Ibid.* "Part 2. Ideal Gas Properties". *J. Phys. Chem. Ref. Data* 1986, 15, 1369.
- (4) *Bulletin of Chemical Thermodynamics*; Freeman, Robert D., Ed.; IUPAC Commission on Thermodynamics (1.2), Thermochemistry, Inc.: Stillwater, OK, issued annually 1960–1985.
- (5) *International Data Series Selected Data on Mixtures. Series A. Thermodynamic Properties of Non-Reacting Binary Systems of Organic Substances*; Kehiaian, Henry, Editor-in-Chief; Thermodynamics Research Center, Texas A&M University: College Station, TX.
- (6) Examples: 20 monographs on compounds of importance in fuel technology, in the *API Monograph Series*, No. 705–724; Publications Di-

- vision, American Petroleum Institute: Washington, DC. Physical Properties, Section 23, *Engineering Data Book*, Vol. II, 10th ed.; Gas Processors Association: Tulsa, OK, 1987.
- (7) Colclough, A. R. "Two Theories of Experimental Error". *J. Res. Natl. Bur. Stand.* **1987**, 92, 167-185.
- (8) Angus, S. *Guide for the Preparation of Thermodynamic Tables and Correlations of the Fluid State*; IUPAC Thermodynamic Tables Project Center, Imperial College of Science and Technology: London, 1983. *Fluids*; IUPAC Thermodynamic Tables Project Center, Imperial College of Science and Technology: London.
- (9) SIR, Inc., a division of InterSystems, Inc., 707 Lake Cook Road, Suite 120, Deerfield, IL 60015.
- (10) Numerica Information Service, Technical Database Services, Inc., 10 Columbus Circle, New York, NY 10019.
- (11) A copy of the following document is available from the authors: Wilhoit, R. C.; Maczynski, A. "Rules for Preparing a COSTAT Message for Transmitting Thermodynamic Data. CODATA Task Group on Critically Evaluated Phase Equilibrium Data"; 15 pp.

Formulation of Isomeric Reaction Types and Systematic Enumeration of Six-Electron Pericyclic Reactions

SHINSAKU FUJITA

Research Laboratories, Ashigara, Fuji Photo Film Co., Ltd., Minami-Ashigara, Kanagawa, 250-01 Japan

Received January 20, 1988

Reaction-center graphs (RCG), reaction graphs (RG), and basic reaction graphs (BRG), all of which are defined as substructures of imaginary transition structures (ITS), represent reaction types of various levels in ascending order of generality. The concept of isomeric RCGs is introduced for the systematic enumeration of the reaction types. The enumeration of six-electron pericyclic reactions is thereby translated into the counting of hexagonal and pentagonal RCGs. This issue is solved by regarding the hexagonal or pentagonal RCG as a derivative of a parent RG and by using Pólya's theorem generalized to be applicable to the cases in which the vertices are obliged to have different modes of substitution.

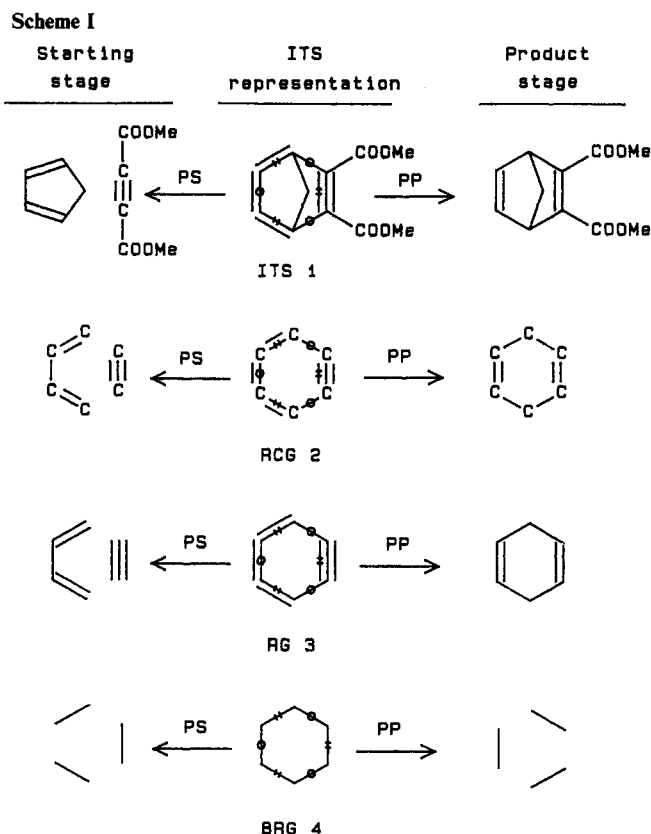
The systematic enumeration of reaction types is fundamentally important in providing organic chemistry with a logical format as well as in generating new possible reactions theoretically.¹ However, few approaches have succeeded, mainly because of three related factors. The first one is the absence of consistent definitions of reaction types.^{2,3} The second factor is the deficiency of the strategy that realizes complete enumeration. The third factor is the lack of mathematical methods of verification.

In previous papers,^{4b,c} we have presented the concepts of *reaction graph* (RG) and *reaction-center graph* (RCG) with respect to the first factor. As for the second and third factors, we have already discussed the enumeration of the RGs. This enumeration, however, has considered only the modes of bond switching and not the variety of reacting atoms. The latter issue requires the enumeration of the RCGs, which will be discussed in the present paper.

FORMULATION OF REACTION TYPES

The unequivocal definitions of reaction types is necessary to the counting of them. Scheme I summarizes the representation, already introduced in the previous papers of this series,⁵ based on the concept of *imaginary transition structure* (ITS). ITS 1 is obtained by superposing the left-hand side onto the right-hand side of the corresponding reaction diagram and by distinguishing the bonds into out- (—/—), in- (—○—), and par-bonds (—).

Each bond of a given ITS is called an *imaginary bond* (or an ITS bond) and consists of out-, in-, and/or par-bonds.⁶ If we consider the imaginary bond as a figure, we can interpret the figure in various ways. In a chemical sense, the imaginary bond \equiv represents an initial single bond that changes into a double bond. From another point of view (i.e., apart from a chemical meaning tentatively), we regard the imaginary bond also as a combination of a par-bond and an in-bond or as an in-bond modified by a single par-bond. Similarly, the imag-



inary bond \equiv , which chemically indicates a triple bond converted into a double bond, is considered to be a combination of two par-bonds and one out-bond as well as an out-bond (—/—) modified by a double par-bond (\equiv). These flexible modes of interpretation open the enumeration to a new treatment.