Environmental Health Perspectives
Vol. 96, pp. 57-60, 1991

A Fact Database for Toxicological Data at the National Institute of Hygienic Sciences, Japan

by Makoto Hayashi,* Masahiro Nakadate,* Takaharu Osada,† Tamotsu Ishibe,† Satoru Tanaka,* Akihiko Maekawa,* Toshio Sofuni,* Yasuko Nakata,* Naoko Kanoh,* Shigeko Hashiba,* Yusuke Takenaka,* and Motoi Ishidate, Jr.*

The computerized fact database for the toxicity data of chemicals was constructed at the National Institute of Hygienic Sciences, Tokyo, Japan (biological database, BL-DB). The BL-DB stores data on mutagenicity, teratogenicity, carcinogenicity, and other toxicological tests of chemicals that appeared in the scientific literature. The BL-DB includes information about chemical identification, test system, results of the assays, and a bibliography. The system consists of five modules: data collection, data maintenance, data search, data downloading, and backup. ADABAS is used as a core database management system. Many kinds of test data are stored with the same formats; therefore, users can retrieve data of different toxicological data by the same manner. A user of the BL-DB can use about 50 kinds of commands to interact with the system, and the majority of fields are defined as search fields, thereby facilitating retrieval of target data through many ways. Currently, there are mainly data for the mutagenicity, especially on the Salmonella/microsome assay and the rodent micronucleus assay. These data can be retrieved and used for structure–activity relationship studies.

Introduction

There are about 70,000 kinds of chemicals surrounding us in the industrialized countries, and a large number of chemicals are introduced into our environment every year. These chemicals have played important roles in developing our lifestyle; on the other hand, it is known that some of these chemicals are hazardous to human health. It is therefore indispensable to understand the toxicity of chemicals not only taken directly into our bodies (e.g., pharmaceutical drugs, food additives, and pesticides), but also the toxicity of any other chemicals to which we might be exposed. In the last several decades, many kinds of toxicological tests have been performed on many chemicals. The realization of the usefulness of this information for chemical safety made it desirable to establish a fact database to collect data that can easily be retrieved.

The fact database for information about toxicity of chemicals was constructed at the National Institute of Hygienic Sciences, Tokyo, Japan (biological database, BL-DB). This database stores test results on mutagenicity, teratogenicity, carcinogenicity, and other toxicological tests of chemicals. Data include chemical identification, explanations about test systems, results of the test (results in tabular form can also be stored), and bibliographic information.

Architecture of the BL-DB

The BL-DB system consists mainly of five modules (Fig. 1): data collection, data maintenance, a data search, a data downloading, and backup. ADABAS is used as a core database management system. The functions and characteristics of modules are as follows: *a*) data collection module: data abstracted from original articles and other sources are input to the working file using data format sheets. This module can collect data from machine-readable media or from direct key in and can also send data to the BL-DB after the formulation of data aided by the data maintenance module. *b*) Data maintenance module: this module formulates data to fit into the main database, BL-DB. It also outputs checklists of working files to update data to satisfy the requirements and restrictions of the BL-DB. *c*) Data search module: this module enables data retrieval by tag keys for name and characteristics of chemicals, test systems (experimental

*Biological Safety Research Center, National Institute of Hygienic Sciences, 1-18-1, Kamiyoga, Setagaya-ku, Tokyo 158, Japan.
†Japan Institute of Synthetic Technology, Hanzomon-Heim, 1-8, Koji-machi, Chiyoda-ku, Toyko 102, Japan.
Address reprint requests to M. Hayashi, Biological Safety Research Center, National Institute of Hygienic Sciences, 1-18-1, Kamiyoga, Setagaya-ku, Tokyo 158, Japan.

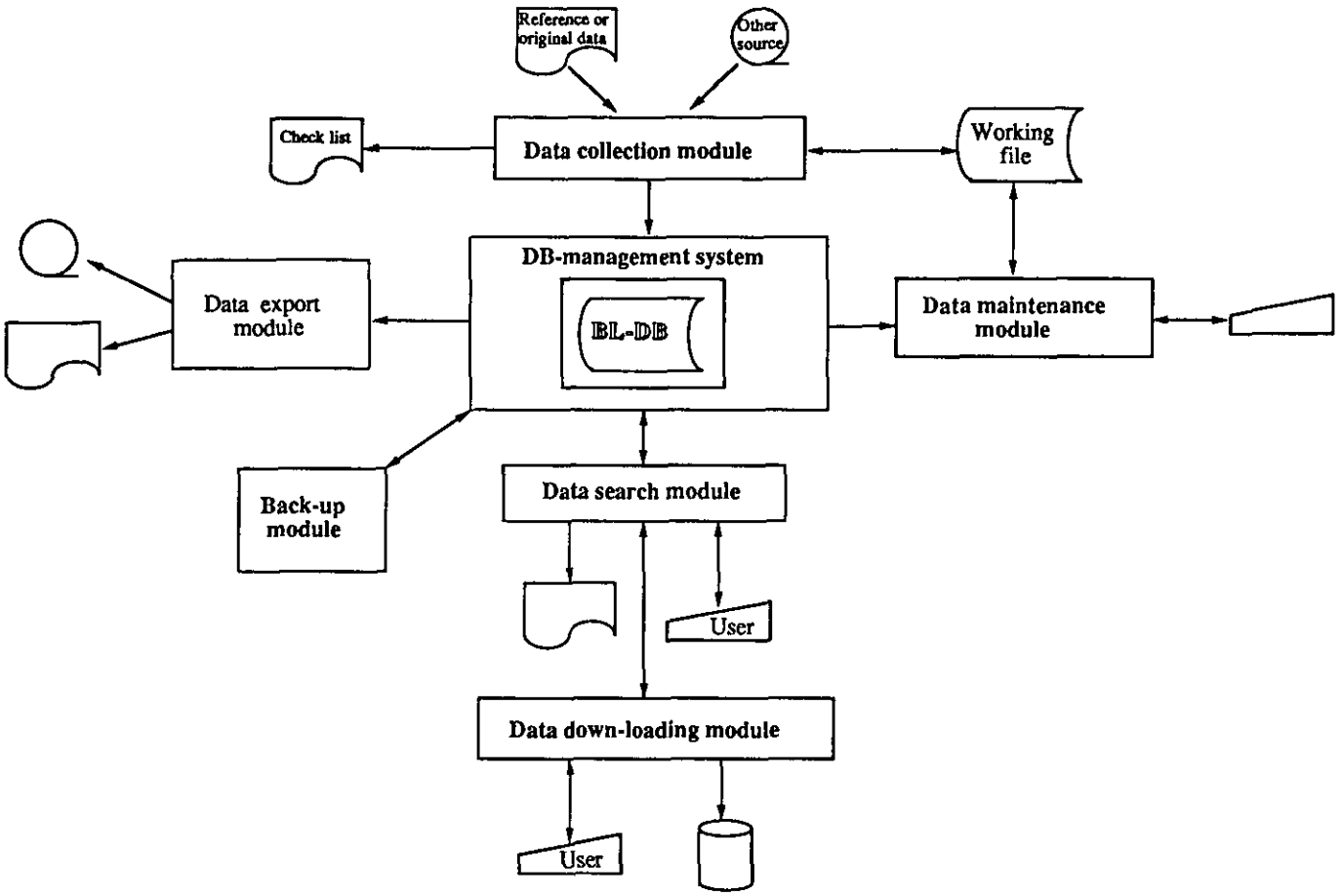

FIGURE 1. Architecture of the biological database.

materials, results, etc.), and also bibliographical information. *d*) Data downloading module: this module downloads data retrieved to conventional sequential file, which can be used for further modifications. *e*) Back-up module: makes a back-up file and recovers it when system is damaged. These modules are assembled logically to realize a total fact database system.

File Configuration

The BL-DB database has three main files, a substance file, a test file, and a bibliographic file, which are linked to each other by keys as shown in Figure 2. A record of a substance file is linked

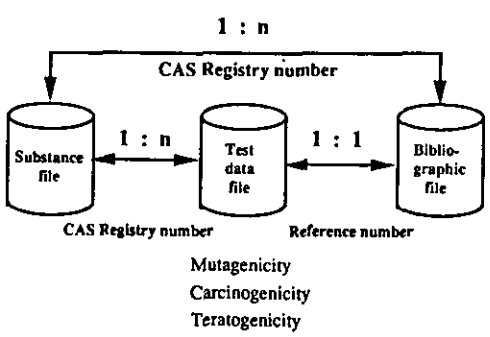

FIGURE 2. File configuration of the biological database.

linked to several records of a test file by Chemical Abstract Service Registry Number (CAS number). A bibliographic record is linked to several test records by a reference number, which serves as a link key. Each chemical is also linked to several bibliographic records by a CAS number.

Field Configuration

Each file consists of fields that are retrieval units in the BL-DB as described below.

Substance File. The substance file consists of the following fields:

1. Chemical name (S0): a common name that appears in the reference.
2. CAS number (S1): the number that identifies the chemical. It is a key tag linking a chemical and test data files and also linking a chemical and reference files. The replaced CAS number(s) is also stored and records can be accessed through the old number(s).
3. Registry number of BL-DB (S2): the identity number of the BL-DB.
4. Registry number of Registry of Toxic Effects of Chemical Substances (RTECS) (S3): the registration number of the U.S. National Institute for Occupational Safety and Health.
5. Chemical name (S4): in the BL-DB, the Chemical Ab-

stracts 9th Collective Index Name (9CI) is used as the formal name of a chemical substance.

6. Synonyms (S5): IUPAC name, a common abbreviation, a trade name, Chemical Abstracts 8th Collective Index name, a color index name and number, etc.
7. Chemical structure (S6): Wiswesser Line-Formula Chemical Notation (WLN) for each chemical is stored to serve in the study of structure–activity relation.
8. Molecular formula (S7): Hill method. Molecular weight is calculated by the system.
9. Use of chemical (S8): drug, food additive, pesticide, etc.
10. Classification of chemical (S9).
11. Toxic dose (SA).
12. Melting point (with unit) (SB).
13. Boiling point (with unit) (SC).
14. Related CAS number (SD): CAS number(s) of free form of a chemical.
15. Comments (SE).

Test Data File. The test data file contains the following files:

1. Chemical name (D0): if the test article is a mixture, several names of chemicals can be stored in this field separated by three spaces.
2. CAS number (D1): link key to the substance file.
3. NIHS number (D2).
4. Purity (D3): purity, maker, lot number, and other information about the chemical used in the experiment.
5. Experiment system-1 (D4): global classification of test systems, e.g., mutagenicity test, reproduction test, carcinogenicity test.
6. Experiment system-2 (D5): precise classification of test systems, e.g., gene mutation test, micronucleus test, teratogenicity test.
7. Material-0 (D6): global classification of an experimental material, e.g., bacteria, yeast or fungi, insects, plants, mammalian cells *in vitro*, mammals *in vivo*.
8. Material-1 (D7): common name or scientific name of experimental materials, e.g., *Salmonella typhimurium*, *Allium cepa*, mouse, rat, dog.
9. Material-2 (D8): the strain of an experimental material, e.g., TA100, TA98, ICR, F344.
10. Material-3 (D9): characteristics of experimental materials and/or target tissues of experiment, if defined, e.g., streptomycin resistance, bone marrow cells, peripheral blood, etc.
11. Age of animals (with unit) (DA): when animals were used as experimental materials.
12. Body weight (with unit) (DB).
13. Route of administration (DC): when animals are used as experimental materials, e.g., oral or po, intubation, intraperitoneal, or ip, inhalation, or ih.
14. Method of experiment (DD): brief description of test methods.
15. Solvent (DE): solvent or vehicle used in the experiment.
16. Dose or concentration (with unit) (DF): individual or range of dose levels tested.
17. Final judgment (DJ): positive or negative. Minimum effective or maximum noneffective dose and also symptoms and other observations are described.
18. Grade (DK): the result of evaluation of the data by some formal organizations or a rough qualitative grade of the experiment (subjective).
19. Statistical method (DL): the statistical method used to evaluate the test data, if available.
20. Comments (DM).
21. Subreference number (DN): equal to the lowest record number that appeared from the reference.
22. Reference number (DN): ISSN number, year of publication, volume number, and first page, e.g., 0123-4S67/91/0012-0345.
23. Test data in tabular form (DG, DH): test data can be stored in tabular form in this file.

Bibliographic File. The bibliographic file contains the following fields.

1. Reference number (R1): Same reference number as in test data file.
2. Author(s) (R2).
3. Title of the article (R3).
4. Subtitle of the article (R4).
5. Journal name (R5).
6. CODEN (R6): an acronym for the references.
7. Japan Information Center of Science and Technology (JICST) code (R6).
8. Year, volume, number, pages (R8).
9. Language (R9).
10. Institute (RA): raw data can also be saved in this system. In such a case, (RA), (RB), and (RC) should be defined.
11. Name(s) (RB): name(s) of investigator.
12. Date studied (RC).
13. Key word(s) (RD): the words that appear in the title of the article are automatically cited as key words. Any other key words can be saved.
14. Abstracter(s) (RE).
15. Comments (RF).

Search Methods

A user of the BL-DB can use commands to interact with the system. About 50 kinds of commands are ready for use. In Figure 3 an example of retrieval result is shown. Files can be opened and closed by commands and a user can move to another file with results previously retrieved. A user may search data with a search tag together with wild card(s). The search can be made using numerical data, and the system can compare the number with data in the BL-DB and determine if the number is equal to, greater than, less than, or falls within a certain range. The data set can be sorted, displayed, and printed.

There is a need to search for a set of properties at the same time, thus the data set created in each search step is kept during the session. Any one of the sets can be "ANDed," "ORed," or "NOTed" with another set, saved in user's work file, and used to generate a subset by extracting appropriate records.

Characteristics of the BL-DB

The BL-DB is a computerized fact database that can store and serve not only qualitative toxicological test data but also quantitative data. Moreover, the BL-DB can treat precise test data in tabular form. Many, possibly all, fields are defined as search fields, thereby facilitating retrieve of target data through many ways. Many kinds of test data can be stored with the same for-

```
S: *****   Welcome to Biological (BL) Database   *****
   *****                        DATE 91/01/09 10.50 *****

S: Session Started                        NIHS copyright

U: FR BENZ&
[S01]S:      141 substance(s)
U: AND EXP MN
[D01]S:       30 data
U: AND RES +
[D02]S:       19 data
U: SHOW 9

#0001 BL041000539
COMPONENT INF( 1)
COMP= Trenimon
RN  = 68-76-8  PUR  = Bayer   SOLV= Isotonic saline 0.9%
EXP = MU
EXP = MN
MAT = Mammals
MAT = Mouse
MAT = Albino CD (Charles River, Paris)
MAT = Bone marrow cells
AGE = 10-20(wk)  WGT = 26-40
ROT = ip
SCJ = 2*(24hr)6hr
DOSE= 0-0.25mg/kg
NUMA= 4(M+F)
TRET= Kastaenbaum & Bowman
JUDG= +
    = +  MNPCE   MED  <0.016mg/kg
REFR= 0165-1218/71/0012-0417        SUB  = 8803564B
```

FIGURE 3. An example of retrieval result.

mats. Therefore, users can retrieve data of different toxicological tests by the same manner. This system downloads retrieved data sets into the multiple purpose working file to enable further analysis of the data set using another application software. Files are separated into a substance, test data, and bibliography files to allow easy expansion of data to other test systems.

Limitations

Currently, there are data for mutagenicity, teratogenicity, and carcinogenicity in the BL-DB. Data for the mutagenicity, however, especially on Salmonella/microsome mutation assay and the micronucleus assay, are dominant. Of about 20,000 test data on 1700 chemical substances by the mutation assays, 1600 pieces of data on 350 chemicals are from the Salmonella/microsome assay and approximately 1000 data on 300 chemicals are from the micronucleus assay and are available at present. These data are retrieved and used for the structure–activity relationship study.

There are several limitations. First, the budget for this program is limited and makes it difficult to fully maintain and regularly update the BL-DB. New data are not abstracted and put into the database periodically. To overcome this situation, we have asked members of Mammalian Mutagenesis Study group, a suborganization of the Environmental Mutagen Society of Japan, to help evaluate, abstract, and input data on the micronucleus assay, including those articles that appeared in Japanese.

Second, it is sometimes difficult to abstract data from references. Although we respect the judgment of the authors, there is disagreement of opinions among authors and abstracter(s). At the moment, we do not have any authorized committee to evaluate data, which is a desirable and necessary step in establishing the total system.

Third, BL-DB is a stand alone system; therefore, there is no way to access it on-line. There are now plans to make a hard copy data book on the micronucleus assay data as a first step to minimize this shortcoming.

This study was supported by funds from Science and Technology Agency of Japan.