PHYTOTOX: A Database Dealing with the Effect of Organic Chemicals on Terrestrial Vascular Plants

CHRISTOPHER L. ROYCE, JOHN S. FLETCHER,* and PAUL G. RISSER†

University of Oklahoma, Norman, Oklahoma 73019

JAMES C. MCFARLANE

Environmental Research Laboratory, Corvallis, Oregon 97330

FRANK E. BENENATI

Environmental Protection Agency, Washington, D.C. 20460

Received July 18, 1983

A new database, PHYTOTOX, dealing with the direct effects of exogenously supplied organic chemicals on terrestrial vascular plants is described. The database consists of two files, a Reference File and Effects File. The Reference File is a bibliographic file of published research papers. The Effects File contains chemical- and plant-response information derived from the papers listed in the Reference File. We anticipate that government, academic, and industrial users will find the PHYTOTOX database an invaluable tool in predicting and modeling the effects of organic compounds on plants and their ecosystems.

INTRODUCTION

Application of organic chemicals to large acreages of terrestrial vegetation has become a common practice in agriculture and forestry over the last 30 years.[1] Approximately 5×10^9 lb (1 lb = 0.45 kg) of organic chemicals is applied annually throughout the world to reduce or eliminate the presence of weeds, insects, and plant pathogens in cultivated crops.[2] It is estimated that in the U.S. the use of pesticide chemicals reduces crop loss by approximately 9%, but even with their use, there remains a 33% loss annually to the damaging influence of plant pests.[3] Thus, there is compelling motivation for the introduction of new agricultural chemicals that will reduce these agricultural and horticultural losses. To accomplish this without imposing environmental or health hazards on the public presents a major challenge to both industry and government. Success in this effort depends in part on effective utilization of extensive research data published on the influence of organic chemicals on terrestrial plants. However, it is virtually impossible for anyone to gain maximum benefit from this large pool of potentially useful information because it is dispersed among hundreds of journals published in many countries and several languages. It is our goal to make these data readily accessible through a computerized database, PHYTOTOX. This database, currently under development, is designed to facilitate both research and regulatory efforts directed toward the development and marketing of new organic chemicals that are environmentally safe. This includes not only agricultural chemicals but also other chemical products as well that may be inadvertently disseminated into the environment.

PHYTOTOX has been designed so that it is compatible with the Chemical Information System (CIS)[4] maintained by NIH and EPA. As PHYTOTOX nears completion, it is anticipated that it will be incorporated into CIS as a part of the CIS-subcomponent SPHERE (Scientific Parameters for Health and the Environment, Retrieval and Estimation).[5] At that time, the full potential of PHYTOTOX will be realized because it will then be linked through Chemical Abstracts Service Registry Numbers with the Structure and Nomenclature Search System (SANSS),[6] which is currently a part of CIS. The combination of SANSS and PHYTOTOX will provide a system that relates the structure of applied chemicals to the biological response of recipient terrestrial plants. Although similar databases have been assembled by some industrial firms,[7,8] to our knowledge the SANSS–PHYTOTOX system will be the first structure–activity database system dealing with higher plants that will be available for common usage by industry, government, and the general public.

DATABASE DESCRIPTION

Design of the Database. The PHYTOTOX database consists of two files, a Reference File and an Effects File. The Reference File is a bibliographic listing of published papers dealing with the influence of organic chemicals on terrestrial plant growth and development. The Effects File contains chemical names, plant species names, and other selected information pertaining to experimental design and plant response. The format of the database permits it to be searched from a host of starting points such as plant species name, chemical name, physiological response, etc. The versatility of available search strategies makes the database extremely useful in evaluating our present knowledge on environmental questions and ascertaining the experimental basis for this knowledge.

The database has been formatted in the GIPSY processing system[9] developed at the University of Oklahoma. GIPSY has been used extensively and described thoroughly by The United States Geological Survey.[10] Two important GIPSY terms related to the structure of the PHYTOTOX database are "label" and "record". Every text (data) line in the database is identified with a one to seven character GIPSY label. Groups of related labels and their data are associated in clusters called records. The uses of labels and records will be illustrated in the following sections.

Reference File. The Reference File was developed first and serves two functions. It lists the scientific papers to be read in developing of the Effects File, and in the completed database it will serve as a computer-linked literature citation file so that the original publications on which effects records are based can easily be identified.

† Present address: Illinois Natural History Survey, Champaign, IL 61820.

0095-2338/84/1624-0007$01.50/0

Table I. Search Strategy Used to Secure Pertinent Titles from the National Agricultural Library Computer Database

search terms used to identify appropriate studies	exclusion terms used to eliminate inappropriate studies
herbicid$[a]	biological control
weed$ and control$	interactions
phytotox$	mixtures
allelopath$	insect studies
	mammal studies
	bird studies
	fish studies
	aquatic plants
	microorganisms
	diseases

[a] $ signifies that the database was searched with both the singular and plural form of the search term.

Table II. Subject Categories That Were Searched in *Biological Abstracts*

agronomy	phytopathology
ecology[a]	nonparasitic diseases
plant[a]	plant physiology
economic botany	enzymes[a]
food technology	growth, differentiation[a]
fruits, nuts, and vegetables	growth substances
general methods[a]	metabolism
preparation, processing, and storage[a]	nutrition[a]
forestry	photosynthesis
genetics and cytogenetics	reproduction[a]
plants	respiration[a]
horticulture	translocation, accumulation
morphology and anatomy of plants[a]	water relations
pest control	

[a] After one-third of the 872 volumes had been searched, these categories were no longer searched because of the few appropriate citations that they contained.

The literature search covering the years 1926–1981 has been completed. Approximately 9200 published papers were identified that described investigations dealing with the influence of exogenously supplied chemicals on plant growth and development. All of the identified papers satisfied three selection criteria: a terrestrial plant was studied, organic chemicals were applied, and direct effects were evaluated. In order to keep the database current, personnel at the University of Oklahoma will update the database on a regular basis contingent upon financial support from either government or private sources.

The titles were identified from three sources: The Oak Ridge National Laboratory Report EIS-155/V1, The National Agricultural Library Computer Database (1970–1981), and *Biological Abstracts* (1926–1981). The number of useful titles secured from each of these sources were 1007, 1845, and 6731, respectively. Different methods were used to secure the titles of relevant publications from each of these sources. The Oak Ridge report was screened for appropriate papers by reading the abstract information in the report. The search strategy used with the National Agricultural Library Computer Database is shown in Table I. A manual search of *Biological Abstracts* was conducted to screen all of the titles and abstracts in the 872 volumes covering the period from 1926 to 1981. Categories listed in the abstracts that were screened are shown in Table II. Although time consuming, the manual search was necessary because many titles would not have been obtained through computer searching since only the last 12 years of *Biological Abstracts* are included in the Biosis Database.

The citation information collected from the identified papers has been entered into the computerized Reference File. The six GIPSY labels used in the reference file are listed and defined in Table III. An example reference file record is given in Figure 1.

Table III. GIPSY Labels Used in the Bibliographic Section of PHYTOTOX

label	definition
REFNO	file number
SENAUTH	senior author
YEAR	year of journal publication
OTHAUTH	other authors (optional label)
ARTITLE	article title
REFRNCE	journal citation
REFSTAT[a]	processing status

[a] This label will be deleted in the finished database.

Table IV. GIPSY Labels Used in the Effects File of PHYTOTOX

label	definition
EFFNO	file number
SPECIES	a single experimental plant taxon
CHEMICL	a single experimental organic chemical
CASNO	Chemical Abstracts Service (CAS) Registry Number
DOSAGE	a single chemical concentration
EFFECTS	coded symptom data
CATEGRY	coded experimental design data
REFILNO	bibliographic source file number
GRADE	usefulness and/or quality of source

```
FILE NUMBER -   R02526
SENIOR AUTHOR -   MURRAY, D.S.
YEAR -  73
OTHER AUTHOR(S) -
  P.W. SANTELMANN, H. GREER.

TITLE -
  DIFFERENTIAL PHYTOTOXICITY OF SEVERAL DINITROANILINE HERBICIDES.
REFERENCE -
  AGRON J     65:34-36
STATUS: HAVE -  Y-R
    ET
```

Figure 1. Sample bibliographic record from the PHYTOTOX database.

```
FILE NUMBER -   008316.1
PLANT SPECIES -
  GOSSYPIUM HIRSUTUM L. "COTTON" 'WESTBURN'

CHEMICAL -
  "PROFLURALIN"
   N,N-PROPYL-N-CYCLOPROPYLMETHYL-4-TRIFLUOROMETHYL-2,6-DINITROANILINE 'CGA
   10832'

CAS NUMBER -  26399-36-0

DOSE -
  1.12 KG/HA

EFFECT(S) -
  PL INJ 10

CATEGORY -
  MF MT 0016 IN PL CV LIQ LO NON
REF. CITE -   R02526
    C
```

Figure 2. Sample of an effects record from the PHYTOTOX database.

Effects File. Each complete record in the Effects File contains information about the effect(s) of the application of one concentration of a single organic chemical on a particular plant species as reported in one paper. Thus, any paper may yield many records, depending upon the number of species and chemicals used and the range of concentrations used for each chemical.

Each record in the Effects File contains five to nine labels and their associated text (Table IV). The text associated with six of the labels is in plain language, whereas the text associated with three of the labels, EFFECTS, CATEGRY, and GRADE, is encoded. A sample of a complete plant effects record as it appears when it is retrieved from the computer is shown in Figure 2.

Each record contains a plant name listed with the label SPECIES. The standard format used in PHYTOTOX for plant nomenclature is *Gossypium hirsutum* L. "cotton" 'Westburn'. The formal latinized name is listed first, the

Table V. Examples of Symptoms and Their Codes as Used in the PHYTOTOX Database

symptom	code	symptom	code
abscisic acid decrease	AAD	injury	INJ
abscisic acid increase	AAI	maturation delayed	MAD
abscisic acid no effect	ASN	meiotic abnormality	MAB
abscission increase	ABS	membrane damage	MEM
auxin decrease	AUD	mitotic rate decrease	MRD
chlorophyll increase	CPI	necrosis, marginal	MNC
chlorosis	CHL	necrotic lesion	NEC
control	CON	none	NON
cover decrease	CVD	parthenocarpy increase	PCI
deformation	DEF	photosynthesis decrease	PSD
desiccation	DES	protein synthesis increase	PRI
DNA synthesis rate increase	DNI	respiration decrease	RSD
dormancy induction	DOI	senescence retarded	SNR
dry mass decrease	DMD	sex expression change	SEX
floral induction	FLI	transpiration increase	TRI
germination decrease	GRD	water content decrease	WAD
harvest yield increase	HYI		

common name is listed second, in quotes, and other identifiers are added last and set off with apostrophes. Each record contains a chemical name associated with the label CHEMICL. The standard format for presenting chemical names in PHYTOTOX is "Profluralin" N,N-propyl-N-cyclopropylmethyl-4-trifluoromethyl-2,6-dinitroaniline 'CGA 10832'. The common name is listed first in quotes, the formal name (as given in the report) follows the common name, and any additional identifying names are listed last and set off with apostrophes. An additional means of identifying chemicals, the Chemical Abstracts Service (CAS) Registry Number, is listed in each record in PHYTOTOX with the label CASNO.

The concentration of the chemical applied is listed with the label DOSAGE. The concentration is given in the units used in each published paper, and abbreviations are standardized. Not all experimental concentrations reported in a study are necessarily incorporated into the database. If no symptoms were noted when increasing amounts of a test chemical were supplied, then this result is listed in the database for only the highest concentration reported. If a species was effected by an applied chemical, then only certain selected concentrations are entered into the database. The concentrations entered include the lowest concentration triggering each unique symptom observed, the lowest concentration responsible for producing the maximum effect, and the midpoint concentration if the response is nonlinear.

Symptoms and symptom sites are listed with the label EFFECTS. These data are encoded to reduce storage requirements and facilitate the use of a controlled vocabulary. Every symptom is listed with a symptom site, identifying the location on the plant where the symptom was manifested. Symptoms are quantified, if possible, by expressing the response as a percent of the control. More than one symptom may be reported in one record. The database includes 160 possible symptoms, reflecting 70 different processes (e.g., transpiration, photosynthesis, etc.). Twenty-one different symptom sites are used. The spectrum of the symptoms and symptom sites used in the database was predicated on the information present in the papers examined. Some of the symptoms and their codes are given in Table V. The symptom sites and their codes are given in Table VI. A sample EFFECTS entry for one record is PL INJ 10 (Figure 2). This entry states that plant injury was 10% greater in the treatment group than in the control group.

Nine aspects of the experimental design are encoded and listed with the label CATEGRY as shown in Table VII. Care has been taken that none of the codes used for CATEGRY conflict with those used for EFFECTS. A typical entry coded

Table VI. Sympton Sites and Their Codes Used in the PHYTOTOX Database

symptom site	code	symptom site	code
entire plant	PL	hypocotyl	HY
shoot	SH	bud	BD
meristem (apical or axillary)	ME	bulb	BB
		flower/inflorescence	FL
embryo	EM	strobilus (mega-, micro-, etc.)	SR
leaf	LF		
cotyledon	CN	fruit	FR
coleoptile	CP	root	RT
stem	ST	cell	CL
tuber	TU	organelle	OR
corm	CO		
rhizome	RH		

Table VII. Information Encoded in the Database under the Label CATEGRY

(1) phenological stage of the plant at the time of chemical application (e.g., seedling) (9)[a]
(2) phenological stage when the symptom was observed
(3) time between application and symptom(s)
(4) condition of plant material at time of application (e.g., intact) (5)
(5) application site (e.g., leaf) (19)
(6) manner in which the plant was maintained during treatment (e.g., greenhouse) (7)
(7) physical form in which the chemical was applied (e.g., liquid) (5)
(8) method of application (e.g., spray) (10)
(9) presence of chemical application aid (e.g., surfactant)

[a] Values in parentheses indicate the number of different possibilities that are recognized.

for the CATEGRY label would be MT MT DO16 IN PL CV LIQ LO NON (Figure 2). This entry states that entire (PL), mature (MT), intact (IN) plants were treated with a liquid (LIQ) chemical added to their growth medium (LO) and the symptoms were noted 16 days (DO16) later during mature growth (MT) of the plants. In addition, the code group reveals that the plants were maintained under cultivated field conditions (CV), and no application aid (NON) was used to facilitate the action of the chemical provided.

The text associated with the label GRADE is an alphabetical character denoting the usefulness of the scientific paper from which data were taken. This entry also serves as a flag to researchers using the database to indicate whether or not the published paper contains additional information that is not included in the PHYTOTOX database. The codes that are used with the GRADE label are A, B, C, and D. A paper that contains all of the types of data that are needed in an effects record, but little or no additional useful data, is graded C. Papers containing a substantial amount of additional data that are directly or indirectly related to the general theme of PHYTOTOX are graded either B or A, depending on the amount and quality of the additional data. Grade D papers are those that lack the minimum information required to make a complete effects record or where the experimentation was poor and perhaps invalid.

Current and Future Status. The PHYTOTOX database is currently maintained on an IBM 3081 system at the University of Oklahoma where the database is being developed. The database is accessed via a teleprocessing time-sharing system (TSO) through 300- or 1200-baud dial-in telephone lines or via 7200-baud dedicated telephone lines on the university campus. The entire database is on TSO disk to facilitate rapid interactive search and retrieval.

All data entry, editing, and on-campus database use are done interactively with the IBM 3081 mainframe computer with IBM 3278 teleprocessing terminals (at 7200 baud) or with Apple II computers coupled with D. C. Hayes Micromodem

II modems (at 300 baud). The data entry software is written in SNOBOL4, a text-oriented symbolic language.

In-house use of the partially completed database has clearly demonstrated its usefulness in environmental assessment work and toxicology research. Future incorporation of PHYTOTOX into SPHERE[5] will greatly amplify the usefulness of the database since it will then be linked to other related databases such as SANSS[6] and CHEMFATE.[12] The interactive use of these databases will be valuable in predicting how a chemical released into the environment will behave and whether or not it will come in contact with a plant species in sufficient concentration to cause an adverse or toxic effect. Realizing that such evaluations are only predictions and that testing must often follow, it is also worth noting that the PHYTOTOX database can serve a second major role in designing tailored experiments for specific chemicals. In this manner, the database is expected to prove useful in the environmental assessment of new commercial chemicals.

ACKNOWLEDGMENT

Development of this database has been supported through cooperative agreements CR-807391 and CR-810195 with the U.S. Environmental Protection Agency. Thanks are extended to the many students who have been involved with the evaluation of the research papers used in preparing this database and to Tom Weaver for his assistance with the GIPSY processing system. This article has not been subjected to the U.S. Environmental Protection Agency's peer and policy review and, therefore, does not necessarily reflect the views of the Agency, and no official endorsement of trade names or commercial products should be inferred.

REFERENCES AND NOTES

(1) Eichers, T. R. In "CRC Handbook of Pest Management in Agriculture"; Pimentel, D., Ed.; CRC Press: Boca Raton, FL, 1980; Vol. II, p 3.
(2) Pimentel, D. In "CRC Handbook of Pest Management in Agriculture"; Pimentel, D., Ed.; CRC Press: Boca Raton, FL, 1980; Vol. I, p 3.
(3) Pimentel, D.; Krummel, J.; Gallahan, D.; Hugh, J.; Merrill, A.; Schreiner, I.; Vittum, P.; Koziol, F.; Back, E.; Yen, D.; Fiance, S. In "CRC Handbook of Pest Management in Agriculture"; Pimentel, D., Ed.; CRC Press: Boca Raton, FL, 1980; Vol. II, pp 45, 46.
(4) Milne, G. W. A.; Heller, S. R. "NIH/EPA Chemical Information System". *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 204-211.
(5) "Scientific Parameters in Health and the Environment, Retrieval and Estimation: A Requirement Analysis and Examination of Alternatives"; CRC Systems Incorporated: Fairfax, VA, 1981; EPA Contract 68-01-4795.
(6) Milne, G. W. A.; Heller, S. R.; Fein, A. E.; Frees, E. F.; Margaret, R. G.; McGill, J. A.; Miller, J. A.; Spiers, D. S. "The NIH-EPA Structure and Nomenclature Search System". *J. Chem. Inf. Comput. Sci.* **1978**, *18*, 181-186.
(7) Brown, H. D.; Costlow, M.; Cutler, F. A., Jr.; Demott, A. N.; Gall, W. B.; Jacobus, D. P.; Miller, C. J. "The Computer-Based Chemical Structure Information System of Merck Sharp and Dohme Research Laboratories". *J. Chem. Inf. Comput. Sci.* **1975**, *16*, 5-10.
(8) Bond, V. B.; Bowman, C. M.; Davison, L. C.; Roush, P. F.; Young, L. F. "Applications of the Wiswesser Line Notation at the Dow Chemical Company". *J. Chem. Inf. Comput. Sci.* **1982**, *22*, 103-105.
(9) "General Information Processing System (GIPSY) Manual"; University of Oklahoma Office of Information Systems Programs: Norman, OK, 1982.
(10) Calkins, J. A.; Keefer, E. K.; Ofsharick, R. A.; Mason, G. I.; Tracy, P.; Atkins, M. "Description of CRIB, the GIPSY Retrieval Mechanism, and the Interface to the General Electric MARK III Service". *Geol. Surv. Circ. (U.S.)* **1978**, No. 755-A.
(11) Ross, R. H.; Kemp, H. T.; Pyon, M. G.; Hammons, A. S.; Ensminger, J. T. "Chemicals Tested for Phytotoxicity"; Oak Ridge National Laboratory: Oak Ridge, TN, 1979; EIS-155/V1.
(12) Howard, P. H.; Sage, G. W.; Lamacchia, A. "The Development of an Environmental Fate Data Base". *J. Chem. Inf. Comput. Sci.* **1982**, *22*, 38-44.

Central Patents Index Chemical Code: A User's Viewpoint†

EDLYN S. SIMMONS

Merrell Dow Pharmaceuticals Inc., Cincinnati, Ohio 45215

Received August 8, 1983

Subscribers to Derwent Publications Ltd.'s Central Patents Index can use the Chemical Fragmentation Code to identify patents describing a chemical compound whether its structure is expressed generically or specifically and whether the compound is claimed or merely described as novel. Compounds are encoded via molecular fragments—elements, carbon chains, rings, and functional groups—which are overcoded on a single record to define a Markush group and are retrieved by encoding chemical structures with the same fragment codes, expressing Markush groups by the Boolean OR and combining fragments with the ORBIT LINK operator. Examples are presented in which retrieval with the Chemical Fragmentation Code is compared with the retrieval of patent references by alternative search techniques.

Because much of the information published about chemical compounds appears only in chemical patents, patents are valuable resources whenever chemical information is needed. Chemical patents are especially important for searchers employed by organizations involved in the development or marketing of chemical compounds and processes. Only patents can tell us whether a compound or process is protected by patent claims. Patents are also vitally important for evaluating the patentability of new inventions, for valid patents can be obtained only for inventions that have not previously been taught or suggested in the prior art, and the prior art includes all of the world's patents. But patents are useful as information sources only to the extent that they are indexed so that the compounds they describe can be identified.

Modern chemical inventions, especially those involving new chemical compounds, are typically defined in terms of complex generic structures in what is called Markush format, so indexing the compounds in patents is not as simple as indexing compounds in the journal literature. A Markush formula such as formula I in Figure 1 includes one or more variable substituents defined in terms of a group of alternative molecular substructures, which is called a Markush group. The position of substitution in a Markush formula can also be variable. The name Markush is derived from a ruling in 1925 on a patent application filed by an inventor named Eugene A. Markush,[1] in which the U.S. Patent Office declared that this format was a legally acceptable way to define a genus of compounds. A

† Presented at the 185th National Meeting of the American Chemical Society, Seattle, WA, March 24, 1983.

0095-2338/84/1624-0010$01.50/0