

# FATE, the Environmental Fate Constants Information Database

Heinz P. Kollig\* and Brenda E. Kitchens

Environmental Research Laboratory, U.S. Environmental Protection Agency, Athens, Georgia 30613-0801

Karen J. Hamrick

Computer Sciences Corporation, c/o U.S. Environmental Protection Agency, Athens, Georgia 30613-0801

Received June 30, 1992

An online database, FATE, has been developed for the interactive retrieval of kinetic and equilibrium constants that are needed for assessing the fate of chemicals in the environment. The database contains values for up to 12 parameters for each chemical. As of December 1991, FATE contained values for about 200 chemicals. FATE is unique in that experimental data are extracted only from primary references. Pertinent experimental conditions are included in the database to support the credibility and applicability of a value. Transformation products are included when they are available. A new computer program is used to extrapolate measured hydrolysis rate constant data to a standard format. Acidic, basic, and neutral contributions are combined to calculate the overall hydrolysis rate constant,  $k_h$ , and the half-life of the chemical at 25 °C and pH 7. The data are reported as second-order acidic and basic rates and a first-order neutral rate at 25 °C. A new computerized expert system will be applied to compute fate constant values that are more reliable than many measured values. The expert system has the capability of crossing chemical boundaries to cover all organic compounds.

## INTRODUCTION

The fate of chemicals in the environment is determined by physical, chemical, and biological processes that include advective transport, dispersion, sorption, volatilization, hydrolysis, oxidation, phototransformation, aerobic and anaerobic biotransformation, and bioconcentration. Kinetic and equilibrium constants that have been determined for these processes are used as input into mathematical models that estimate the transport and transformation of chemicals. This information subsequently is used to assess the risks posed by chemicals to ecosystems and human health and to manage their release and disposal.

The need for rate and equilibrium (i.e., fate) constants for chemicals that have potential environmental impact has grown in tandem with the production of new chemicals by industry. Considering the total number of chemicals under production, relatively few of the required fate constants have been measured experimentally and published, and many of the published fate constants are of questionable reliability or applicability. It is imperative that data that are used for ecological and human risk assessments for regulatory purposes be reliable for the risk assessments to have validity.

Availability and validity of fate constants obtained from scientific journals vary, sometimes considerably.<sup>1</sup> It is common to locate several values for a given process constant for a specific chemical and to find no values for hundreds of other chemicals for which data are needed. Data evaluation systems have been suggested<sup>2-4</sup> as a mechanism to assist in deciding which one of many values might be the most reliable. However, the reliability of a fate constant value in a primary reference can be evaluated only by examining the research protocol and experimental conditions provided by the author. (A primary reference is defined as the source who presented the original research.)

The inherent complexity in measuring fate constants makes the measurement process difficult and expensive. The need for data will never be completely satisfied through laboratory measurements because of the time and expense involved.

Therefore, reliable computational techniques are needed to estimate physical and chemical process constants. Computational techniques can be used to generate values rapidly at a fraction of the costs required to measure them experimentally.

FATE was developed as an online database that provides users with reliable and environmentally realistic fate constants. The quality of the data is enhanced by (a) applying objective screening criteria to determine whether a value should be entered into the database; (b) entering experimentally determined literature values from primary sources only, along with pertinent experimental conditions and reaction products; and (c) entering into the database computed values obtained by estimation techniques that use fundamental chemical structure theory incorporated in the expert system SPARC (SPARC Performs Automated Reasoning in Chemistry)<sup>5</sup> and conventional techniques based on property-reactivity correlations in the QSAR (quantitative structure-activity relationships)<sup>6</sup> system.

## SOURCES AND LIMITATIONS OF DATA

**Experimental Data.** Numerous computerized databases are being developed to provide experimental fate constant data that are more accessible to the scientific community than those obtained through literature searches. As data for these databases have been compiled or used, it has become obvious that there are problems with either too little data for specific chemicals or too many data that do not corroborate other measurements for the same chemical. To reliably extract measured kinetic and equilibrium constants from the scientific literature, extensive knowledge is required of the processes under investigation and the conditions under which the investigations were conducted.<sup>4,7</sup>

Decades of experiments have been conducted and reported using investigative criteria that were different from the stringent criteria that are required for risk assessments today. Reported aqueous solubility values, partition coefficients, and other measured parameters contained in the literature may range over several orders of magnitude.<sup>1,8</sup> For example, a

**Table I.** Experimental Octanol–Water Partition Coefficients for *p,p'*-DDT and Dieldrin from Literature Survey

| chemical name    | $K_{ow}$          | chemical name | $K_{ow}$          |
|------------------|-------------------|---------------|-------------------|
| <i>p,p'</i> -DDT | $9.5 \times 10^3$ | dieldrin      | $1.2 \times 10^4$ |
|                  | $1.2 \times 10^5$ |               | $2.5 \times 10^5$ |
|                  | $8.2 \times 10^6$ |               | $1.6 \times 10^6$ |

literature survey for the octanol–water partition coefficients for *p,p'*-DDT and dieldrin (Table I) shows values from three publications ranging from  $9.5 \times 10^3$  to  $8.2 \times 10^6$  for DDT and from  $1.2 \times 10^4$  to  $1.6 \times 10^6$  for dieldrin.

Literature documentation of experimental conditions occasionally may include vague phrases such as 'distilled water' when a pH value is now required to extrapolate the data, or 'room temperature' instead of a precise temperature. Frequently, experimental conditions cannot be extrapolated to environmental conditions without considerable mathematical manipulation. In the repeated citation of data from other-than-primary sources, there is always the possibility that data were miscalculated; transposed; or published with signs missing, decimal point placed incorrectly, or any number of other problems. It is important, therefore, to obtain fate constant data from primary sources. It is equally important to be consistent in the use of exponents and the number of significant figures when reporting the data to avoid misinterpretation.

An example of a publication error, demonstrating the importance of obtaining values from primary sources, was found with the octanol–water partition coefficient,  $K_{ow}$ , for acenaphthene. A secondary publication was found to have a log value of  $-2.02$ , with a citation included for the primary reference. The value of  $-2.02$  was viewed as questionable because of our knowledge of the  $K_{ow}$  values of structurally similar compounds. The primary source was located and found to contain a value of  $3.98$  for the log  $K_{ow}$  of acenaphthene. Clearly, a user without knowledge of the process could have used a value that was off by 6 orders of magnitude.

In many situations, a user requiring data for risk assessment modeling would prefer to be given one "best estimate" fate constant value for each chemical because resources and expertise may not be available for determining the most accurate value. Although accuracy cannot be determined by literature review, the thoroughness of the documentation of a report gives some indication of the confidence with which a reported value can be used in an environmental risk assessment.<sup>4</sup> For these reasons, the goal of the FATE database is to provide measured data from primary references that are applicable to environmental risk assessments, including as much information as possible about the experimental protocol.

**Estimated Data.** During the past several years, extensive research has been conducted to develop methods for estimating fate constants because of the difficulty and expense of laboratory measurements.<sup>9</sup> The largest compilation of property estimation methods using structure–activity correlations has been published by Lyman et al.<sup>10</sup> Often these relationships are applicable only within the limitations of the family of chemicals that was used to develop the correlation coefficients. For other chemicals, the estimated values may be off by 1 or more orders of magnitude. Therefore, the evaluation of data must be done with an appreciation of the range and limitations of the estimation techniques.

QSAR is an interactive chemical database and hazard assessment system designed to provide basic information for the evaluation of the fate and effects of chemicals in the environment. The QSAR system is a composite of databases

**Table II.** Fate Parameters with Symbols Used for FATE Database

| symbol             | fate parameter   |
|--------------------|--|
| $H_c$              | Henry's law constant   |
| $k_h$              | hydrolysis rate constant                                       |
| $pK_a$             | ionization constant  |
| $K_{ow}$           | octanol–water partition coefficient                            |
| $K_{oc}$           | organic carbon normalized sediment–water partition coefficient |
| $k_d$              | direct (aqueous) photolysis rate constant                      |
| $K_p$              | sediment–water partition coefficient                           |
| $P_v$              | vapor pressure   |
| $S_w$              | water solubility   |
| $\epsilon_\lambda$ | molar absorptivity   |
| $\phi_r$           | aqueous photolysis reaction quantum yield                      |
| $k_{bio}$          | biodegradation rate constant                                   |

containing measured values obtained from the literature and a state-of-the-art QSAR (quantitative structure–activity relationships) model library capable of estimating chemical properties, behavior, and toxicity, based on conventional estimation techniques derived from the work of Lyman.<sup>10</sup> The QSAR system contains automated estimation routines for 13 properties that have been modified from structure–activity correlations. Data may be retrieved by entering a CAS Registry Number or SMILES (simplified molecular identification and line entry system)<sup>11,12</sup> notation for most chemicals.

The expert system SPARC is being developed by collaborating scientists at the U.S. EPA's Environmental Research Laboratory—Athens and the University of Georgia. This system uses algorithms based on fundamental chemical structure theory and allows estimation of values for any parameter that depends on molecular structure. Unlike methods based solely on property–reactivity correlations, this system combines the principles of quantitative structure–activity relationships, linear free energy theory, and perturbation theory from quantum chemistry to estimate process constants for all organic chemicals, without limitations. Although not fully developed, SPARC already computes values for many chemicals that are more reliable than experimentally determined values. SPARC will eventually contain estimation routines for most of the 12 parameters that are included in the FATE database.

## FATE DATABASE

The FATE database has been developed with FOCUS,<sup>13</sup> a fourth-generation language, as a menu-driven, read-only captive system. It is installed on the VAX cluster of the Environmental Research Laboratory—Athens and is available for on-line use by the public.

Currently, FATE contains published fate constant data that have been extracted from primary references for approximately 200 chemicals. FATE also contains selected fate constants derived from computational techniques. We believe that the addition of data computed with SPARC will expand the number of values in FATE rapidly.

**Database Files.** The FATE database consists of three data files. The CAS File contains CAS Registry Numbers (Chemical Abstract Service), molecular formulas, SMILES notations, and chemical and common names. The REF File contains reference numbers and complete citations. The FATES File is cross-referenced to the other files and contains the data for 12 fate parameters (Table II).

**Methods of Data Retrieval.** Interactive users of the FATE database system may query the database by CAS Registry Number, by SMILES notation, by molecular formula, by a substring of the preferred chemical name, by a substring of

```

Fate Data, References as of 06/23/92 Page 1
PF7 to scroll backward RETURN to go forward PF3 to abort report
-----
CAS Number: 79345 FATE Reference: C00000165
CAS Prefix:
CAS Name : Ethane, 1,1,2,2-tetrachloro-

SMILES : ClC(Cl)C(Cl)Cl

FATE Code : 05, hydrolysis rate constant
Analytical Method: GLC Estimating Program:
Medium: buffered dist. H2O pH: see comments
Experimental Temperature: 95.00 C
Products: [79-01-6]
Comments: 1st order rates were meas. over the pH range 5 to 9
          at 11 temp. Data were extrapolated to 1st- and 2nd-
          order rates at 25 C with the RATE program. Ea(base)
          was estimated as 21.2 kcal/mol or 88.8 kJ/mol.

-----Fate Values-----
half-life : 98 day (25 C)
k(acid) : 0.0/M-yr (25 C)
k(base) : 2.6E7/M-yr (25 C)
k(h), pH 7 : 2.6/yr (25 C)
k(neutral) : 0.0/yr (25 C)

FATE Reference: C00000165
Authors: Cooper, William J.; Mehran, Mostafa;
         Rlusech, David J.; Joens, Jeffrey A.

Date: 1987
Title: Abiotic transformation of halogenated organics. 1.
       Elimination reaction of 1,1,2,2-tetrachloroethane
       and formation of 1,1,2-trichloroethene.

Citation: Environ. Sci. Technol. 21(11):1112-1114.

```

Figure 1. Sample report of hydrolysis for 1,1,2,2-tetrachloroethane, CAS No. 79-34-5.

the common names, by the reference number, or by primary or secondary authors from the reference publications.

**RATE Program.** Measured hydrolysis data that are entered in the FATE database are analyzed initially with RATE,<sup>14</sup> a FORTRAN program that was developed at the Environmental Research Laboratory—Athens. RATE is used to extrapolate data to a standard format of second-order acidic and basic rates and a first-order neutral rate at 25 °C. The acidic, basic, and neutral contributions are combined to calculate the overall hydrolysis rate constant,  $k_h$ , and the half-life of the chemical at 25 °C and pH 7. Use of the RATE program has been described in a previous publication.<sup>14</sup>

#### DATA RETRIEVAL FROM FATE

The FATE database can be accessed by any user requesting an account and a password. Initial contact with the database is established by logging onto the restricted FATE account on the Athens VAX. Reports from FATE can be captured and printed from the personal computer of the user.

Figure 1 provides a sample fate data and reference report for the hydrolysis rate constant,  $k_h$ , for 1,1,2,2-tetrachloroethane, CAS Number 79-34-5. Currently, the FATE database contains data for hydrolysis from two references, data for the octanol–water partition coefficient from two references, and one value for the water solubility of 1,1,2,2-tetrachloroethane.

#### DISCUSSION

The development of the FATE database evolved from the need for fate constants to be used for chemical risk assessment. Literature searches for measured data revealed that few values were available, that many publications lacked sufficient documentation to determine data credibility, and that many data were determined under environmentally unrealistic conditions. These factors made the extrapolation of data to the environmental conditions required by environmental models for fate assessment difficult or impossible. More importantly, few studies documented the products of the degradation processes. Unfortunately, environmental risk assessment must account for the “persistent” chemicals and not merely the parent compound or intermediate products of a reaction process.<sup>16</sup> Moreover, fate constants being used in

the regulatory process for risk assessments must be reliable and applicable.

The reliability and applicability of literature data can be assessed by obtaining primary references only and evaluating the entire experimental process described in the publication in an objective manner. Currently, we evaluate publications using a set of reasonably objective evaluation criteria for 12 fate parameters that were published by Kollig.<sup>4</sup> These criteria also include corroborative information. The author who determined the data in the laboratory is actually the only one who can assure the reliability, whereas the applicability may be determined from the experimental information given in the publication. Knowledge of the process is a must. We can only evaluate what is detailed in the publication. If information is not provided, the evaluator must consider the data to be of undemonstrated reliability or applicability, even though the work in question may have been carried out in a scientifically acceptable manner. Finally, a call to the author may be necessary to clarify any doubts.

The quickest and least expensive way of obtaining environmental fate constants is through accessing a database. Most database managers extract data from the literature (primary and secondary). In many cases, this is done by junior personnel to keep costs low. The database furnishes the user what the literature reports, without regard to environmental reliability and applicability. Unfortunately, the literature contains few environmental fate constants. This is primarily because many of them, especially kinetic constants, are difficult to determine and the process is costly. We have estimated that the average cost of determining a fate constant is about ten thousand dollars. With more than ten million chemicals in existence, the cost becomes unbearable, and many years would be required to complete the task.

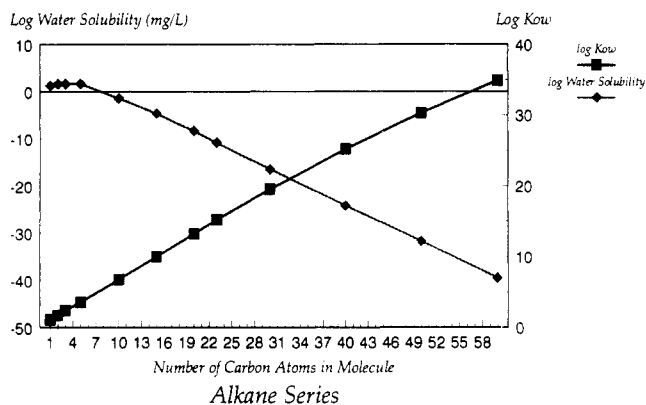
The solution to this dilemma is computational techniques. We have estimated that the average cost of computing a fate constant is about \$10, and many can be computed in a short period of time. Table III shows the highest and lowest literature values compiled by Mackay et al.<sup>17</sup> for four fate parameters of five chemicals from high to low hydrophobicity. It is obvious that the ranges, generally, increase with hydrophobicity. Analytically, it becomes more difficult to determine an accurate value, and at the extremes, we cannot determine one at all. Corroboration becomes an important factor, and computational techniques should be able to estimate an acceptable value.

SPARC has been tested with the alkane series up to 60 carbon atoms (hexacontane). This series should exhibit a regression curve for fate parameters with increasing carbon atoms in the molecule. Figure 2 shows the water solubility and the log of the octanol–water partition coefficient ( $K_{ow}$ ) values computed by SPARC for this series. A literature survey for fate values revealed fewer data with increasing carbon atoms. As carbon atoms increased, so did the ranges of values for these fate data, sometimes by several orders of magnitude. Some values decreased at times rather than increasing as the molecule increased in size. No literature values were found for solubility and  $K_{ow}$  from triacontane ( $C_{30}H_{62}$ ) to hexacontane ( $C_{60}H_{122}$ ). These factors prevented us from plotting literature data with SPARC data as a comparison. However, as Figure 2 shows, SPARC data indicate smooth regression curves, which makes us believe that these data are more reliable than literature data, especially in the area of high hydrophobicity.

A support activity was initiated at the Environmental Research Laboratory—Athens to provide kinetic and equi-

**Table III.** Ranges of Literature Data of Four Fate Parameters for Five Chemicals with High to Low Hydrophobicity from Mackay et al.<sup>17</sup>

| chemical   | water solubility, mg/L   | vapor pressure, Pa  | Henry's constant, Pa·m <sup>3</sup> /mol                        | log <i>K</i> <sub>ow</sub>  |
|--|--|---|---|-----------------------------|
| decachlorobiphenyl<br>PCB-209<br>CAS No. 2051-24-3   | 7.2 × 10 <sup>-2</sup> –2.12 × 10 <sup>-8</sup><br><i>n</i> = 24 | 4 × 10 <sup>-5</sup> –2.2 × 10 <sup>-8</sup><br><i>n</i> = 18     | 12.46–100<br><i>n</i> = 3                                       | 5.28–11.20<br><i>n</i> = 35 |
| pentachlorobiphenyl<br>PCB-101<br>CAS No. 37680-72-3 | 5.94 × 10 <sup>-2</sup> –5 × 10 <sup>-4</sup><br><i>n</i> = 34   | 4.03 × 10 <sup>-3</sup> –5.27 × 10 <sup>-4</sup><br><i>n</i> = 26 | 7.09–217<br><i>n</i> = 8  | 4.12–7.64<br><i>n</i> = 43  |
| 2-chlorobiphenyl<br>PCB-1<br>CAS No. 2051-60-7       | 0.685–49.6<br><i>n</i> = 24                                      | 0.260–2.50<br><i>n</i> = 17                                       | 28.9–74.6<br><i>n</i> = 3                                       | 3.75–5.7<br><i>n</i> = 33   |
| 1,3-dichlorobenzene<br>CAS No. 541-73-1              | 4.81 × 10 <sup>1</sup> –1.50 × 10 <sup>2</sup><br><i>n</i> = 41  | 2.43 × 10 <sup>2</sup> –3.89 × 10 <sup>2</sup><br><i>n</i> = 11   | 1.19 × 10 <sup>2</sup> –4.72 × 10 <sup>2</sup><br><i>n</i> = 12 | 3.13–3.95<br><i>n</i> = 48  |
| benzene<br>CAS No. 71-43-2                           | 6.65 × 10 <sup>2</sup> –4.82 × 10 <sup>3</sup><br><i>n</i> = 78  | 3.75 × 10 <sup>3</sup> –1.32 × 10 <sup>4</sup><br><i>n</i> = 11   | 4.41 × 10 <sup>2</sup> –7.40 × 10 <sup>2</sup><br><i>n</i> = 31 | 1.56–2.69<br><i>n</i> = 84  |

**Figure 2.** Regression of log water solubility and log *K*<sub>ow</sub> values for the alkane series computed by SPARC.

librium constants and their transformation products for the assessment of environmental transport and transformation of high-priority chemicals. This activity involves literature searches for measured data and postulating transformation pathways and expected products. When measured data are not available, laboratory measurements of selected fate constants are initiated, or computational techniques are applied as another alternative for data.

The FATE database was developed to eliminate many of the problems that scientists involved in this support activity experienced when they attempted to obtain kinetic and equilibrium constants from the literature. As a result, the FATE database contains values that can be used with confidence for up to 12 parameters. Data have been screened for applicability to environmental assessment, and only literature data from primary sources have been entered. Transformation products have been included when they have been identified in the publications. Chemical hydrolysis rate constants have been extrapolated to a standard format with a computer program that determines the overall hydrolysis rate constant, *k*<sub>h</sub>, the half-life of the chemical at 25 °C and pH 7, the second-order acidic and basic rates, and a first-order neutral rate at 25 °C. In addition, the FATE database contains values that have been computed with the expert system SPARC and the QSAR system for selected fate constants.

Future emphasis with the FATE database will be to enter data computed with SPARC as the expert system is developed

for additional parameters. The expert system has the potential capability of crossing chemical boundaries to provide estimates for selected fate constants for all organic chemicals and will generate reliable values for a fraction of the expense and time involved in generating measured values.

## REFERENCES AND NOTES

- (1) Bowman, B. T.; Sans, W. W. Determination of octanol–water partition coefficients (*K*<sub>ow</sub>) of 61 organophosphorus and carbamate insecticides and their relationship to respective water solubility (*S*) values. *J. Environ. Sci. Health* **1983**, *B18* (6), 667–683.
- (2) Howard, P. H.; Sage, G. W. Development of indicators of data documentation quality for environmental measurements data. Developed by the Syracuse Research Corp., New York, for the Chemical Manufacturer's Assoc., Washington, D.C., under Contract 010114.185; 1986; SRC F0089-01.
- (3) Creeger, S. M.; Whetzel, N. K.; Fletcher, C. L. Standard evaluation procedures: Hydrolysis, aqueous photolysis, aerobic soil metabolism, soil photolysis, soil column leaching. U.S. Environmental Protection Agency: Washington, D.C., 1985. EPA 540/9-85-013, -014, -015, -016, -017.
- (4) Kollig, H. P. Criteria for evaluating the reliability of literature data on environmental process constants. *Toxicol. Environ. Chem.* **1988**, *17* (4), 287–311.
- (5) Karickhoff, S. W.; Carreira, L. A.; Melton, C.; McDaniel, V. K.; Vellino, A. N.; Nute, D. E. Computer prediction of chemical reactivity—the ultimate SAR. U.S. Environmental Protection Agency: Athens, GA, 1989; EPA/600/M-89/107.
- (6) QSAR was developed jointly by the U.S. EPA Environmental Research Laboratory—Duluth, Montana State University Center for Data Systems and Analysis, and the Pomona College Medicinal Chemistry Project.
- (7) Mill, T.; Walton, B. T. How reliable are database data? *Environ. Toxicol. Chem.* **1987**, *6*, 161–162.
- (8) Chiou, C. T.; Freed, V. H.; Schmedding, D. W.; Kohnert, R. L. Partition coefficient and bioaccumulation of selected organic chemicals. *Environ. Sci. Technol.* **1977**, *11* (5), 475–478.
- (9) Heller, S. R. Property prediction. *Ind. Chem.* **1988**, July.
- (10) Lyman, W. J.; Reehl, W. F.; Rosenblatt, D. H. *Handbook of Chemical Property Estimation Methods: Environmental Behavior of Organic Compounds*; McGraw-Hill Book Co., New York, 1982.
- (11) Anderson, E.; Veith, G. D.; Weininger, D. SMILES: a line notation and computerized interpreter for chemical structures. U.S. Environmental Protection Agency: Duluth, MN, 1987; EPA/600/M-87-021.
- (12) Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36.
- (13) FOCUS was developed by Information Builders, Inc. Copyright, 1987.
- (14) Hamrick, K. J.; Kollig, H. P.; Bartell, B. A. Computerized extrapolation of hydrolysis rate data. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 511–514.
- (15) McMurry, J. *Organic Chemistry*; Brooks/Cole Publishing Co.: Belmont, CA, 1984; p 112.
- (16) Kollig, H. P. A fate constant data program. *Toxicol. Environ. Chem.* **1990**, *25* (2–3), 171–179.
- (17) Mackay, D.; Shiu, W. Y.; Ma, K. C. *Illustrated Handbook of Physical-Chemical Properties and Environmental Fate for Organic Chemicals*; Lewis Publishers, Inc.: Chelsea, MI, 1992; Vol. 1.