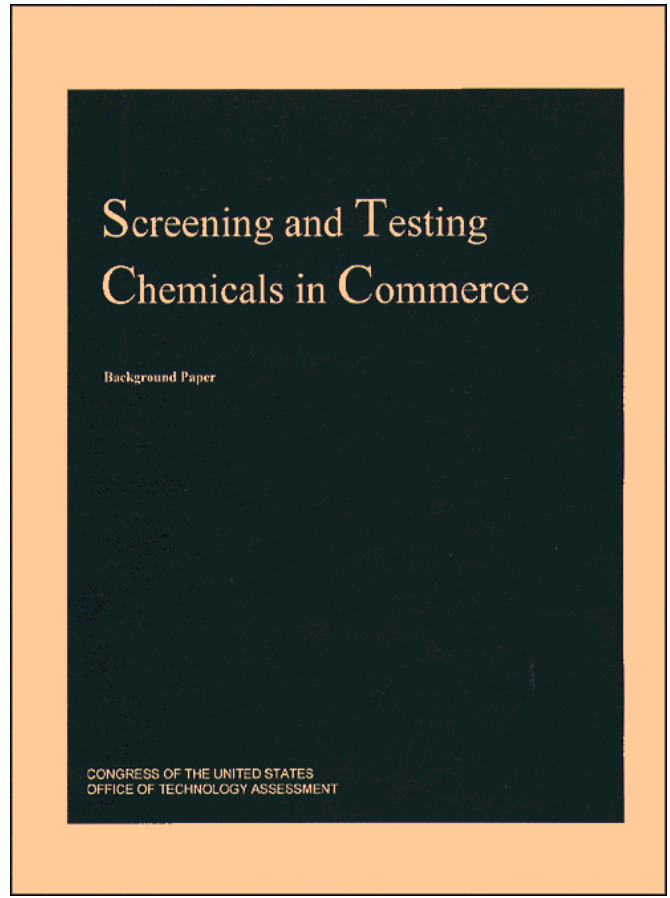
Screening and Testing Chemicals in Commerce

September 1995

OTA-BP-ENV-166



Foreword

In 1994 the Senate Environment and Public Works, Subcommittee on Toxic Substances, Research and Development asked the Office of Technology Assessment (OTA) to carry out a study on the Toxic Substances Control Act (TSCA). Enacted in 1976, TSCA gives the Environmental Protection Agency (EPA) authority to screen both new and existing chemicals-in-commerce to protect workers, consumers, and the environment. The Senate Subcommittee asked OTA to see if there are technologies or new approaches that would allow a more rapid screening of the existing chemicals-in-commerce for possible negative effects on human health and the environment.

This background paper comes from a workshop held by OTA in April 1995. OTA invited experts from industry, academia, and government who are involved with toxicity testing and screening chemicals. The individual chapters of this report were written by participants in the workshop. Each chapter discusses a specific type of testing or screening method. Every chapter has been reviewed by at least two outside reviewers for accuracy and completeness. After revisions, the final versions are produced here.

The report reviews some of the many test technologies and techniques available for screening chemicals-in-commerce for toxicity. Some of the test technologies, such as those for predicting carcinogenesis, are well established and results can be fairly clearly linked to real health effects in humans. Results of other tests, such as those for detecting neurotoxicity, are presently less clearly linked to actual human health effects. Structure-activity analysis, a class of techniques used to predict the toxicity of unknown compounds based on knowledge of related chemicals, may be especially useful for screening large numbers of compounds. However, predictive methods and computer modeling of toxicity will never be a complete substitute for real toxicity data.

OTA appreciates the assistance and support it received for this effort from many contributors and reviewers. They provided OTA with valuable information and important insights critical to the completion of this background paper. OTA, however, remains solely responsible for the contents of this report.

ROGER C. HERD W

Director

Contributors

Errol Zeiger
Workshop Chairman
NIEHS

James Bond
CIIT

Robert Chapin
NIEHS

Rick Corley
Dow Chemical

George Daston
Procter & Gamble

Elin Eysenbach
Procter & Gamble

Anne Fairbrother
Ecological Planning and Toxicology, Inc.

Spencer Farr
Xenometrix, Inc.

Alan Goldberg
Johns Hopkins Univ.

Joseph Gorsuch
Eastman Kodak

Bert Hakkinen
Procter & Gamble

Jean Harry
NIEHS

James Huff
NIEHS

Nancy Kerkvliet
Oregon State University

Neil Krivanek
DuPont

Michael Luster
NIEHS

James MacGregor
SRI International

Larry Needham
Centers for Disease Control

John O'Donoghue
Eastman Kodak

Sally Perreault-Darney
NIEHS

Richard Purdy
3M

Ann Richard
U.S. EPA

Gary Rosenthal
Somatogen, Inc.

Barbara Shane
Louisiana State Univ.

Judson Spalding
NIEHS

Steve Tannenbaum
Massachusetts Inst. of Tech.

Gilman Veith
U.S. EPA

Pauline Wagner
U.S. EPA

John Walker
U.S. EPA

Maurice Zeeman
U.S. EPA

Project Staff

Clyde Behney
Assistant Director, OTA
Health, Education, and the Environment

Robert Niblock
Environment Program Director

Mark Brown
Project Director

Yolanda Rivas
Research Analyst

Michael Bowes
Senior Analyst

David Butler
Analyst

ADMINISTRATIVE STAFF

Kathleen Beil
Office Administrator

Kimberly Holmlund
Administrative Secretary

Nellie Hammond
Administrative Secretary

Reviewers

Daniel Acosta
University of Texas, Austin

Romauldo Benigni
Istituto Superiore di Sanita

Robert Boethling
U.S. EPA

William Breslin
Dow Chemical

Rick Brown
Dow Chemical

Dave Brusick
Corning Hazelton

Tim Buckley
U.S. EPA

Yvonne Freund
SRI International

John Giesy
Michigan State Univ.

Ralph Gingell
Shell Oil

John Heddle
York University

Claude Hughes
Bowman Gray School of Medicine

Jerry Lynch
Exxon Biomedical Sciences

William Meylan
Syracuse Research Corp.

Robert Przgoda
Exxon Biomedical Sciences

Chuck Ruffing
Eastman Kodak

Harry Salem
Edgewood Research

David Seyler
Lilly Research Laboratories

Contents

1
Introduction

The focus of the chapters presented in this background paper are screening and test strategies for reviewing the Toxic Substances Control Act (TSCA) inventory of existing chemicals in commerce.[1] The screening problem poses a challenge, not only in terms of the numbers of chemicals that have undergone little testing or review, but also in terms of the many exposure routes and ecological and health endpoints of potential concern. The foremost goal of a review strategy must be to identify the chemicals that pose the greatest potential ecological and health risks, and allocate limited testing resources to better characterize these risks.

BACKGROUND

In 1994, the Senate Environment and Public Works, Subcommittee on Toxic Substances, Research and Development asked the Congressional Office of Technology Assessment to carry out a study of Existing Chemicals Program under the Toxic Substances Control Act (TSCA, PL 94-469).

Congress originally enacted TSCA in 1976. Administered by the U.S. Environmental Protection Agency (EPA), the law gives EPA authority to screen and require further testing of both new and existing chemicals in commerce as necessary to protect public health and the environment. TSCA states, "It is the policy of the United States that . . . adequate data should be developed with respect to the effect of chemical substances and mixtures on health and the environment and that the development of such data be the responsibility of those who manufacture and those who process such chemicals and mixtures. "

The task of addressing the large number of existing chemicals has proven to be daunting. A 1984 study by the National Research Council (2) concluded that no toxicity data was available for almost 80 percent of the chemicals in general commerce and only 10 percent of substances had test data that were adequate for conducting a health-hazard assessment. In 1994, the Government Accounting Office reported that the EPA has fully reviewed only about 2 percent of the existing chemicals in commerce (1).

The current estimates are that approximately 70,000 chemicals have been used in U.S. commerce since 1976. Of these, roughly 30,000 are polymers that present little health risk. Another 25,000 are produced in low volume (less than 10,000 lbs/year), with some no longer in production. Thus, it is certainly debatable whether all 70,000 chemicals in commerce present equal concerns. Still, there remain some 15,000 chemicals that are produced in significant volumes. with approximately 3 - 4,000 produced in excess of 1,000,000 lbs/year. For perhaps thousands of these chemicals of potential concern, toxicity and exposure data remain inadequate for risk assessment.

The complete toxicological evaluation of thousands of chemicals would be both time consuming and extraordinarily expensive. Full toxicological evaluations for a single chemical can cost

[1]**The phrase "chemical in commerce" is used to mean all chemical substances that are potentially regulated under TSCA. TSCA covers most chemicals except drugs, pesticides, tobacco, food products, food additives, and radioactive materials. The inventory of existing chemicals includes some 60,000 substances registered as being in commerce when TSCA was passed, plus others that have subsequently been reviewed by EPA as new chemicals.**

Table 1-1: Workshop Topics

Carcinogenicity	**Environmental toxicity** (ecological endpoints)
Dermal and ocular toxicity	Multiple endpoints (integrated test strategies)
Immunological toxicity	Exposure assessment (biomarkers)
Neurological toxicity	Structure-activity methods
Reproductive and developmental toxicity	

up to $2,500,000, involve several thousand test animals, and take five years to complete. An effective screening strategy must cheaply identify those chemicals that pose the greatest potential risk, before committing limited resources.

THE WORKSHOP

On April 24-25, 1995, OTA held a workshop to address the question of whether there were technologies that could be used to rapidly screen existing chemicals in commerce for effects on human health and the environment. The purpose of the workshop was to compile and review current and developing systems that may be relevant to the needs of TSCA.

We invited panels of experts to cover various specific testing endpoints and general screening approaches (listed in table l-l). Each panel was chosen to include individuals from a mix of academic, government, and industry backgrounds. The individual chapters of this background paper were written by the panels that participated in the workshop. Each panel was asked to address the following questions in their chapters:

- What are the best currently used methods to identify chemicals of concern and their health effects?
- Cost and time are critical considerations for an evaluative strategy. What faster and cheaper screening technologies are available that can inform the review process or set priorities for further testing? Address the issue of using these assays to evaluate various number of chemicals (100, 1,000, 10,000, etc.).
- What are the tradeoffs in using the cheaper and faster screens? Consider confidence as well as ambiguities of results and reproducibility between different laboratories?
- For the specific endpoints (e.g. carcinogenicity), what is the contribution from receptor-based assays, SAR approaches, and mechanism-based assays?
- For the specific endpoints, how efficiently can the screening tests be integrated into an overall screening and test strategy for a comprehensive evaluation of a chemical?
- Finally, test technology is a discipline in constant development, and test strategies should be designed to adopt technological innovations. What new developments might we expect for test technologies in the next decade?

The efforts to answers to these questions are contained in this background paper and are worthy of an audience among legislators, regulators, and the informed public, as well as toxicologists. Although several of the papers are quite technical, all offer considerable insight as to current regulatory practices and scientific capabilities.

REFERENCES

1. Guerrero, P. F., U.S. General Accounting Office, “Toxic Substances Control Act: EPA’s Limited Progress in Regulating Toxic Chemicals,” testimony at hearings before the Senate Subcommittee on Toxic Substances, Research and Development, GAO/T-RCED-94-212, May 17, 1994.
2. National Research Council, Commission on Life Sciences, *Toxicology Testing: Strategies to Determine Needs and Priorities*, (Washington, DC: National Academy Press, 1984).

2

Multiple Endpoints and Integrated Test Strategies

Neil Krivanek and Rick Corley

ABSTRACT: *The basic questions of what toxic endpoints test for and the appropriate time to conduct these tests has remained an issue for the regulatory and toxicology community since the creation of TSCA. In order to maximize the return of multiple endpoint screening tests, we have examined several important components of these tests. Initial selection of compounds for testing should consider potential human exposure and production volumes.*

Simple acute studies in rodents should precede multi-dose studies. For multiple endpoint screening properly designed 28-day exposure studies can generate data for most endpoints. We have examined some of the boundaries which impact quality and interpretation of test data.

This paper will discuss some key interrelated items in toxicology screening testing for multiple endpoints that are pertinent to this OTA workshop on the evaluation of the existing chemicals review program which is administered by the EPA.

The scope of this paper is first to describe the current state for evaluating multiple toxicity endpoint in mammalian systems and secondly describe the important issues we see that need to be addressed in screening tests for existing chemicals under TSCA.

Screens, by their nature, involve a series of trade-offs or compromises. They need to be sensitive enough to identify subtle hazards, selective enough to minimize false positives and negatives yet be manageable enough (cost and time) to evaluate large numbers of materials. For the purposes of this paper, a screen can consist of either a battery of studies or a single study which incorporates numerous endpoints that can be completed within a reasonable period of time (i.e. less than one year or budget cycle) at a reasonable cost per chemical. Such an exploratory screen should provide sufficient information to identify potential systems affected (e.g. respiratory, cardiac, digestive, reproductive, neurological, immunological, etc.), its severity and dose-responsiveness to make a preliminary assessment of potential risk (e.g. margins of safety). Appropriately designed screens may also be used to direct further research to either more fully characterize an effect or to establish the relevancy of an effect to exposed populations, human or otherwise.

The focus on this manuscript is on the initial assessment (screen) for potential toxicity to a variety of organ systems as the first in a possible tiered evaluation.

The main question which needs to be consistently asked is "what" endpoints to test for and "when" to conduct these tests. For industrial chemicals, including those which come under TSCA, the goals are to safely manufacture, use, and dispose of these materials. These toxicity data are used to develop programs for safe handling and use, for occupational exposure limits, appropriate warnings for use, and appropriate information on Material Safety Data Sheets (MSDS's).

TESTING TRIGGERS

What we would like to discuss is what toxicity information triggers certain types of testing. To date most testing has been done on a case-by-case basis. The types of tests have been largely triggered by the toxicity of the chemical and by structure-activity relationships. When to test has generally been driven by a combination of potential exposures and production volume, with em-

phasis on exposure. However within the last few years, strategies have been emerging in several parts of the world, notably Europe, which link a certain amount of testing with a certain volume of production of the chemical without an assessment of exposure. This particular issue of what to test for and when to test has also been controversial since the implementation of TSCA.

To illustrate, I will list two examples of priority setting for testing. This is shown in table 2-1. This is a comparison of the Screening Information Data Set (SIDS) from Organization for Economic Cooperation and Development (OECD) and the development of a criteria document by National Institute of Occupational Safety and Health (NIOSH). This side by side comparison shows that each contains the same basic elements of the initial exposure assessment. This is part of the first step to decide what is important for prioritization. Both efforts emphasize chemical use patterns, type and extent of workplace exposure, and number of workers exposed.

The concept of a production volume trigger for defining testing strategies is shown in table 2-2 and 2-3. In table 2-2, for 1 ton/year with and without a 5 ton cumulative trigger the information required focuses on physical chemical properties, acute animal and aquatic testing, and limited repeated dose studies in animals for up to 28 days. In table 2-3, for volumes of greater than 100 tons/year to less than 1000 tons/year with 500 tons cumulative, testing is more comprehensive.

Table 2-1: Comparison between OECD SIDS and U.S. NIOSH Data Requirements for Chemicals-in-Commerce

SIDS **Initial exposure assessment**	**NIOSH** ***Criteria documents &*** **preliminary prioritizations**
Identification of occupational or professional uses of the chemical or products in which it is contained, and of use in consumer products	Frequency of use (occupations, processes, number of industries)
Uses in consumer products	
Function of chemical (for each consumer product identified)	
Weight fraction of chemical (actual or recommended)	
Form or product (as marketed, e.g., aerosol, powder, liquid)	
Workplace exposure, frequency and duration of such exposure	Extent of exposure and background exposures
Number of workers (in range of situations including manufacture, maintenance, and use)	Number of workers exposed
Quantities per media (time dimension of release, type of release, and uncertainties in estimates)	Substitution; interactions
An indication of measured exposure levels (expressed in an appropriate statistical form, e.g., geometric mean and standard deviation)	Technical feasibility of controls, quality of available data, severity and type of adverse health effects
An overview of monitoring data in the environment (with specifications of conditions)	Availability of sampling and analytical methods
Any additional information that will help to focus the exposure assessment	Other considerations

Table 2-2: Overview of Base Set Testing Requirements: 6th Amendment vs. 7th Amendment to the Dangerous Substances Directive

6th Amendment Annex VII (trigger: 1 ton/year)	7th Amendment Annex VII, Part A (trigger: 1 ton/year or 5 tons cumulative)
UV, IR, NMR spectra	UV, IR, NMR spectra
Methods of detection	Methods of detection
Melting point	Melting point
Boiling point	Boiling point
Relative density	Relative density
Vapor pressure	Vapor pressure
Surface tension	Surface tension
Water volubility	Water volubility
Fat volubility	—
$P K_{ow}$	**PK_{ow}**
Flash point	Flash point
Flammability	Flammability
Explosive properties	Explosive properties
Auto-flammability	Self-ignition temperature
Oxidizing properties	Oxidizing properties
—	*Granulometry (particle size distribution)*
Acute oral LD_{50}	Acute oral LD_{50}
Acute inhalation LC_{50} or acute cutaneous LD_{50}	Acute inhalation LC_{50} or acute cutaneous LD_{50}
Skin irritation	Skin irritation
Eye irritation	Eye irritation
Skin sensitization	Skin sensitization
28-day sub acute study	Repeated dose toxicity (28-days)
Ames assay	Mutagenicity, bacterial (reverse mutation) test
Non-bacterial mutagenicity	Chromosomal aberration or damage
	Toxicokinetic behavior assessment
—	*Reproductive screening test*
Fish acute toxicity	Fish acute toxicity
Daphnia acute toxicity	Daphnia acute toxicity
—	*Algal growth inhibition*
	Bacterial inhibition
Biodegradation	Biodegradation
Abiotic degradation	Abiotic degradation
—	*Adsorption/desorption screening test*

Developmental/reproductive endpoints, extra mutagenicity, toxicokinetics, and environmental repeated dose and bioaccumulation are added.

For volumes greater than 1000 tons/year 5000 cumulative, Level 2 of testing is activated (table 2-4). This includes chronic effects (including carcinogenicity, second species developmental, further toxicokinetics, organ specificity, and other species in the ecological sphere such as birds and other fishes. Besides production volumes, it is important to consider the potential for human exposure when deciding on testing strategy. Although specific quantitation of this aspect has not been incorporated into a specific regulation, it

Table 2-3: Additional Requirements with 100 to 1,000 Tons Annual Production and 500 Tons Cumulative Production

Level 1 Attachment 8 (Complementary tests)
• Further physical & chemical testing
• Fertility study (one or two generation)
• Teratogenesis (one species)
• Subchronic and or chronic toxicity study
• Additional mutagenesis
• Screening for carcinogenesis
• Toxicokinetics
• Prolonged toxicity to Daphnia
• Phytotoxicity (higher plant)
• Bioaccumulation (preferably fish)
• Further degradation tests (if poor degradation found)
• Further absomtion/resorption

Table 2-4: Additional Requirements with Greater than 1,000 Tons Annual Production or 5000 Tons Cumulative Production

Level 2 Attachment 8 (Complementary tests)
• Fertility (3 generation)
• Chronic toxicity
• Carcinogenicity
• Teratogenesis (2nd species)
• Further toxicokinetics
• Complementary studies to determine organ toxicity
• Complementary tests on accumulation, degradation mobility and absorption/desorption
• Complementary tests on fish
• Toxicity to birds
• Complementary tests on other organisms

Table 2-5: Corollaries of Descriptive Toxicity Testing

- Tests are designed to *characterize* toxic effects
- Tests are not designed to *demonstrate safety*
- There are no set of tests that have to be conducted on *every* chemical-in-commerce
- Tests are dependent upon:
  - Use
  - Potential exposure
  - Chemical/physical properties
  - Structural analogs

Table 2-6: Components of Descriptive Toxicity Study

- Relevant route of administration
- Primarily conducted with rats or mice
- Typically three or more dose levels, plus controls
- Includes a batter of clinical observations
- Clinical pathology
- Necropsy
- Microscopic evaluation of tissues
- May include other special tests

should not be disregarded. For example, a large volume chemical used only as a site limited intermediate has limited potential for human exposure, while a low volume chemical which becomes a component of a consumer product has a significant potential for human exposure and, therefore, a likely different testing strategy.

ANIMAL TESTING

Before addressing some of the issues pertaining to testing strategy, I would like to describe the simple, acute toxicity testing in mammals (table 2-5).

For the simple descriptive tests, there are basic components (table 2-6). Over the years there have been clear improvements. The first one is the characterization of the test material. For pure compounds purity can be in the 99+ % range. For technical grade material, the amount and content of each component can be determined to very low levels such as ppm. In general, impurities at >0.1% must be identified. Another important feature is stability of the material, knowing that composition doesn't change from time of manufacturer to completion of tests is critical.

The test species that are used: rats, mice, guinea pigs, rabbits and dogs have become more consistent; one can select a particular strain of animal and generally they remain healthy during the course of the study. Similar upgrades with emphasis on consistency and quality are found in animal food, water, housing, and lack of disease.

Table 2-7: Subacute Tests

- Range finding for subchronic studies
  - Potential target organs
  - Palatability/application limits
  - Oral, dermal, and inhalation
- Primarily in the rat, mouse, (rabbit, dog)
- 5/sex/dose, 3 dose levels plus control
- Limited clinical pathology
- Gross and partial histopathology
  - Liver, kidney, lung, skin, and gross lesions

Table 2-8: Subchronic Tests

- Most complete "short-term" studies
- Primarily in the rat, mouse, rabbit, and dog
- Range-finding for chronic studies
  - Identifies target organs
  - Oral, dermal, and inhalation
- Rodents: 10/sex/dose, 3 dose levels plus controls
  - Recovery groups
- Full in-life clinical observations
- Full clinical pathology
- Gross and full histopathology (top dose versus control)

For acute toxicity testing data development there have been improvements as well. Two examples are worth mentioning. In the classic evaluation for lethal effects, the LD50 and LC50 are considered as the basic values. However, this exact value is being superseded by the Approximate Lethal Dose (ALD) and Approximate Lethal Concentration (ALC). These latter two values provide enough information for most determinations and the fine tuning to an LD50 has become less important (2). Also ALDs and ALCs use fewer animals thus addressing some animal welfare concerns. Further development in this area has been the use of a limit dose test, which is aimed at getting basic information. OECD recommends a dose of at least 2,000 mg/kg; if mortality is observed a full study could be done (5). Lethality data remains important, but this starting point may not need to be so well defined.

The second example is the determination of irritation. In the classic paradigm, there was dermal testing in rabbits and eye testing in rabbits. Dermal testing remains generally unchanged, but significant changes in eye testing strategy have been made. The dermal response in rabbit is used to determine whether or not an eye test will be conducted. Even when eye tests are conducted, less material is used, 0.1 ml vs. 1 ml amounts. Furthermore, if the pH is less than 2 or greater than 11.5 eye testing is not performed, because it would be painful to the animal and results, based upon past experiences, would most likely show the material to be at least a severe irritant.

Beyond acute studies other endpoints are only addressed with multi-dose and exposure studies.

The subacute and subchronic tests have not changed format significantly (tables 2-7 and 2-8) Their main use is to determine multiple exposure effects, i.e. cumulative toxicity and begin to focus on identifying target organs. The major evaluation improvements over acute studies are whole animal observations while under test and histopathological evaluation of tissues for microscopic effects. The data collected are shown in table 2-9, along with the data analysis, and data interpretation.

An example of detailed organ evaluation is shown for two commonly examined tissues - liver and kidney (table 2-10). Besides microscopic evaluation, cellular enzymes and other cellular components are analyzed. A specific for these organs can be evaluated. The additional end points shown in this table are not usually done routinely, but often become part of a research

Table 2-9: Data Collection and Analysis

Data Collection

- Measured endpoints
- Clinical signs
- Body and organ weight
- Hematology
- Pathology
  - Gross lesions
  - Histopathology

Data Analysis

- Statistics and multiple comparisons
- Good Laboratory Practices

Table 2-10: Multiple Endpoints for a 90-day Study

Target Organ	Core endpoints	Additional endpoints
Liver	Histopathology Clinical chemistry ALT, AST, AP, glue, chol., bili, prot., alb., GGT, triglycerides Organ weight Clinical signs	Additional stains Electron microscopy Cell proliferation Enzyme levels
Kidney	Histopathology Clinical chemistry BUN, protein, electrolytes Urinalysis protein, pH, specific gravity Organ weight	Additional stains Electron microscopy Cell proliferation Enzyme levels

investigation examining mode or mechanism of action.

The overall above discussion briefly describes the basic mammalian evaluation of an existing chemical. The chronic exposure component for which all practical purposes is aimed at determining whether or not the chemical has carcinogenic potential is beyond the scope of this discussion.

ISSUES

The second portion of this document addresses several issues which have emerged with the implementation of TSCA and have an impact on how testing is carried out.

The following tables were prepared to give a view of costs for conducting the various studies (table 2-1 1). For comparison purposes, a previous publication in 1973 (3) shows that costs of toxicity tests were about one-tenth of what they are today. This averages out to over a 10% increase on an annual basis. Table 2-12 shows how long a study needs to be run, ie. exposure duration in order to develop adequate information on a particular endpoint and table 2-13 addresses the various non-cancer endpoint of general interest. From these two tables, duration of exposure of at least 28 day provides appropriate data in the rat model. The 14 day study is likely insufficient in duration to reach a steady state for metabolism and lesion development and the 90 day study may not provide that much more information. A 90 day study covers the male rat sperm cycle of about 60 days, whereas 28 days might have limitations. These tables are useful in helping decide what duration of testing should be considered for a chemical and raises a fundamental issue in experimental design which focuses on length of exosure. Thus, we find that a 28 day study would

Table 2-11: Cost of Various Laboratory Studies for Acute and Chronic Toxicity

Test	cost ($1000)
Acute battery • Oral LD50 • Inhalation LC50 • Dermal LD50 • Eye Irritation • Skin Sensitivity	30-40
Mutagenicity battery • Ames • CHO/HGPRT • Mouse micronucleus	40-60
Repeated exposure • oral, dermal, or inhalation	35- 90*
Subchronic • oral, dermal, or inhalation	120- 200'
Metabolism	50-250
Developmental	120-160
Reproduction	350-500
Chronic/Oncogenicity	600-1200

*Costs for each route of exposure

Table 2-12: Duration of Test Exposures Needed to Generate Adequate Information on Various Endpoints

	Exposure duration (with rats)			
Endpoint	**Acute**	**14 Day**	**28 Day**	**90 Day**
Lethality	+	+	+	+
Clinical signs	+	+	+	+
Toxic signs time course	-	+	+	+
Body weight and food data	-	+	+	+
Hematology		+	+	+
Gross necropsy	*	+	+	+
Clinical chemistry	-	+	+	+
Histopathology	*	+	+	+
Target organ		*	+	+
Dose-response	-	+	+	+

+ = Test needed; - = Test not needed; ± = Test maybe needed

be an appropriate screen for longer term, dose level selections for special studies and perhaps more importantly for describing systemic toxicity.

A previously publication comparing 6 month studies to longer term ones, suggests that 6 months is adequate for identifying non-cancer endpoints vs. 12 months or longer studies (1,4). An up-to-date comparison of 14 vs 28 vs 90 day duration studies is needed so that a data base is established to make case for shorter duration studies, with the caveat that a comprehensive 28 day study can be sufficient for determining repeated exposure effects.

The implementation of the Good Laboratory Practice (GLP) standards under TSCA published in FR November 19, 1983 and revised in 1989, FR August 17, 1989, added several layers of details, especially documentation of procedures and protocols to the conduct of studies under TSCA with GLPs. However, these efforts add significantly to the testing costs, measured in time, required to complete a study, amount of documentation, and quality assurance compliance activities. An unintended side effect of this burden has been a tendency to raise the threshold for deciding whether to conduct a study which, for screening tests, could be counter-productive.

Table 2-13: Test Exposures Needed to Generate Information on Various Non-Cancer Endpoints

	Exposure duration (with rats)			
Endpoint	**Acute**	**14 Day**	**28 Day**	**90 Day**
Neurotoxicity	±	+	+	+
Immunological	-	±	+	+
Reproductive		±	+	+
Pharmacokinetics	-	+	+	+
Mutagenicity	-	+	+	+

+ = Test needed; - = Test not needed; ± = Test maybe needed

However, with GLP in place the overall quality of studies is increased in a way measured by having sufficient detail to available to describe all aspects of the study.

An issue which we raise are the costs (defined both in absolute dollars, as well as utilization of finite resources) of doing testing under TSCA. This is tied in with a second issue of data gaps versus data needs.

As we have heard from several discussions during this conference, there are approximately 60,000 chemicals on the TSCA inventory and about 10,000 of those materials are out in commerce. Of that 10,000 approximately 1,000 have sufficient toxicity data developed to make judgment for risk assessments, although there is considerable variability in the amount and type of information.

The number of additional endpoints which can be added on to any study will have limitations. Two items that begin to put boundaries on add-ons are the ability to properly manage the logistics of the study and secondly, the interpretation of the data. The first item will be affected by GLP's and the second brings into light whether the experimental design was appropriate for addressing the particular toxicity endpoint.

We would also like to share some experiences with you concerning TSCA Section 4 test programs. These programs require significant amounts of time to develop and once finalized their implementation is attached to a timeline for completion and submittal of reports to the appro-

priate TSCA office. Once received by the agency internal review takes place, but the timeline for that is not known to the parties responsible for conducting the studies. Generally, there is no conclusionary meeting between the agency and parties affected about the study results and any associated issues which may have been raised by those studies. Examples of this are also found with TSCA Section 4 test rules, such as triethylene glycol ethers, cyclohexanone and the phenylene diamines. We would propose that data interpretation and recommendations should be discussed by both parties so that there is a value beyond than just completing the test requirements. The opportunity for this has been consistently missed.

CONCLUSION

Only when properly designed and conducted toxicity studies are carried out, can effective strategies for predicting potential hazards of chemicals be realized. We believe that simple, acute studies, followed by 28 days multi-dose rodent studies by appropriate exposure routes can address the majority of toxicity endpoints for further prioritization of testing. As part of the initial strategy, the selection of existing chemicals under TSCA should include evaluation of potential human exposure and production volume triggers.

REFERENCES

1. Betton, G. et al., *Human and Experimental Toxicology* 13:221-232, *1994.*
2. Brown, V.K.H., "Acute Toxicity Testing - a Critique," Chapter 3 in *Testing for Toxicity,* J.W. Garrod, ed. (Taylor and Francis Ltd., 1981).
3. Ghering, P.J., Rowe, V.K. and McCollister, S.B. *Food and Cosmetic Toxicology* 11:1097-1110, 1973.
4. Lumley, C.E. and Walker, S.R. Reg. *Toxicol. Pharmacology* 6:66-72, *1986.*
5. *Offical Journal of the European Communities,* L383QA, V35, Dec 29, 1992

3

Carcinogenicity

James T. MacGregor, Barbara S. Shane, Judson Spalding, and James Huff

The objective of the carcinogenicity and genotoxicity assay working group was to assess the methodology available for predicting and identifying the human carcinogenicity of chemicals that are subject to review and testing under the Toxic Substances Control Act (TSCA), with emphasis on existing chemicals. Specific objectives were to 1) identify existing assays for assessing or predicting the human carcinogenicity of chemicals, 2) indicate which assays are the most reliable for the prediction of human carcinogenicity, and 3) assess the reliability of low cost predictive assays. This mandate included an assessment of genotoxicity assays in the context of their predictive value for carcinogenicity, but did not include an assessment of other health implications of genotoxicity. The objectives are summarized in table 3-1.

ASSAYS IN HUMANS

Epidemiology

Human studies are extremely valuable because they measure the endpoint of concern directly, i.e., induction of human cancer by exposure to chemicals or environmental agents (80, 99). However, such studies have major limitations that restrict their applicability in the context of the TSCA mandate to protect human health by preventing exposure to those chemicals that pose the greatest potential for inducing cancer under the conditions of their actual use. The most significant limitation is that they cannot be used to identify potential carcinogenic agents before exposure occurs, because human epidemiologic studies can be conducted only after sufficiently large populations have been exposed. They are, however, extremely important for assessing the health impacts of *existing* exposures in human populations.

Disadvantages of epidemiologic studies are their relative insensitivity and the difficulty of proving causality. Epidemiologic studies are always subject to uncontrolled factors that can confound the interpretation. They are also expensive and time consuming. Their power is greatest when combined with results from laboratory data that demonstrate similar effects under more rigorously controlled conditions. Finally, such studies can be conducted only when it is possible to identify a reasonably large defined population with a documented exposure to specific agents, and such populations are often difficult or impossible to identify.

Genetic Biomarkers of Cancer

Major strides have been made in determining the molecular basis of human cancer, and this knowledge may soon lead to a greatly improved ability to monitor the induction of cancer in human populations and individuals. Specific molecular alterations in DNA have been associated with certain human cancers (13), including mutations that activate cellular oncogenes or inactivate tumor suppressor genes. Translocations at specific chromosomal sites are also believed to activate or inactivate key genes in the process of cancer development.

As inexpensive methods for monitoring these molecular changes become available and the

The authors appreciate the comments of Dr. Michael Cimino and Dr. Ann Richard, who contributed to the sections on current TSCA testing requirements (MC) and structure-activity analysis (AR). The contributions of these individuals do not represent an official position of the EPA.

- Assess current assays for predicting and identifying human carcinogenicity of chemicals subject to TSCA review
  - Identify the most reliable assays
  - Evaluate low-cost predictive assays and assess their reliability
- Consider the predictive value of genotoxicity assays for carcinogenicity, but not other health effects related to genotoxicity

exact mechanisms of their role in cancer progression become known, studies of these molecular changes in humans should allow more effective monitoring of exposed human populations. At present, however, such methods are not yet practical for studies of chemically induced cancer in humans. The advantages and limitations of direct studies of carcinogenicity in human populations are summarized in table 3-2.

Biomarkers for Nonspecific Genetic Damage

Because genetic alterations in DNA are believed to be involved in the initiation phase of cancer, monitoring genetic damage in "reporter" genes in humans is a useful method of monitoring exposure to genotoxic agents. Demonstration that exposure to chemicals can induce genetic damage in humans is strong evidence that the exposure may pose a significant carcinogenic risk, even when the monitored genetic loci are not necessarily involved directly in the process of carcinogenesis. For those chemicals known to induce cancer via genetic damage (i.e., for genotoxic carcinogens), monitoring genetic damage in humans provides a means of estimating the carcinogenic risk associated with a particular exposure.

Table 3-3 lists biomarkers that have been used to monitor genetic damage in humans. Details of the available methodologies and summaries of their applications have been reviewed (24, 38, 54, 62). These biomarkers provide direct measures of the types of genetic damage believed to be involved in the induction of carcinogenesis, but they are more useful when the mechanism of action of the chemical under study has been elucidated in laboratory studies. When the mechanism of cancer induction by the chemical under study is known, it is possible to estimate risk quantitatively by using these surrogate biomarkers of damage *in lieu* of direct measurements of cancer incidence.

The most widely used of these biomarkers is cytogenetic damage, or the occurrence of chromosomal aberrations. Micronuclei serve as an alternate screen for chromosomal aberrations or cellular chromosome loss. Several methods now

Table 3-2: Direct Assessment of Human Carcinogenicity

- Provides direct evidence of human carcinogenicity
- Restricted to epidemiologic studies, which have important limitations
  - Difficult to prove causality
  - Insensitive
  - Expensive
  - Time consuming
  - Requires a large defined population with documentable exposure
  - Retrospective (damage already incurred)
  - Not practical for screening
- In the near future, assays for unique damage in oncogenes or tumor suppressor genes involved in cancer development may facilitate human studies of cancer induction

Table 3-3: Human Biomarkers of Exposure and Risk

- Available biomarkers
  - Cytogenetic damage (chromosomal aberrations, micronuclei, aneuploidy)
  - Mutations in surrogate tissues(e.g., *hprt,* glycophorin A, HLA-A, T-cell receptor mutations)
- Mutations in oncogenes or tumor suppressor genes
  - DNA adducts
- Uses and Limitations
  - Useful for monitoring exposures to known genotoxic carcinogens
  - Useful for assessing population risk following exposure to genotoxic agents
  - Indirect relationship to carcinogenesis limits interpretation (esp. for individual)
  - Not practical for assessing risk from uncharacterized agents or tumor promotors

allow direct measurement of mutations in human populations, but an important limitation of the mutation and cytogenetic studies in humans is that damage can be measured in only a few tissues – principally cells that can be obtained by blood sampling. Available assays include the measurement of *hprt*, HLA-A, and T-cell receptor mutations, mutations in lymphocytes, and glycophorin A mutations in erythrocytes. The mechanisms of cancer induction are now beginning to be understood at the molecular level, and in a few cases it is already possible to measure mutations in oncogenes and suppressor genes that are believed to be linked directly to carcinogenesis. Unfortunately, the ability to apply these assays to human populations is extremely limited because, in general, it is not possible to obtain samples of the many different tissues in which cancer may arise.

Sensitive new methodologies have been developed for measuring the interaction of the chemical with DNA to form adducts. The most notable are ^{32}P-postlabeling (75) and immunological methods (50, 51) for specific DNA abducts. In general, these assays are also limited to accessible tissues. They are useful for monitoring exposures when the chemical interacts directly with DNA, and they are useful for estimating the risk associated with exposure to agents that have already been characterized as genotoxic carcinogens. The major limitations of these assays are that 1) they are limited to a few tissues, 2) the endpoints are related only indirectly to cancer (cancer is a multistep process and many defense systems can modify the progression of damage that leads to cancer), and 3) certain of the assays are specific to individual chemicals (e.g., immunoassay).

ANIMAL CARCINOGENICITY BIOASSAY

Assessment of cancer risks to humans can, in practice, be conducted only for a small fraction of the chemicals subject to TSCA regulations. Therefore, these assessments are most often based on data from laboratory animal studies (35). Among the available laboratory tests for assessing carcinogenic potential, the rodent cancer bioassay is generally considered the most reliable predictor of human cancer hazard (37). Virtually all the known human carcinogens are carcinogenic in animals, and those characterized as potent DNA-reactive (genotoxic) carcinogens show excellent interspecies concordance (3, 84). These highly reactive genotoxic carcinogens are generally potent multisite carcinogens, and are generally considered to be the most hazardous class of carcinogens. Approximately one-third of the known human carcinogens were first discovered to be carcinogenic in animals and were later shown to be carcinogenic in humans (36, 98).

One of the major advantages of direct animal cancer bioassays over other shorter term predictive assays is that the animal model is closely related to the human in terms of anatomy, absorption, metabolism, uptake, and pharmacokinetics as well as in the histology of the tumors in various tissues. Thus, it is possible in this

Table 3-4: Advantages and Limitations of the Rodent Carcinogenicity Bioassay

Advantages

- Most reliable method of predicting human carcinogenicity
- Most known human carcinogens are carcinogenic in animals
- Interspecies concordance is good, especially for multisite carcinogens
- Approximately 3070 of human chemical carcinogens first identified in animals
- Can relate effect to exposure for quantitative extrapolation to humans (if dose-response determined)
- Biologic model closely related to human (absorption, metabolism, uptake, pharmacokinetics)

Limitations

- Expensive and lengthy (5 years and >$1.0 M/rodent species)
- Metabolism, pharmacokinetics, tissue defenses, and DNA repair responses may differ quantitatively from human
- Impractical to evaluate low-doses or multiple exposure routes (high-dose data maybe misleading)
- Quantitative extrapolation to humans (and other species) is imprecise
- High-dose bioassay can cause cancer by species - and/or tissue-specific mechanisms not relevant to humans (e.g., -2-globulin nephropathy, saccharin-induced bladder tumors, halogenated organics-induced mouse liver tumors)

model to make quantitative dose response predictions. Extrapolation from rodent cancer bioassay data is the method used by most regulatory agencies, including the EPA, to quantitatively assess human risk. The mathematical model and assumptions involved in such extrapolations have been summarized recently by Fan and Howd (18).

The rodent cancer bioassay also has some important limitations, principally time and expense. A rodent bioassay requires approximately 5 years and costs $500,000-700,000 even with straightforward exposure regimens in a single species. Specialized expertise and laboratory space are also required.

Although the rodent provides a relatively reasonable model for the processes of metabolism, pharmacokinetics, tissue defense, and tumor development, there are often significant quantitative differences in these parameters between the human and the rodent. Metabolic pathways and kinetics often differ between humans and rodents. These important differences may influence the potency of the carcinogen and the shape of the dose response curve (47).

Expense and time restraints make it impractical to. evaluate such important factors as the total shape of the dose response curve (in particular, the low dose portion of the response curve) and limit the range and type of exposures and other experimental variables that can be evaluated. Consequently, the quantitative extrapolations of risk to humans made from bioassays are often imprecise.

Additionally, cancer development in animal models can involve mechanisms that are unlikely to occur in humans. For example, the genotype of the rodent strain can influence the chemical effect, or the high dose regimens used in animal studies can lead to an increased tumor frequency due to enhanced cell proliferation (1, 34, 61, 91, 92, 102). Some well known examples include alpha-globulin nephropathy (unique to male rat kidney) (90), saccharin-induced bladder tumors (due to crystallization in the bladder at very high doses) (14), and the high incidence of liver tumors observed in mouse liver after treatment with certain agents (such as organochlorine compounds) (57, 60). The major advantages and limitations of the rodent cancer bioassay are summarized in table 3-4.

PREDICTIVE METHODS AND MODELS

The rodent assays may provide the most reliable prediction of human carcinogenic hazard, but time and cost factors limit the number of chemicals that can be evaluated in these systems. Therefore, much effort has been devoted to developing low-cost short-term assays that can be

Table 3-5: Predictive Methods and Models

- Chemical structure-activity analysis and mechanism-based inference
- *In vitro* genotoxicity and cell transformation assays
- Short- or mid-term animal models
  - Accelerated tumor development models
    - Strain A mouse
    - TG.AC mouse
    - p53+/- mouse
    - DNA repair-deficient mice
    - Initiated short-term model
  - Biomarkers of preneoplastic tissue growth
    - e.g., -glutamyltranspeptidase-positive preneoplastic foci
- *In vivo* genotoxicity assays
  - Endogenous or transgenic reporter genes for mutation detection
  - Chromosomal aberrations and aneuploidies/micronuclei
  - Unscheduled DNA synthesis (UDS)
  - DNA adducts
  - DNA strand-breaks
  - DNA damage-inducible genes

used to predict carcinogenic potential. Table 3-5 summarizes the major types of predictive methods and provides examples of some important assays.

Structure-Activity Analysis and Mechanism-Based Inference

Structure-based models for predicting chemical carcinogenicity attempt to take advantage of the currently available rodent bioassay carcinogenicity data, which represent hundreds of millions of dollars in testing investment. Carcinogenicity is one of the most complex, yet most widely studied toxicity endpoints from the perspective of structure-activity relationships (SAR) and structure-based mechanism inference. The unifying mechanistic paradigm underlying much of this work is the electrophilic theory of chemical carcinogenesis, which proposes that genotoxic chemical carcinogens form reactive electrophilic intermediates that intercalated, adduct, or otherwise alter or damage DNA (63).

Several models for predicting carcinogenicity based on chemical properties or biochemical indicators have been developed for noncongeneric chemicals, i.e., diverse chemical structures. One published prediction method is based on an experimentally measured, biochemical indicator of electrophilicity, i.e., an electron attachment rate constant (7). COMPACT, a computer-based prediction method, models oxidative P-450 metabolism of a chemical in terms of calculated structural and electronic features as a presumed condition for formation of a reactive electrophile (46). Other computerized prediction programs rely on statistical "discovery" of chemical features significantly associated with carcinogenicity, where model predictions are based on the presence or absence of chemical fragments (CASE) and/or values of calculated molecular properties (TOPKAT, ADAPT) [for reviews, see 22,45, 76].

Ashby (2) has formulated a list of "structural alerts" for use in predicting chemical carcinogenicity, i.e., structural features that are likely to be associated with formation of electrophilic intermediates and whose presence in a molecule provides an alert to potential carcinogenicity. Tennant et al. (95) used such alerts, in conjunction with available short-term test data, subchronic toxicity data, and organ pathology data from the rodent bioassay, as the basis for an "expert intuition" approach to carcinogenicity prediction.

In a recent National Toxicology Program prospective prediction exercise (NIT-44), this approach had an accuracy of >80% in predicting the outcome of the NTP rodent bioassay of 44 previously untested chemicals—an accuracy better than many short-term bioassays (4). This performance was judged significantly better than the 55%-70% accuracy achieved by the computerized SAR models that participated in the exercise, e.g. TOPKAT, CASE and COMPACT (32, 101). However, there were some important distinctions between these approaches: the "expert intuition" approach was neither automated nor easily applied by "non-experts" and required biological test data, whereas the statistical approaches were fully automated and based on information pertaining only to chemical structure.

The special problems and issues surrounding the application of statistical SAR approaches to the modeling of noncongeneric carcinogenicity data bases have been discussed (2, 76). Computerized SAR models such as CASE, TOPKAT, and COMPACT have the advantages of wide applicability, no requirement for availability of the chemical or biological testing, and the potential for generating insight into mechanisms of carcinogenicity. However, due to the dissociation of the results from mechanistic interpretation and the large uncertainties associated with prospective predictions, such methods are not currently in use by EPA for TSCA screening, and the use of such methods in isolation from expert judgment and oversight is not recommended.

On the other hand, SAR and mechanism inference have contributed greatly to understanding of the molecular basis for chemical carcinogenicity when applied to more narrowly defined classes of carcinogens, such as PAHs, nitroaromatics, or certain PCBs. Chemical class, mechanism-based SARs applied with expert judgment are relied on heavily within EPA for carcinogenicity screening of a wide range of new and existing TSCA chemicals (107).

ONCOLOGIC is a recently developed, computerized expert system for carcinogenicity prediction that represents an ambitious attempt to reproduce the cancer prediction expertise of the Structure-Activity Team within EPA. For the chemical classes covered in ONCOLOGIC, early indications are that this method accurately reproduces such expertise.

Hybrid approaches to modeling chemical carcinogenicity data are now being used more extensively in an effort to improve prospective prediction accuracy (see e.g. 6, 12). Inclusion of short-term bioassay data or subchronic toxicity data, when available, provides elements of the complex biological interaction that may be difficult to model by structure alone. In the absence of such data, approaches such as ONCOLOGIC that combine elements of successful SARs, current knowledge of mechanisms, and human judgment appear most promising. The accuracy of such models is expected to improve over time with increased understanding of mechanisms of carcinogenesis and with SAR model refinement resulting from additional data and prospective prediction exercises. See the chapter on SAR/modeling in this volume for a more in-depth discussion of the general requirements and limitations of SAR modeling for use in toxicity prediction.

In Vitro Genotoxicity and Cc// Transformation Assays

In the early 1970s there was great enthusiasm that *in vitro* assays for genetic damage would be an effective and inexpensive means of identifying the carcinogenic potential of chemicals. This confidence was spurred by reports that mutagenicity in *in vitro* assays that incorporated mammalian metabolic enzymes for the activation of metabolically dependent electrophilic carcinogens was an excellent predictor of carcinogenicity *in vivo (59).* Unfortunately, later studies that incorporated larger numbers of chemicals showed that the overall concordance (agreement between the tests in both positive and negative results) between carcinogenicity *in vivo* and mutagenicity *in vitro* was considerably less than that observed in the more limited early studies, i.e., approximately 59%-66% rather than the 90°/0 implied by earlier studies (94, 109). However, the *predictivity* (percentage of agents positive in mutagenicity

assays that are carcinogenic) of positive responses in assays such as the Ames *Salmonella* mutation assay has been found to be quite good, approximately 89°/0. Thus, agents found to be mutagenic in multiple short-term assays are quite likely to prove carcinogenic, but a lack of mutagenicity in *in vitro* assays does not provide strong assurance of noncarcinogenicity (93).

Nonetheless, those carcinogens that are inherently DNA-reactive and are potent multisite carcinogens *in vivo* (those believed to be the greatest carcinogenic hazard) are generally mutagenic in *in vitro* genotoxicity assays. *In vitro* assays are therefore an extremely useful means of identifying potential carcinogens, but there is a substantial risk of misclassification.

In addition, induction of cancer is mechanistically complex, and *in vivo* factors such as metabolism, pharmacokinetics, and tissue specific defenses and proliferation rates often result in marked tissue specificity of carcinogenesis *in vivo. The* complexity of these *in vivo* factors generally makes it impossible to obtain reliable quantitative estimates of the human carcinogenic risk based solely on data from *in vitro* assays (52, 55). These assays therefore are best used as screening assays to provide an initial qualitative assessment of potential carcinogenic hazard.

In vitro mammalian cell transformation assays have undergone extensive study for screening chemicals for potential carcinogenic activity. Among transformation assays, the primary Syrian hamster embryo (SHE) cell assay and the BALB/c 3T3 mouse embryo cell line have been the most extensively used for identifying the potential carcinogenic activity of chemicals.

A change in morphological phenotype is the measured endpoint for chemical activity in both assays, although the transformed phenotype must ultimately be related to the ability of the cells to produce neoplastic growth in suitable recipient animals. The relationship between cellular transformation and genetic alterations is now becoming clear, and suggests an important role for such assays (9, 10, 1 1). Retrospective studies that evaluated the activity of carcinogens and noncarcinogens that had previously been identified in the National Toxicology Program's standard two-year rodent bioassay indicated that positive activity in the transformation assays was most highly correlated with electrophilic and/or mutagenic activity (42, 58).

Although the mammalian cell transformation assays are able to correctly identify nonelectrophilic carcinogens, more work is required to. develop protocols that can correctly discriminate between nonelectrophilic carcinogens and noncarcinogens. The report by Matthews et al. (58) indicated that the BALB/c 3T3 transformation assay was able to discriminate between nonmutagenic *(Salmonella-* negative) carcinogens and noncarcinogens and thus complement the *Salmonella* mutagenicity assay. Similar results were reported for the Syrian hamster embryo assay (23,74). All these results support the conclusion of Swierenga and Yamasaki (91) that cell transformation assays appear to respond to both genotoxic and nongenotoxic carcinogens. The SHE cell transformation system has been used successfully to detect several chemical carcinogens that are not typically identified in short-term assays and often have been considered to be nongenotoxic. These include diethylstilbestrol, 17-estradiol, asbestos, amitrole, arsenic, and reserpine.

However, for the purposes of screening, additional validation studies would be required before these assays could be recommended.

Accelerated Tumor Development Models

The development of tumors in humans and other animal species is the culmination of a multistage process. It is believed that the contributing components of this process are multiple gene mutations in cellular protooncogenes, loss of tumor suppressor gene function, alterations in the regulation of gene expression, and the "time factor", which can be one-half to two-thirds of the human/animal lifespan. The sequence of these events in the development of specific tumors is unknown.

Mutations in the family of cellular ras protooncogenes and mutations or loss of function

in the tumor suppressor p53 gene are associated with a wide range of human tumor types and chemically induced or spontaneous tumors in animal models. The TG.AC transgenic mouse line, which carries inactivated v-Ha-ras gene, has the properties of genetically initiated skin, and skin papillomas are readily induced by promoting agents as well as mutagenic and nonmutagenic carcinogens (43, 88, 89). A mouse line deficient in the p53 gene has been shown to be sensitive to mutagenic carcinogens, and chemically induced tumors are detected as early as 20-26 weeks of treatment (15, 16, 21). Another approach is to "initiate" the animals by pretreating with multisite carcinogens (27). The results from the first phase of a validation study performed in the TG.AC and p53 (+/-) mouse models have been summarized recently (96). These validation studies in the TG.AC transgenic and heterozygous p53 (+/-) knockout mouse lines are being extended by the NTP to determine their ability to discriminate between known rodent carcinogens and noncarcinogens.

The advantage of using transgenic or gene deficient animal (mouse) models to evaluate chemicals for their potential carcinogenic activity is that the presence of an altered gene (oncogene) or absence of a specific tumor suppressor gene can significantly reduce the "time factor" required to observe a tumorigenic effect (from 72-96 weeks to 20-26 weeks). With these models, fewer animals are required per treatment group, and the cost and time required to determine the tumor endpoint are reduced significantly.

Although models of this type have the potential to shorten both the time required and the expense of the traditional rodent cancer bioassay, these models are not sufficiently validated and cannot replace the traditional bioassay at this time.

Biomarkers of Preneoplastic Tissue Growth

A number of biochemical markers are presently being used to monitor neoplastic disease in humans. One example is the prostate specific antigen (PSA). An elevation in the level of this antigen in the blood indicates cell growth that may lead to prostate cancer. As far as is known, this marker is specific for prostate cancer (44, 108). In contrast, three other detectable markers are released from various tumors. Carcinoembryonic antigen (CEA) is elevated in patients with cancer of the breast (82), lung, kidney, pancreas, stomach, and colon (8, 105); CA 125 indicates ovarian, pancreatic, breast, and lung cancer (83), and -fetoprotein (ALF) is indicative of liver and lung cancer (19) and, to a lesser extent, ovarian and testicular cancer. However, CEA can be increased in the serum of patients with benign tumors (79) and varies widely in the human population, thus resulting in a high incidence of false positives.

More recently, attention has focused on the p53 gene, which is involved in restricting cell replication following DNA damage (28, 29, 33, 87). Mutations in p53 alter its normal function and can result in uncontrolled replication of damaged cells. Although, p53 is mutated in 50°/0 of tumors from various tissues, it does not indicate a specific type of cancer (77). This marker may have more relevance in monitoring the prognosis of cancer because mutations in p53 have been correlated with metastatic cancer (48).

In Vivo Genotoxicity Assays

During the last decade, major advances have been achieved in the technology of measuring mutations, chromosomal interchanges, and aneuploidy in tissues *in vivo.* These advances have provided sensitive and rapid methods of measuring the key genetic endpoints associated with neoplastic development in any tissue of interest in laboratory animals. These methods include development and validation of transgenic and endogenous reporter genes that allow direct measurement of mutation in tissues from animals (reviews by 25, 54, 64), improved methodologies for measuring *stable* (balanced) chromosomal aberrations and aneuploides (55, 78, 100), improved methods for surrogate markers of DNA damage such as micronucleus induction (53, 56) or DNA strand breakage (97), and damage-specific inducible responses to DNA damage or intracellular oxidative damage (20, 31, 81).

Table 3-6: Correlation Between Mutagenicity and Carcinogenicity in Tissues of B6C3F1 Mice and F344 Rats

			CARCINOGENICITY		MUTAGENICITY *IN VIVO*		
Agent	**Tissue**	**Sex/Species**	**Dose**	**Care. Activ.[a]**	**Dose**	**Muta. Activ.[a]**	**Refs.**
2-AAF	liver	fem. mouse	60 ppm	2-3	75 ppm	2.7	**49, 72**
1,3-Butadiene	bone marrow	male mouse	625 ppm	C	1250 ppm	4.8	**26,69, 71**
Aflatoxin B1	liver	rat	15 ppb	C	0.25 mg/kg	18.3	**73,** 104
Aflatoxin B1	liver	mouse	1000 ppb	NC or WC	2.5 mg/kg	NM	73
Dimethylnitrosamine	liver	male mouse	4 mg/kg	C	4 mg/kg	10	39,40,65
Methylmethane-sulfonate	liver	male mouse	30 mg/kg	NC	20 mg/kg	NM	65
2,4-Diaminotoluene	liver	male mouse	200 ppm	NC/WC	1000 ppm	2	17, 67
2,6-Diaminotoluene	liver	male mouse	200 ppm	NC	1000 ppm	NM	17,68
Benzene	spleen, lymphoma	male mouse	100 mg/kg	C	750 mg/kg	1.7	30, 70
Benzene	lung	male mouse	100 mg/kg	C	750 mg/kg	1.2 (NM)	30
o-Anisidine	liver	mouse	5000 ppm	NC	750 mg/kg	NM	66, 86
o-Anisidine	bladder	mouse	5000 ppm	C	750 mg/kg	2.1	41, 86

[a]Carcinogenic and mutagenic activity given as increase over control or qualitatively: C = Carcinogenic; NC = Not carcinogenic; WC = Weakly carcinogenic; and NM = Not mutagenic.

The development of labeled hybridization probes that are specific for individual chromosomes or regions of chromosomes has made possible the development of assays that can detect stable chromosomal aberrations and aneuploidy. These same methods have also increased the sophistication of *in vivo* micronucleus assays by allowing determination of whether micronuclei arise from chromosome breakage or loss of whole chromosomes (55, 56.). Although data using these assays are limited, they do suggest that these new *in vivo* assays can predict the carcinogenic activity of chemicals more effectively than has been possible with *in vitro* assays.

One major advance is the development of transgenic animal models with "reporter" genes that allow the measurement of mutations in essentially all tissues of the animal. This advance is of major importance because the systems for measuring mutations *in vivo* were previously limited to one or a few tissues, whereas cancer induction is known to be highly tissue specific. These new transgenic models provide the first opportunity to determine if mutations in specific target tissues are correlated with the development of tumors in those same tissues (a necessary feature if the predictive model is to be used in risk assessment).

The potential of one of these new transgenic mutagenesis assays [the "Big Blue" mouse, with a *lacI* reporter gene; (85)] to predict tumorigenesis is illustrated by the data in table 3-6. This table compares the induction of mutations in a "neutral" reporter gene *(lacI)* with induction of cancer in specific target tissues. Excellent quantitative correlation between the induction of mutations and development of tumors is shown when the comparison is made for specific target tissues,

Table 3-7: In Vitro and In Vivo Genotoxicity Assays

- *In vitro assays*
  - Advantages
    - Useful prescreen for potent genotoxic (DNA-reactive) carcinogens (esp. direct-acting)
    - Rapid
    - Relatively inexpensive
  - Disadvantages
    - Overall predictivity for carcinogenicity of nonelectrophilic agents is poor
    - Predictivity not quantitative
    - Does not model uptake, metabolism, distribution, pharmacokinetics *in vivo*
- *In vivo* genotoxicity **assays**
  - Potentially very useful, but not yet sufficiently evaluated for predictivity of carcinogenesis
  - Tissue specificity of carcinogens limits utility of single-tissue assays (e.g., conventional cytogenetics, micronucleus)
  - General models of initiation/promotion paradigm of carcinogenicity are not established

but only a limited number of agents has been tested to date. The agents included in this table show marked selectivity in tissue site and species sensitivities to carcinogenicity, and similar selectivity is observed for mutation induction in these same target tissues.

In contrast, *in vitro* models fail to predict these quantitative selectivities. For example, MMS and DMN both methylate liver DNA to a similar extent at the doses used, yet only the hepatocarcinogen (DMN) induces mutations in liver at carcinogenic doses. This difference is attributed to a differential spectrum of methylated adducts and a markedly higher stimulation of cellular proliferation by DMN, which facilitates fixation of mutations (65). Currently used *in vitro* testing schemes do not adequately predict these types of *in vivo* differences. Thus, these *in vivo* transgenic assays have a strong potential to provide low-cost predictivity of carcinogenicity by genotoxic chemicals. Evaluation of the predictive value of these assays, using a wide range of classes of chemical carcinogens and noncarcinogens, should be given a high priority by funding agencies.

In addition to the above methods already established *in vivo,* it is now known that there are many different mechanisms of repair and control of DNA damage, and that many of the genes that control these responses are inducible. Simple assays to assess many of these responses *in vitro are* already available, and *in vivo* methodologies are being developed. When available, these methods will provide additional indicators of genotoxic damage *in vivo* and will expand our understanding of the nature of genetic damage and repair by carcinogens (56).

Table 3-7 summarizes the advantages and limitations of currently available assays for genotoxic damage *in vitro* and *in vivo.*

In summary, the advantages of *in vitro* assays are their low cost and speed of performance; they have proved useful as screening assays for ranking hazards and are particularly effective at identifying potent DNA-reactive (genotoxic) carcinogens, especially those not requiring metabolic activation. The major disadvantage is that the *in vitro* systems do not model uptake, metabolism, distribution, and pharmacokinetics *in vivo, so* that it is not possible to make quantitative predictions based on them.

The *in vivo* genotoxicity assays are potentially very useful, but these assays have not yet been evaluated systematically to determine their overall predictability for a wide range of chemical classes. Our working group recommends that such a systematic evaluation be undertaken. Assays that are not restricted to specific tissues are expected to be the most valuable, because those that are restricted to specific tissues, such as

the micronucleus assay and conventional cytogenetic analysis, cannot be expected to serve as a surrogate for all the tissues in which cancer can occur.

A major need is the development of models that predict the potential to induce cancer via nongenotoxic mechanisms. As specific genes involved in the carcinogenic process continue to be elucidated, models for evaluating the factors that modify progression of initiated cells into metastatic tumors will be developed. One example of such a model that has already proved to be useful is the TG.AC mouse model discussed above.

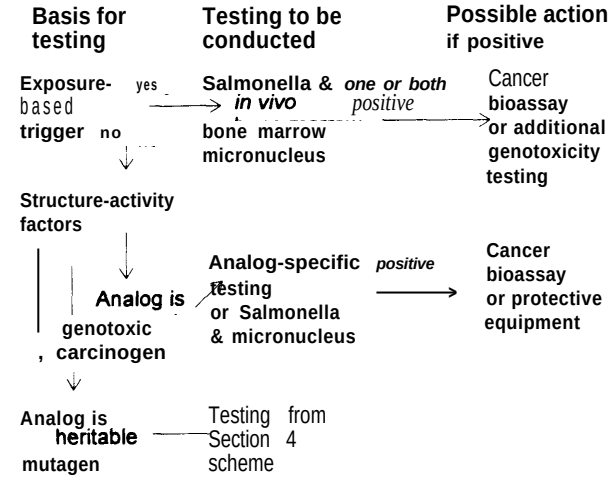
CURRENT TSCA GENOTOXICITY AND CARCINOGENICITY TESTING REQUIREMENTS

The current genotoxicity and carcinogenicity testing schemes for chemicals subject to regulation under TSCA are summarized in figure 3-1 (for new chemicals) and figure 3-2 (for existing chemicals). In each case, the weight of available evidence is considered by the EPA; the arrows in figures 3-1 and 3-2 therefore indicate the steps that are considered under the circumstances indicated rather than a mandate to perform the indicated assays. These testing schemes have been discussed by Auletta et al. (5).

New chemicals that meet specified volume and exposure criteria require testing in two short-term genotoxicity assays (in addition to short-term toxicity and ecological effects testing). The two genotoxicity tests are the Ames *Salmonella* mutagenicity assay and an *in vivo* bone marrow micronucleus test (see figure 3-1). If positive response(s) are obtained, additional genotoxicity testing and/or a cancer bioassay may be required. If both are negative, then a cancer bioassay is unlikely to be required by EPA unless strong evidence (such as chemical structural alerts) suggest carcinogenic potential.

New chemicals may require testing under two other conditions, both depending on structure-activity considerations. If there is a weight-of-evidence argument that the chemical may be a

Figure 3-1: EPA's Mutagenicity Test Scheme for New Chemicals under TSCA

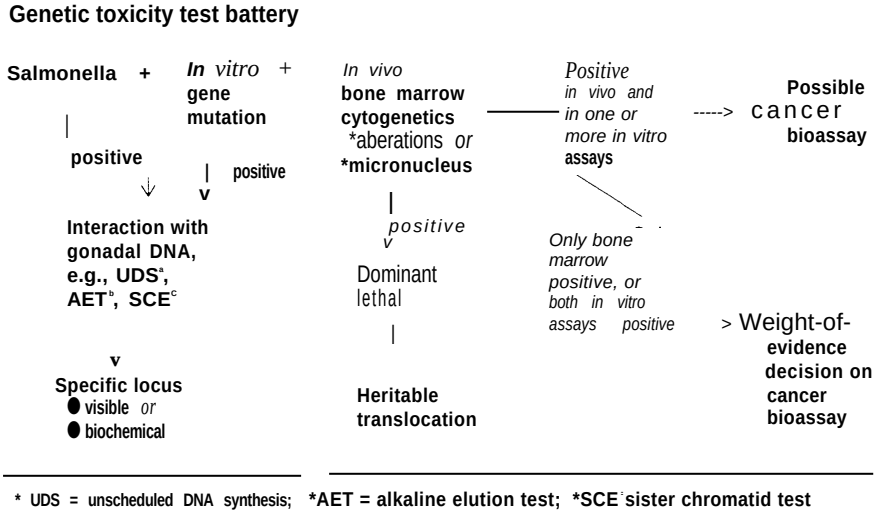


potential carcinogen, genotoxicity testing may be required in assay systems for which positive genotoxicity data on analogs exist. Positive responses for the new chemical generally require a cancer bioassay and/or use of protective equipment to limit exposure.

The third situation for testing of new chemicals involves agents for which data on analogs indicate the potential for heritable mutagenicity, but there is minimal concern for potential carcinogenicity. Such chemicals may require testing similar to that shown in the three-tier scheme for existing chemicals (see discussion below and figure 3-2).

For existing chemicals, the priorities for testing are set by the Interagency Testing Committee. If genotoxicity testing is deemed appropriate, a battery of three short-term genotoxicity tests are required: the Ames *Salmonella* mutagenicity assay, an assay for mutagenicity in mammalian cells *in vitro,* and an *in vivo* bone marrow cytogenetics assay (either a micronucleus or a chromosomal aberration test; see the first row of figure 3-2). If the *in vivo* assay and a minimum of one of the *in vitro* assays are positive, then a rodent cancer bioassay may be required. If both *in vitro* assays or any one assay is positive, then all available data, including test results from other toxicity endpoints, structure-activity relation

Figure 3-2: EPA's Mutagenicity Test Scheme for Existing Chemicals under TSCA



ships, production volume and/or exposure, are reviewed to determine if a cancer bioassay is warranted. If all assays are negative, then a cancer bioassay is unlikely to be required by EPA unless other available data suggest a cause for concern. Positive results in the first tier of mutagenicity testing may trigger additional genotoxicity testing in the second and third tiers for potential heritable mutagenicity (see second and third rows of figure *3-2).*

In view of current knowledge, these testing schemes are considered by the Working Group to be reasonable requirements. They allow reasonable protection against exposure to carcinogens, given the cost constraints of evaluating the many thousands of chemicals subject to TSCA regulations. However, as the test methodologies described above become better validated as predictors of carcinogenic potential during the next few years, strong consideration should be given to including them in the decision tree for carcinogenicity evaluation.

CONCLUSIONS

The principal conclusions of the Working Group are as follows:

- The current OPPT testing scheme is reasonable in view of currently available technology.
- Human epidemiological studies provide the only available direct measure of human carcinogenicity. They provide information that is extremely valuable, but are relatively insensitive, expensive, lengthy, and usually retrospective.
- The chronic rodent carcinogenesis bioassay appears to be the best available assay for predicting human carcinogenicity, but it is expensive and lengthy and therefore practical only for agents with widespread high exposure potential.
- SAR methods provide a highly cost-effective approach to identifying agents with carcinogenic potential, especially if the agents are related to structural analogs with known carcinogenic activity.
- In vitro assays are useful for identifying DNA-reactive carcinogens, especially those that are direct-acting, at relatively low cost. The overall concordance between carcinogenicity assays and in vitro geno-

toxicity is relatively poor and not quantitative.

- In vivo mutagenesis assays applicable to multiple tissues are relatively new and data are still limited. They are potentially very useful but require "validation" as predictors of carcinogenesis. Existing data suggest that predictivity for DNA-reactive carcinogens may be very good.
- Rapid tumor development models are potentially useful. Current data suggest the utility of this type of assay for specific types of carcinogen (e.g., skin cancer in TG.AC mouse). Such assays require further development and "validation" with a variety of classes of carcinogens and noncarcinogens.

The following technologies have a high potential for more cost-effective prediction of carcinogenic potential of chemicals in the near future:

- Reporter genes for mutagenicity in vivo (e.g., Big Blue, MutaMouse, MutaMetrix Mouse).
- Models with rapid tumor development (such as defense knockout [p53, DNA repair], activated oncogenes [TG.AC, pim mouse], etc.).
- Animals with specific oncogene targets.
- Reporters of cell-system- or damage-specific response.
- Markers of cell proliferation and/or tumor growth (CDKs, tumor markers, PCNA).
- Probes that facilitate chromosomal aberration or aneuploidy screening in vivo.
- Transgenic animals with human-like metabolic capacity.
- Improved structure-activity predictions based on improved modeling and more reliable databases.

REFERENCES

1. Ames, B.N. and Gold, L. S., "Too Many Carcinogens: Mitogenesis Increases Mutagenesis," *Science* 249:970-971, *1990.*
2. Ashby, J., "Two Million Rodent Carcinogens? The Role of SAR and QSAR in their Detection," *Mutat. Res.* 305:3-12, *1994.*
3. Ashby, J. and Tennant, R. W., "Definitive Relationships Among Chemical Structure, Carcinogenicity and Mutagenicity for 301 Chemicals Tested by the US NTP," *Mut. Res.,* 257:220-306, 1991.
4. Ashby, J. and Tennant, R. W., "Prediction of Rodent Carcinogenicity for 44 Chemicals: Results," *Mutagenesis* 9:7-15, *1994.*
5. Auletta, A. E., Dearfield, K.L., and Cimino, M. C., "Mutagenicity Test Schemes and Guidelines: U.S. EPA, Office of Pollution Prevention and Toxics and Office of Pesticide Programs," *Environ. Mol. Mutagen.* 21:38-45, *1993.*
6. Bahler, D., and Bristol, D. W., The Induction of Rules for Predicting Chemical Carcinogenesis in Rodents," *Intelligent Systems for Molecular Biology,* L Hunter, J Shavlik, and D Searls (eds.), (Menlo Park, CA: AAAI/MIT Press, 1993).
7. Bakale, G. and McCreary, R. D., "A Physico-Chemical Screening Test for Chemical Carcinogens: the k_e Test," *Carcinogenesis* 8:253-264, *1987.*
8. Ballesta, A. M., et al., "Carcinoembryonic Antigen in Staging and Follow-up of Patients with Solid Tumors," *Tumor Biol.* 16:32-41, *1995.*
9. Barrett, J. C., "Relationship Between Mutagenesis and Carcinogenesis," *Mechanisms of Environmental Carcinogenesis: Role of Genetic and Epigenetic Changes, 1:129-142,* J. Barrett (cd.), (Boco Raton, FL: CRC Press, 1987).
10. Barrett, J. C., Tsutsui, T., and Ts'o P., "Neoplastic Transformation Induced by a Direct Perturbation of DNA," *Nature 274: 229-232, 1987.*
11. Barrett, J.C. and Wiseman, R. W., "Molecular Carcinogenesis in Humans and Rodents," *Prog. Clin. Biol. Res.* 376:1-30, *1992.*
12. Benigni, R., et al., "Electrophilicity as Measured by Ke: Molecular Determinants, Relationship with Other Physical-Chemical and Quantum Mechanical Parameters, and Ability to Predict Rodent Carcinogenicity," *Carcinogenesis* 13:547-553, 1992.

13. Brugge, J., et al. *Origins of Human Cancer: A Comprehensive Review,* (Plainview, NY: Cold Spring Harbor Laboratory Press, 1991).
14. Cohen, S.M., and Ellwein, L.B. "Cell Proliferation in Carcinogenesis," *Science 249: 1007-1011, 1990.*
15. Donehower, L.A., *FASEB J., 5: 225-229, 1993.*
16. Donehower, L. A., et al., "Mice Deficient for P53 are Developmentally Normal but Susceptible To Spontaneous Tumors," *Nature* 356:215-221, *1992.*
17. Dycaico, M.J., Rogers, B.J., and Provost, G. S., "The Species Specific Difference of Mutation Sensitivity of Transgenic lambda/laci Rats," *Environ. Molec. Mutagen. 25,* Suppl. 25:13, 1995.
18. Fan, A. and Howd, R., "Risk Assessment of Environmental Chemicals," *Ann. Rev. Pharmacol. Toxicol.* 35:341-368, *1995.*
19. Ferrigno, D., Buccheri, G. and Biggi, A., "Serum Tumor Markers in Lung Cancer: History, Biology and Clinical Applications," *Eur. Respir. J.* 7:186-197, *1994.*
20. Fornace, A. J., et al., "Genotoxic-Stress-Response Genes and Growth-Arrest Genes," *Ann. NY Acad. Sci.,* 663:139-53, *1992.*
21. French, J. E., et al., "Short Term Carcinogenesis and Mutagenesis Studies with p53 Deficient (+/-) and/or F1 lambda Liz a1pha:p53 Deficient (+/–) Mice," *Environ. Molec. Mutagen. 25,* Suppl. 25:16, 1995.
22. Frierson, M. R., Klopman, G., and Rosenkranz, H. S., "Structure-Activity Relationships SARs among Mutagens and Carcinogens: A Review," *Environ. Mutagenesis* 8:283-327, *1986.*
23. Gibson, D. P., et al., "Detection of Aneuploidy-Inducing Carcinogens in the Syrian Hamster Embryo (SHE) Cell Transformation Assay," *Mutat. Res.* 343:7-24, 1995.
24. Gledhill, B. L., and Mauro, F. New *Horizons in Biological Dosimetry, (New* York, NY: Wiley-Liss, 1991).
25. Gorelick, N.J., "Overview of Mutation Assays in Transgenic Mice for Routine Testing," *Environ. Molec. Mutagen.* 25:218-230, 1995.
26. Gunz, D., Shephard, S. E., and Lutz, W. K., "Can Nongenotoxic Carcinogens be Detected with the lacI Transgenic Mouse Mutation Assay?," *Environ. Molec. Mutagen.* 21:209-211, *1993.*
27. Hagiwara A., et al., "Correlation Between Medium-term Multi-organ Carcinogenesis Bioassay Data and Long-term Observation Results in Rats," *Japan J. Cancer Res.* 84(3):237-245, *1993.*
28. Harris, C., "Chemical and Physical Carcinogenesis: Advances and Perspectives," *Cancer Res.* 51:5023s-5044s, *1991.*
29. Harris, C.C. and Hollstein, M., "Clinical Implications of the p53 Tumor-suppressor Gene," N. *Engl. J. Med.* 329:1318-1327, *1993.*
30. Hayward, J.J., et al., "Differential in vivo Mutagenicity of the Carcinogen-Noncarcinogen Pair 2,4-and 2,6-Diaminotoluene," *Carcinogenesis* (in press, 1995).
31. Herrlich, P., Angel, P., and Rahmsdorf, J. H., "The Mammalian Genetic Stress Response," *Adv. Enzyme Regul.* 25:485-504, *1986.*
32. Hileman, B. "'Expert Intuition' Tops in Test of Carcinogenicity Prediction," *Chem. & Engin. News 71(25):35-38, 1993.*
33. Hollstein, M., et al., "p53 Mutations in Human Cancers," *Science* 253:49-53, 1991.
34. Huff, J. E., "Chemical Toxicity & Chemical Carcinogenesis. Is There a Causal Connection? A Comparative Morphological Evaluation of 1500 Experiments," *Mechanisms of Carcinogenesis in Risk Identification,* 116:437-475, H. Vainio, et al., (eds) (Lyon, France: IARC Sci. Pub., 1992).
35. Huff, J. E., "Chemicals and Cancer in Humans: First Evidence in Experimental Animals," *Environ. Health Perspect.* 100:201-210, 1993a.
36. Huff, J.E. "Issues and Controversies Surrounding Qualitative Strategies for Identifying and Forecasting Cancer Causing Agents in the Human Environment," *Pharmacol. Toxicol.* 72(1):12-27, 1993b.
37. Huff, J. E., Haseman, J.K. and Rail, D.P. "Scientific Concepts, Value, and Significance

of Chemical Carcinogenesis Studies," *Ann. Rev. Pharmacol. Toxicol.* 31:621-652, *1991.*

38. Hulka, B. S., Wilcosky, T. C., and Griffith, J.D., *Biological Markers in Epidemiology,* (Oxford, U.K.: University Press, 1990).

39. International Agency for Research on Cancer, *IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans. Volume 7, Some Anti-thyroid and Related Substances, Nitrofurans, and Industrial Chemicals,* (Lyon, France: IARC Sci. Pub., 1974).

40. International Agency for Research on Cancer, *IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans. Volume 17, Some N-nitroso Compounds,* (Lyon, France: IARC Sci. Pub., 1978).

41. International Agency for Research on Cancer, *IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans. Supplement 7. Overall Evaluation of Carcinogenicity: An Updating of IARC Monographs. Volumes 1-42,* (Lyon, France: IARC Sci. Pub., 1987).

42. Jones, C. A., et al., "An Interlaboratory Evaluation of the Syrian Hamster Embryo Transformation Assay Using Eighteen Coded Chemicals," *Toxicology In Vitro* 2:103-116, *1988.*

43. Leder, A., et al., "v-Ha-ras Transgene Abrogates the Initiation Step in Mouse Skin Tumorigenesis: Effects of Phorbol Esters and Retinoic Acid," *Proct. Natl. Acad. Sci. USA* 87:9178-9182, *1990.*

44. Lee, W. R., Giantonio, B., and Hanks, G. E., "Prostate Cancer," *Curr. Probl. Cancer* 18:295-357, *1994.*

45. Lewis, D. F. V., "Computer-assisted Methods in the Evaluation of Chemical Toxicity," *Reviews in Computational Chemistry,* K.B. Lipkowitz and D.B. Boyd (eds.), (New York, NY: VCH Publishers, Inc., 1992).

46. Lewis, D. F. V., Ioannides, C., and Parke, D. "A Retrospective Evaluation of COMPACT Predictions of the Outcome of NTP Rodent Carcinogenicity Testing," *Environ. Health Perspect.* 103:178-84, *1995.*

47. Lijinsky, W., "Species Differences in Carcinogenesis, " *in vivo* 7:65-72, *1993.*

48. Lipponen, P, et al., "p53 Protein Expression in Breast Cancer as Related to Histopathological Characteristics and Prognosis," *Int. J. Cancer 55:51-56, 1993.*

49. Littlefield, N.A., et al., "Effects of Dose and Time in a Long-term Low-dose Carcinogenicity Study," *J. Environ. Pathol. Toxicol.* 3:17-34, *1979.*

50. Lehman, P.H.M., Laauwerys, R., and Sorsa, M. "Methods of Monitoring Human Exposure to Carcinogenic and Mutagenic Agents," *Monitoring Human Exposure to Carcinogenic and Mutagenic Agents, 59: 423-427,* A. Berlin, et al. (eds.) (Lyon, France: IARC Sci. Pub., 1984).

51. Lehman, P. H. M., et al., *Molecular Dosimetry of Genotoxic Damage: Biochemical and Immunochemical Methods to Detect DNA-damage in vitro and in vivo. TIPS-FEST Supplement, (New* York, NY: Elsevier, 1985).

52. MacGregor, J. T., "Environmental Mutagenesis: Past and Future Directions," *Environ Molec Mutagen,* 23(suppl.24):73-77, *1994.*

53. MacGregor, J. T., et al., "The *in vivo* Erythrocyte Micronucleus Test: Measurement at Steady State Increases Assay Efficiency and Permits Integration with Toxicity Studies," *Fund. Appl. Toxicol.* 14:513-522, *1990.*

54. MacGregor, J. T., et al., "Monitoring Environmental Genotoxicants," *Methods for Genetic Risk Assessment,* D. Brusick, (cd.) (Boca Raton, FL: Lewis Publishers, 1994).

55. MacGregor, J. T., et al., "Integration of Cytogenetic Assays with Toxicology Studies," *Environ Molec Mutagen.* 25:328-337, 1995a.

56. MacGregor, J. T., et al., "New Molecular Endpoints and Methods for Routine Toxicity Testing," *Fund. Appl. Toxicol.* 26:156-173, 1995b.

57. Maronpot, R.R, et al., "Liver Lesions in B6C3F1 Mice: The National Toxicology Program Experience and Position," *Arch. Toxicol. Supl.* 10:10-26, *1987.*

58. Matthews, E.J., Spalding, J. W., and Tennant, R. W., "Transformation of Balb/c-3T3 cells:

V. Transformation Responses of 168 Chemicals Compared with Mutagenicity in *Salmonella* and Carcinogenicity in Rodent Bioassay," *Environ Health Perspect Suppl.* 101(2):347-482, *1993.*

59. McCann, J., et al., "Detection of Carcinogens as Mutagens in the *Salmonella/Microsome* Test: Assay of 300 Chemicals," *Proc. Natl. Acad. Sci. USA* 72:5135-5139, *1975.*

60. Melnick, R.L., and Huff, J.E., "Liver Carcinogenesis is not a Predicted Outcome of Chemically Induced Hepatocyte Proliferation," *Toxicol. Indust. Health* 9:415-438, *1993.*

61. Melnick, R. L., et al., "Cell Proliferation and Chemical Carcinogenesis: a Symposium Overview," *Mol. Carcinog.* 7:135-138, *1993.*

62. Mendelssohn, M. L., Peeters, J. P., and Normandy, M.J., *Biomarkers and Occupational Health,* (Washington, DC: Joseph Henry Press, 1995).

63. Miller, J.A. and Miller, E. C., "Ultimate Chemical Carcinogens as Reactive Mutagenic Electrophiles," *Origins of Human Cancer,* H.H. Hiatt, H.D. Watson and J.A. Winsten (eds), (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1977).

64. Mirsalis, J.C., Monforte, J.A., and Winegar, R. A., "Transgenic Animal Models for Measuring Mutations *in vivo," Crit Rev Toxicol 24: 255-280, 1994.*

65. Mirsalis, J. C., et al., "Induction of Hepatic Mutations in lacI Transgenic Mice," *Mutagenesis* 8:265-271, 1993.

66. National Cancer Institute (NCI), *Bioassay of O-Anisidine Hydrochloride Dihydrochloride for Possible Carcinogenicity (CAS No. 134-29-0), NCI* Tr. No. 89, (Bethesda, MD: National Cancer Institute, 1978).

67. National Cancer Institute (NCI), *Bioassay of 2,4-Diaminotoluene Dihydrochloride for Possible Carcinogenicity (CAS No. 95-80-7), NCI* Tr. No. 162. (Bethesda, MD: National Cancer Institute, 1980a).

68. National Cancer Institute (NCI), *Bioassay of 2, 6- Toluenediamine Dihydrochloride for Possible Carcinogenicity (CAS No. 15481-70-6).* NCI Tr. No. 200, (Bethesda, MD: National Cancer Institute, 1980b).

69. National Toxicology Program (NTP), *Toxicology and Carcinogenesis Studies of 1,3-Butadiene (CAS No. 106-99-0) in B6C3F1 Mice (Inhalation Studies),* NTP Tr. No. 288, (Research Triangle Park, NC: National Toxicology Program, 1984).

70. National Toxicology Program (NTP), *Toxicology and Carcinogenesis Studies of Benzene (CAS No. 71-43-2) in F344/n Rats and B6C3F1 Mice (Gavage Studies),* NTP Tr. No. 289, (Research Triangle Park, NC: National Toxicology Program, 1986).

71. National Toxicology Program (NTP), *Toxicology and Carcinogenesis Studies of 1,3-Butadiene (CAS No. 106-99-0) in B6C3F1 Mice (Inhalation Studies),* NTP Tr. No. 434, (Research Triangle Park, NC: National Toxicology Program, 1993).

72. Provost, G. S., et al., "Evaluation of Mutagenic and Non-mutagenic Compounds Using lacI Transgenic Rodents," *Environ Molec. Mutagen. 23(* Suppl. *23): 55, 1994.*

73. Provost, G. S., et al., "Validation Studies of the lambda/lacI Transgenic Mouse Assay," *The Toxicologist* 15:174, 1995.

74. Przygoda, R.T., McKee, R.H. and Traul, K.A., "The Use of Short Term Assays in the Evaluation of the Dermal Carcinogenic Potential of Petroleum-Derived Materials," *Environ. Molec. Mutagen.* 19:51, 1992.

75. Randerath, K, et al., "Postlabeling Methods for Carcinogen-DNA Adduct Analysis," *Environ. Health Perspect.* 62:57-65, 1985.

76. Richard, A. M., "Application of SAR Methods to Non-congeneric Data Bases Associated with Carcinogenicity and Mutagenicity: Issues and Approaches," *Mutation Research* 305:73-97, *1994.*

77. Riou, G., et al., "The p53 and mdm Genes in Human Testicular Germ-cell Tumors," *Molec. Carcinog.* 12:124-131, *1995.*

78. Robbins, W. A., et al., "Detection of Aneuploid Sperm by Fluorescence *in situ* Hybridization: Evidence for a Donor Difference in Frequency of Sperm Disomic

for Chromosomes 1 and Y," *Am J Human Genet* 52:799-807, *1993.*

79. Ruiball-Morell, A., "CEA Levels in Non-neoplastic Disease," *Int J Biol Markers* 7: *160-166, 1992.*

80. Sankaranarayanan, R., Wahrendorf, J., and Demaret, E., *Directory of On-Going Research in Epidemiology 1994,* IARC Sci. Pub. No. 130 (Lyon, France: IARC Sci. Pub., 1994).

81. Sarasin, A., "SOS Response in Mammalian Cells," *Cancer Invest.* 3:163-174, *1985.*

82. Schmitt, F. C., and Andrade, L., "Spectrum of Carcinoembryonic Antigen Immunoreactivity from Isolated Ductal Hyperplasias to Atypical Hyperplasia Associated with Infiltrating Ductal Breast Cancer," *J Clin Pathol 48:53-56, 1995.*

83. Shabana, A., and Onsrud, M., "Tissue Polypepetide-specific Antigen and CA 125 as Serum Tumor Markers in Ovarian Carcinoma," *Tumor Biol* 15:361-367, *1994.*

84. Shelby, M.D., "The Genetic Toxicity of Human Carcinogens and its Implications," *Mutat. Res.* 204:3-15, *1988.*

85. Short, J. M., et al., "The Use of lambda Phase Shuttle Vectors in Transgenic Mice for Development of a Short Term Mutagenicity," *Mutation and the Environment, Part A, New* York, NY: Wiley-Liss, 1990).

86. Sisk, S. C., et al., "Molecular Analysis of lacI Mutants from Bone Marrow of B6C3F1 Transgenic Mice Following Inhalation Exposure to l,3-Butadiene," *Carcinogenesis* 15:471-477, *1994.*

87. Smith, M. L., et al., "Involvement of the p53 Tumor Suppresor in Repair of U.V. Type DNA Damage," *Oncogene* 10:1053-1059, 1995.

88. Spalding, J. W., et al., "Chemically-induced Skin Carcinogenesis in a Transgenic Mouse Line (TG.AC) Carrying a v-Ha-ras Gene," *Carcinogenesis* 14:1335-1341, *1993.*

89. Spalding, J. W., et al., "The TG.AC Transgenic Mouse Line: An Important *in vivo* Short Term Test Model for Identifying Nongenotoxic Carcinogens," *Environ. Molec. Mutagen. 25,* Suppl. 25:50, 1995.

90. Swenberg, J.A. "Alpha 2 u-globulin Nephropathy: Review of the Cellular and Molecular Mechanisms Involved and their Implications for Human Risk Assessment," *Environ. Health Perspect.* 101, Suppl. 6:39-44, 1993.

91. Swierenga, S. H. H., and Yamasaki, H., "Performance of Tests for Cell Transformation and Gap-junction Intercellular Communication for Detecting Nongenotoxic Carcinogenic Activity," *Mechanisms of Carcinogenesis in Risk Identification,* H. Vaino, et al. (eds.) (Lyon, France: IARC Sci. Pub., 1992).

92. Tennant, R. W., "Stratification of Rodent Carcinogenicity Bioassay Results to Reflect Relative Human Hazard," *Mutat. Res.* 286:111-118, 1993.

93. Tennant, R. W., and Zeiger, E., "Genetic Toxicology: The Current Status of Methods of Carcinogen Identification," *Environ. Health Perspect.* 100:307-315, *1993.*

94. Tennant, R. W., et al., "Prediction of Chemical Carcinogenicity in Rodents from *in vitro* Genetic Toxicity Assays," *Science 236:933-941,* 1987.

95. Tennant, R. W., et al., "Prediction of the Outcome of Rodent Carcinogenicity Bioassays Currently Being Conducted on 44 Chemicals by the National Toxicology Program," *Mutagenesis* 5:3-14, *1990.*

96. Tennant, R. W., French, J. E., and Spalding, J. W., "Identification of Chemical Carcinogens and Assessing Potential Risk in Short Term Bioassays using Transgenic Mouse Models," *Environ. Health Perspect. 103(10)* (in press), 1995.

97. Tice, R.R., et al., "The Single Cell Gel (SCG) Assay: an Electrophoretic Technique for the Detection of DNA Damage in Individual Cells," *Biological Reactive Intermediates. IV. Molecular and Cellular Effects and Their Impact on Human Health,* Witmer, C. R., et al., (eds.) (New York, NY: Plenum, 1991).

98. Tomatis, L., "The Predictive Value of Rodent Carcinogenicity Tests in the Evaluation of Human Risks," *Ann. Rev. Pharmacol. Toxicol. 19:51 1-530, 1979.*

99. Tomatis, L., *Cancer: Causes, Occurrence and Control.* IARC Sci. Pub. No. 100, (Lyon, France: IARC Sci. Pub., 1990).

100. Tucker, J., et al., "Validation of Chromosome Painting as a Biodosimeter in Human Peripheral Blood Lymphocytes Following Acute Exposure to Ionizing Radiation *in vitro," Int. J. Radiat. Biol.* 64:27-37, *1993.*

101. Wachsman, J.T., et al., "Predicting Chemical Carcinogenesis in Rodents," *Environ. Health Perspec.* 101:444-445, *1993.*

102. Weinstein, I.B. "Mitogenesis is Only One Factor in Carcinogenesis," *Science 251:387-388, 1991.*

103. Weinstein, I.B., "Toxicity, Cell Proliferation, and Carcinogenesis. *Mol. Carcinog.* 5:2-3, *1992.*

104. Wogan, G.N., and Newberne, P.M. "Dose Response Characteristics of Aflatoxin B, Carcinogenesis in the Rat," *Cancer Res.* 27:2370-2376, *1967.*

105. Wolmark, N., et al., "The Prognostic Significance of Preoperative Carcinoembryonic Antigen Levels in Colorectal Cells. Results from NSABP (National Surgery Adjuvant Breast and Bowel Project) Clinical Trials," *Annals Surgery* 199:375-382, 1984.

106. Woo, Y-T., et al., *Chemical Induction of Cancer: Structural Bases and Biological Mechanisms,* Vols. 11, III, (New York, NY: Academic Press, Inc., 1985).

107. Woo, Y-T., et al. "Development of Structure Activity Relationship Rules for Predicting Carcinogenic Potential of Chemicals. Proceedings of ATSDR Sponsored Workshop on Screening Technologies," *Toxicology Letters 78:* (in press), 1995.

108. Zagars, G. K., et al., "The Source of Pretreatment Serum Prostate Specific Antigen in Clinically Localized Prostate Cancer," *Int. J. Radiat. Oncol. Biol. Phys.* 32:21-32, 1995.

109. Zeiger, E., et al., "Evaluation of Four *in vitro* Genetic Toxicology Tests for Predicting Rodent Carcinogenicity: Confirmation of Earlier Results with 41 Additional Chemicals," *Environ. Molec. Mutagen. 16,* Suppl. 18:1-14, 1990.

4

Neurotoxicity

Jean Harry, John L. O'Donoghue, and Alan M. Goldberg

ABSTRACT: *Guidelines for testing existing chemicals for neurotoxicity under the Toxic Substances Control Act (TSCA) have been published by the U.S. Environmental Protection Agency (EPA). While the current regulatory process for assessing existing chemicals under TSCA does not use a tier-testing approach, there is widespread support for tier-testing processes. However, there is general disagreement between the Agency and the regulated community over what tests should be used for hazard identification (i. e., first-tier screening test). The regulated community sees the standard toxicity tests which are commonly conducted for systemic toxicity as sufficient for neurotoxicity screening, while the Agency regards its guideline tests as necessary to screen chemicals specifically for neurotoxicity. The guideline tests while frequently referred to as screening tests are expensive and time-consuming and therefore not widely used outside of formal existing chemicals testing programs. Development of true screening tests should be based on a mechanistic understanding of the neurobiological processes which result in neurotoxicity. The most commonly used alternative screening techniques include structure-activity analysis and in vitro methods. In vitro techniques (e.g., primary neuronal cultures, glial cell cultures, organotypic explants) are commonly used today to study mechanisms of neurotoxicity and have the potential for being used for hazard identification. Rapid, inexpensive screening tests would be expected to be useful during the early phases of new product development cycles and thus may have much more pollution prevention potential than existing methods. Such tests may eventually offer methodologies to either replace or complement tests currently used. The complex nature of the nervous system suggests that if in vitro methods gain acceptance as screening tests for neurotoxicity, they will have to be used in batteries of several assays to study multiple endpoints.*

DEFINING NEUROTOXICITY: CONTROVERSIAL BUT CRITICAL

Neurotoxicity is one of several organ-specific endpoints used by regulatory agencies to determine hazards of chemical exposure. Definitions of neurotoxicity have been established by various organizations as the capacity of chemical, biological, or physical agents to cause adverse functional or structural changes in the central or peripheral nervous system **(3, 5, 9,** 10, 15, 16, 18).

In each of these cases, the definition of neurotoxicity is dependent on the controversial interpretation of the word "adverse". Tilson (12) has proposed that the definition of adverse includes alterations from a baseline state that diminishes the ability of an organism to survive, reproduce, or adapt to its environment. It has been suggested that unintended or unwanted effects should also be included under this definition (12). However, such a definition must take into account the possibility that neurobehavioral effects might be produced nonspecifically at high dose levels. Some argue that the definition of neurotoxicity should be defined more in terms of direct nervous system toxicity **(5).**

Clarification of the definition of neurotoxicity is critical to the design of neurotoxicity screening tests, since the designer of screening tests must have a clear understanding of what the testing paradigm is expected to accomplish. For example, tests to detect blurring of vision caused by eye irritation must be designed very differently from those expected to detect vision loss due to methanol intoxication. Interpretation of the results of currently used tests for neurotoxicity can be difficult because the currently used screening tests do not necessarily distinguish between effects which are direct vs. those which are indirect. Direct effects are produced by agents or their metabolizes that produce toxicity

by primarily interacting with target sites in the nervous system. Indirect effects are produced by agents or their metabolizes that produce toxicity by interacting with target sites outside the nervous system. The occurrence of systemic toxicity could complicate the interpretation of functional changes; however, systemic toxicity does not necessarily preclude the use of functional changes in defining neurotoxicity.

A major concern is that nonspecific behavioral changes produced by high-dose level systemic toxicity may be interpreted as providing evidence of neurotoxicity. A well designed neurotoxicity study needs to control for nonspecific toxicity that could impair the assessment of chemically-induced changes in nervous system function. Concerns about indirect effects may be addressed by selecting appropriate dose levels which do not induce a significant degree of general systemic toxicity. In some cases the differentiation between direct and indirect effects may require additional second-tier testing to resolve.

Also of concern is the distinction between effects which are transient vs those which are persistent. Transient effects are those which are considered to be fleeting in time and typically are related to pharmacological processes and the presence of a chemical in the body, while persistent effects have a lifespan which exceeds the lifespan of the chemical in the body. Some transient effects (e.g. seizure activity) are obviously serious, but many others (e.g. changes in enzyme levels or increased rates of whisker twitching) may not have any recognizable consequence, yet in the current risk assessment process, each of these changes could be evaluated as critical endpoints requiring equivalent safety factors.

CURRENT METHODS FOR NEUROTOXICITY TESTING UNDER TSCA

The complexity and integrative nature of the nervous system makes the identification of a single endpoint problematic. As a result, neurotoxic effects are usually measured at multiple levels of nervous system organization, including behavioral, neurophysiological, neurochemical, and neuroanatomical levels. There is general agreement that an assessment of potential neurotoxicity should be based on a number of parameters generated from a variety of tests at relevant dose levels. Historically, morphological methods have been used to detect neurotoxicity; however, assessments of neurotoxic potential can be enhanced by a combination of morphological and functional data. Some neurotoxic agents and pharmacologically active materials (e.g., cholinesterase inhibitors) can cause alterations in the functioning of the nervous system in the absence of morphological changes, thus adding support to an assessment based on different types of endpoints (13).

A number of expert groups has recommended tier-testing strategies for evaluation of chemically-induced neurotoxicity (9, 15). Cage-side observations and the US Environmental Protection Agency (EPA) guideline for a Functional-Observational Battery (FOB) are examples of tests which are considered first-tier tests by the regulated community and the Agency, respectively. The initial phase of a tier-testing strategy is the identification of chemical's capability to produce neurotoxicity at some dose level (i.e., hazard identification). First-tier tests are typified by their capability to assess a large number of animals, usually requiring little or no training of test animals prior to exposure, and generally being relatively simple to perform. The types of observational methods used to detect neurotoxicity (e.g., FOB) have been criticized as labor intensive, subjective, and semi-quantitative. However, the current manner in which clinical signs are collected has also been criticized as being highly variable and poorly documented. Therefore, the development of the FOB has been at least partially driven by efforts to develop methods to place observation of clinical signs under a systematic protocol. Whether first-tier testing is comprised of cageside observations or the FOB, there is widespread agreement that any screening technique should include the following features: 1) the method and endpoints should be

clearly defined, 2) the effects should be quantified using an explicitly stated rating scale, 3) observers should be trained, and 4) a number of endpoints should be assessed to evaluate multiple modalities of nervous system function.

The EPA has considered the inclusion of a quantitative measure of motor activity in the first-tier testing for existing chemicals under Toxic Substances Control Act (TSCA). The Agency's approach on the use of motor activity is based on the large wealth of neurobehavioral pharmacology data using this endpoint. In addition, the fact that motor activity levels can be influenced by the general toxicity of a chemical can be used to aid in the interpretation of observational screening data. However, the use of motor activity as a test for neurotoxicity has been repeatedly rejected by the regulated community which views such tests as having little value for identification of neurotoxicity, prone to interpretation bias, and invalidated as a screening test for neurotoxicity. Other tests which have been included in a first-tier test battery are quantitative measures of limb grip strength and hind limb foot splay. In many situations, functional tests are used in conjunction with other methods including neuropathology.

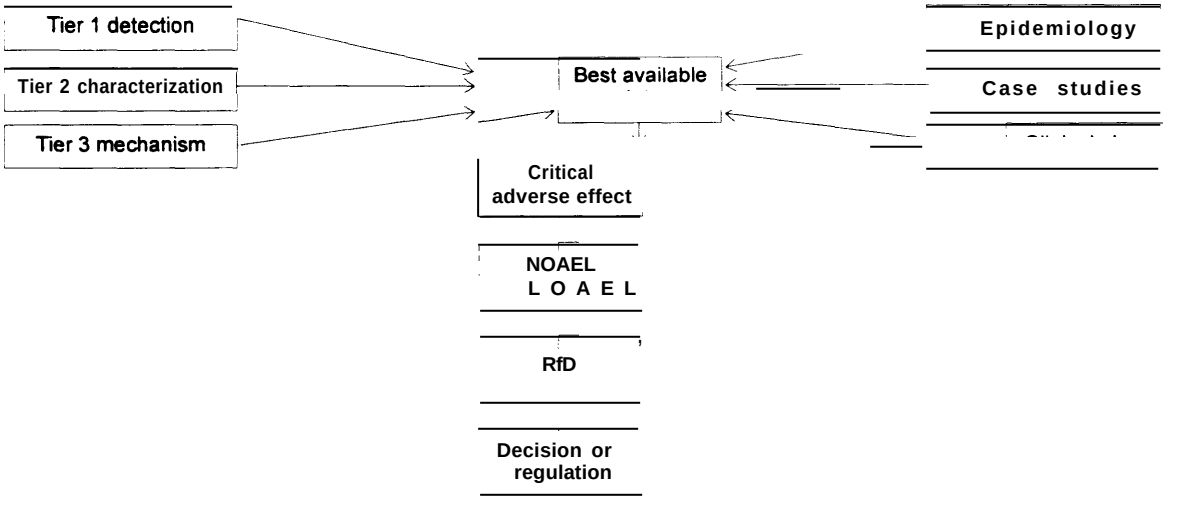
In order to improve identification of agents capable of producing neurotoxic effects, efforts have been made to validate reliable, sensitive measures of neurotoxicity. Increased emphasis on testing for neurotoxicity has been included in the existing chemicals program under TSCA resulting in the development of testing guidelines by the EPA and standardized procedures by the regulated community. Cageside observations for neurological and behavioral changes have been part of toxicological testing practices for many years. The cageside observations and routine pathology studies conducted as part of the data gathering process for systemic toxicity are considered by the regulated community as the first tier for all systemic toxicants, including neurotoxicants. However, given regulatory agency guidelines and the need to provide more quantitative measurements, FOBS have been developed to include more systematic recording of observations. Testing guidelines, such as the FOB, rely on behavioral measures based on the assumption that behavior appears to be the net result of the integrated output of various nervous system processes. A change in such an integrative process could serve as a relatively sensitive indicator of nervous system dysfunction, especially since many of the endpoints assess changes in sensory, motor, and cognitive functions.

Although a number of articles exist in the published literature on the use of observational methods for neurotoxicity testing, not all report equal success in detecting neurotoxic effects, pointing out the need for data on inter- and intralaboratory reliability and interlaboratory sensitivity. The International Programme on Chemical Safety of the World Health Organization is currently sponsoring an international collaborative study on neurobehavioral methods for the FOB, motor activity, and grip strength.

Although observational methods are conceptually the most straightforward, they are also the easiest to confound and can sometimes be difficult to interpret without some internal or external corroboration of results. Given the various biological modalities encompassed in nervous system function and the numerous endpoints used to assess function, questions can arise concerning the significance of a change in a specific endpoint. One of the approaches that has been proposed to deal with such data is to cluster the various observations into functional domains that represent common neurobiological processes (i.e., autonomic function) and generate a composite response score to reflect the functional integrity of a given subset of necrologic processes. This approach would allow data to be evaluated within a small number of neurobiologically meaningful clusters rather than numerous isolated endpoints. While this clustering methodology may be conceptually appealing, widespread acceptance of it will depend on how well the testing community perceives that there is a meaningful biological basis for the clustering.

The second tier of neurotoxicity testing (beyond screening for the potential for neurotoxicity) is generally regarded as providing more

FIGURE 4-1: Use of Data in Regulatory Decision-Making



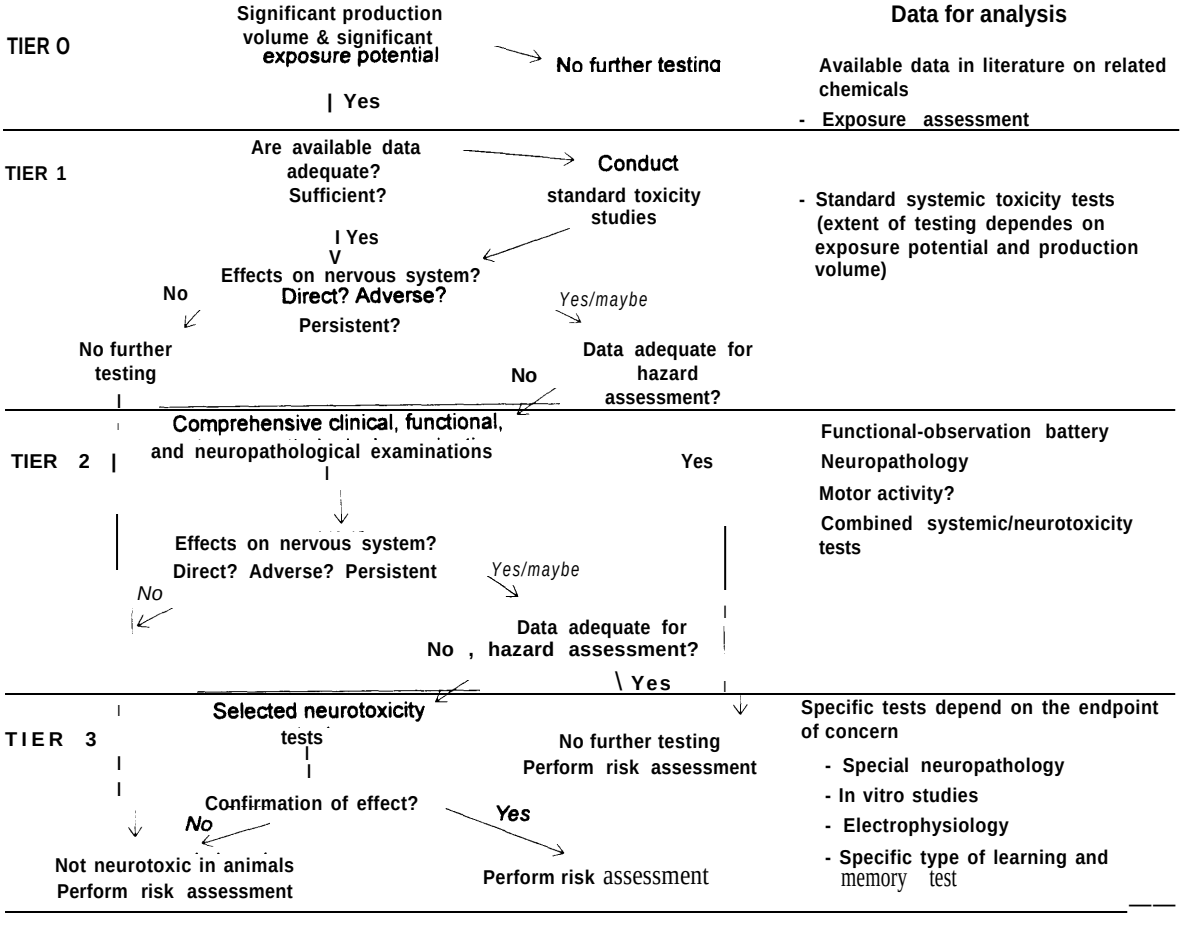
specific test results than those of the first tier and designed to characterize the nature of a chemical's neurotoxicity. The choice of the most appropriate approach and method(s) is dependent on the scientific questions generated by the results of first-tier testing. Such second-tier tests are aimed at objectively quantifying sensorimotor deficits, evaluating cognitive behaviors relating to learning and memory, and assessing performance of complex tasks.

Third-tier testing involves "mechanistic studies" which attempt to establish a detailed profile of a chemical's effect at several levels of nervous system organization (i.e., behavioral, cellular, molecular). Such tests are expected to provide data on enzyme function, ionic balance, transmitter systems, receptor modulation, and the pathogenesis of effects. The value of mechanistic studies cannot be emphasized enough. It is from such studies that understanding of the processes underlying neurotoxicity and specificity of effect is gained. Mechanism of action studies provide the basis for moving beyond empirical structure-activity analysis and being able to rationally prioritize chemicals for testing and most importantly, develop biologically-based models of neurotoxicity.

The EPA has established guidelines to test existing chemicals under TSCA (16). These guidelines include a FOB, motor activity, neuropathology, nerve conduction velocity, and schedule-controlled operant behavior. A neurotoxicity screening battery (17) combining the FOB, motor activity, and neuropathology guidelines into a single screening battery is now required for registration and reregistration of pesticides. However, the EPA does not at present use a tier-testing strategy within the TSCA regulatory context. For example, current test rules are promulgated with a full battery of tests with no guidance on how to use tests in a tiered manner. Likewise when testing is completed for a chemical, all test results from hazard identification, characterization, and mechanism-based studies are considered together (figure 4-1) to determine a critical adverse effect (the most sensitive endpoint). The critical effect could be identified from any of the data available (including the FOB) and the risk assessment process then uses this effect to support regulatory decision making.

Within the regulated and basic science communities, the concept of a tiered approach to testing has received wide support. A scheme for

FIGURE 4-2: Test Strategy for Neurotoxicity



SOURCE: Adapted from Eisenbrandt, D. L., et al., "Evaluation of the Neurotoxic Potential of Chemicals in Animals," *Food and Chemical Toxicology* 32:655-669, 1994.

using data collected by tiered testing (figure 4-2) which begins with the collection and analysis of data from standard toxicology tests has been published (4); Perhaps the strongest disagreements that the regulated community have with the present regulatory approach to neurotoxicity testing are that the data from standard tests are underutilized as a first-tier test for neurotoxicity and that relatively nonspecific behavioral signs from existing data have been used to trigger additional testing (which is often nonspecific as well) and risk assessments. The EPA, on the other hand, is concerned that some first-tier approaches involving cageside observations may be insensitive and therefore, subject to frequent false negative results. An additional concern is that cageside observations collected during standard toxicity tests have not been designed to specifically detect neurotoxicity. Incorporation of more systematic, better defined protocols for cageside observations into standard tests may provide a wealth of first-tier type information.

The significant costs associated with current screening methods recommended for existing chemicals under TSCA (FOB, motor activity, and neuropathology) are an obstacle to widespread

use of the methods. For example, addition of the three screening tests to an acute oral toxicity has been estimated to increase median cost of the base test ($21K) by $50K (15). For subchronic tests, the base test cost ($111K) has been estimated to increase by $79K; the base chronic test cost ($308K) has been estimated to be increased by $113K (15). The addition of schedule-controlled operant behavior to a subchronic oral test has been estimated to increase test cost by $64K (15). When neurotoxicity testing is conducted as a independent test, final costs have to include both those associated with the neurotoxicity test and the incremental original base cost. In the future, the use of *in vitro* methods could significantly reduce the costs of first-tier testing.

DESIGN CRITERIA FOR NEW SCREENING TESTS

A realistic assessment of how many chemicals actually might require testing is also important in designing future tests. If all of the approximately 72,000 chemicals on the TSCA inventory were to be tested by the current screening battery (FOB, motor activity, and neuropathology), only for acute effects, the testing bill would be greater than $4B; not to mention the 6 billion animals that would be needed. This is clearly not a feasible approach. However if the number of chemicals were actually more manageable, more significant testing might be achieved. The calculation of a realistic number of chemicals for testing could be made by subtracting the 25,000 polymers in the inventory and the large number of site-limited intermediate chemicals, low production volume chemicals, those with little or no exposure, and those which cannot be tested because of physical-chemical property limitations. With a realistic evaluation of the number of chemicals requiring testing, the magnitude of the problem may be much more manageable than it currently is perceived to be. Clearly, some strategy for prioritization of chemicals for testing needs to be developed. The prioritization process could begin with the EPA list of chemicals reported under the TSCA 8(b) Inventory Update Rule, which includes 9487 chemicals produced in excess often-thousand kilograms/year (1) and set priorities for screening a smaller set of chemicals based on exposure potential.

Screening tests for neurotoxicity should have several characteristics common to testing paradigms for other forms of systemic toxicity. The methods should have a high degree of sensitivity to insure against excessive numbers of false negative results. The method should be specific and produce results which are predictive of a hazard to the nervous system and thus avoid frequent false positive results. The results should be reproducible within and between laboratories. The screen should be cost effective and produce timely results; assays which cost thousands of dollars and take months or years to yield results are not really screening tests. If the screening method is to be widely used, most toxicology laboratories should be capable of performing the method with trained personnel.

Any attempt to design additional screening methods needs to take into consideration that many of the materials tested will not have neurotoxic potential. It is therefore critical that the methodology used be fairly specific for neurotoxicity to prevent a high number of false positive results. Any misclassification of chemicals for neurotoxicity, whether false positive or false negative, will result in some form of unnecessary future cost and wasted resource. The available estimates of the number of chemicals which might be neurotoxic ranges from 5-28°/0 (15). A significant reason for this range of estimates in the number of neurotoxic substances is in how neurotoxicity, and in particular, an adverse effect on the nervous system is defined. Thus the designer of future screening tests will need to decide whether the test is meant to detect any perturbation in baseline function of the nervous system or generally recognized toxic effects on the nervous system.

If a screening test is expected to be used for new chemicals as well as existing chemicals, the test designer should consider where in the product development cycle the chemical is to be tested. The cost associated with current neurotoxicity test

methods are such that the tests are only used for chemicals which are produced in relatively large volumes. However, there are potential uses for methods to screen for neurotoxicity during the product development cycle. Most product development cycles can be broken down into various phases such as product conception, product and process development, commercialization, and post-commercialization. Ideally, screening tests could be used early in product conception and development phases of the cycle as an aid to choosing candidate chemicals for development. Since *in vitro* assays require less test material, they may be particularly useful in early phases of product development when supplies of new chemicals are typically low. Use of such screening tests in earlier phases of product development could support pollution prevention evaluations along with accompanying product efficacy, cost benefit, process development, and other important considerations.

ALTERNATIVES TO EXISTING TEST METHODS

All presently available neurotoxicity test guidelines for existing chemicals use laboratory animals, primarily rats. In order to accomplish screening of large numbers of chemicals, alternative methods need to be developed to reduce the cost and time to complete screening. Such methods include structure-activity analysis as well as *in vitro* methods.

In general, structure-activity relationships (SAR) in neurotoxicology have received relatively little attention, however, SAR is routinely used in the premanufacture notification process for new chemicals and by pesticide and pharmaceutical research groups for the identification of candidate chemicals with neuroactive properties. There are some examples of SAR being used effectively in neurotoxicology. Many of these are based on available mechanistic data for representative chemicals that allow for an understanding of a specific process underlying one type of neurotoxicity which can be generalized to other similar chemicals. For example, the identification of the importance of gamma-diketones for induction of axonopathy led to the screening of chemicals which were gamma-diketones or could be metabolized to gamma diketones for axonopathy using small scale animal screening tests (8) and *in vitro* techniques (14). SAR techniques are currently used empirically to qualitatively identify materials which might be neurotoxic. As currently used, SAR is not able to identify chemicals which are not neurotoxic. However, there is reason to believe that continued work on SAR could lead to much more informative quantitative techniques. SAR offers the potential for development of inexpensive technology that could be used to evaluate large numbers of chemicals before other screening tests are employed.

For development of improved SAR techniques and *in vitro* methods, there is a need to better understand the chemical-biological interactions (mechanisms) that result in neurotoxicity. If *in vitro* tests are mechanistically-based, they are much more likely to be used earlier in the product development process as they will more likely be accepted as reliable predictors of neurotoxicity. Because of the number of chemicals to be evaluated and the complexity of the nervous system, mechanistic *in vitro* studies can be expected to provide results which can be interpreted and extrapolated. Due to the complexity of the nervous system, batteries of *in vitro* tests will be necessary to characterize toxicity and evaluate potential hazard. Even when batteries of *in vitro* tests are available for hazard identification, whole animal tests will probably still be needed to develop data sets adequate for risk assessment. However, the additional information provided by *in vitro* tests may reduce the number of animal required for first-tier testing.

As more is learned about the mechanism of action of neurotoxic chemicals, initial efforts should be directed at refining existing test methods to reduce the number of animals used to evaluate neurotoxicity. Such information would also offer the ability to develop *in vitro* assays that would address specific mechanistic endpoints

Table 4-1: In Vitro Assays for Neurotoxicity

Test System	Endpoint Parameter(s)	(+) Advantages and (-) Disadvantages	Example
Membrane models (erythrocyte and synaptosome membranes)	Effects on integral cell membrance enzymes (AcHE, ATPase)	(+) Useful for mechanistic studies (-) Limited specifically to compounds which effect cell membranes	Carbon disulphide Toluene
Primary neuronal cultures		(+) Possible to study individual neurons (+) Useful for mechanistic studies (-) Neurons are deprived of their normal afferent and efferent targets (-) Maintenance of the cells is difficult (-) No blood-brain barrier	Excitotoxic amino acids NMDA antagonists
Glial cell cultures		(+) Useful for mechanistic studies (-) No blood-brain barrier	Ethanol Alpha-chlorhydrin
Cell lines	Effects on ion channels and interaction with receptors	(+) Useful for studying cell biology (-) Model system that shares only certain features with real neurons or glia	Methylmercury Pyrethroid insecticides
Organotypic explants	Effects on development of the nerve system, on development of neuro-muscular junctions or other morphological endpoints	(+) Useful for mechanistic studies (-) Preparation and maintenance is difficult (-) Neurons are immature (-) Explant is disconnected from its normal afferents	Tellurium Hexacarbon solvents
Rotation-mediated aggregating cultures	Effects on specific transmitter systems, on cell surface recognition and on enzymes	(+) Ease of preparation, reproducibility and representation (+) Appropriate for interdisciplinary investigation (-) Neurons are immature (-) Large quantities of foetuses are required (-) Electrophysiological examination is not possible	Kainic acid 6-hydroxydopamine

SOURCE: Adapted from European Center for Ecotoxicology and Toxicology of Chemicals, Monograph No. 18, *Evaluation of the Neurotoxic Potential of Chemicals* (Brussels: September 1992)

responsible for neurotoxicity. In cell culture systems, it is possible to examine the effects of growth factors, hormones, and chemicals on growth, differentiation, cell-cell interactions, and metabolic activities. In recent years, the advent of molecular biological methods has allowed for cell lines to be developed to examine specific targets as neurotransmitter receptors or specific genes. Because the nervous system is composed of a highly specialized, heterogeneous, yet integrated population of cells, single *in vitro* test systems are unlikely to be able to mimic the responses of the nervous system to a broad range of chemically-induced toxicities. However batteries of *in vitro* tests offer the possibility of developing first-tier screening methods.

Within the area of neurotoxicology, recent evaluations have focused on correlating *in vivo* and *in vitro* endpoints. Although cell culture models have been proposed as systems for neurotoxicity screening, it is the ability to conduct detailed analysis and experimental manipulations that makes such culture systems attractive for the identification and subsequent evaluation of cellular mechanisms underlying neurotoxicity. The major types of nervous systems cultures (table 4-1) that have been useful in assessing neurotoxicity range from clonal cell lines, primary cells, reaggregate cultures, organotypic explants, organ cultures to whole embryos. Each system offers a unique approach to examining toxicant-induced perturbations, however, each system is not without distinct limitations. The emphasis on the use of *in vitro* techniques within neurotoxicology has resulted in the development of model systems which encompass a wide array of basic approaches both as a screening battery for early detection of potential neurotoxicity and to detect basic underlying mechanisms associated with both neural development and functioning. While *in vitro* systems offer unique opportunities to examine detailed cellular events associated with environmental perturbations to the nervous system, the results from such studies need to be viewed in the isolated nature in which they are generated. If a chemical is found, *in vitro,* to have selective neurotoxic properties as compared to general cytotoxicity one may speculate that the chemical would also be neurotoxic *in vivo.* However, no matter how attractive and useful an *in vitro* system appears to be, it is still an artificial system that is isolated from the various biological processes that greatly modulate *in vivo* neurotoxicity. Results from *in vitro* studies using single cell systems are not easily extrapolated to an integrated nervous system. In addition, the interpretation of *in vitro* data collected in the absence of normal metabolic systems and without appropriate toxicokinetic and toxicodynamic information is highly problematic. Given the complicated nature of the interdependent interactions of the various cell types and network processes in the nervous system, it would be unwise to at this time to conclude that a chemical has or does not have neurotoxic potential based upon data from *in vitro* systems alone.

VALIDATION OF NEW SCREENING METHODS

Numerous test methods exist to evaluate the potential for a chemical to produce neurotoxic effects by alteration of specific organization processes in the nervous system (2, 11). The question of validation of these systems remains a difficult problem. For example, many laboratories have ongoing projects to develop methods for screening chemicals, however, assays that have been found to be useful and predictive in one laboratory for a specific purpose and in an isolated environment may not be considered "validated" for broad screening purposes by other laboratories. Such assays can include both *in vivo* behavioral screening assays and mechanistically-based *in vitro* tests. The success of such tests is critically dependent upon the level of expertise and training that exist within any one laboratory. In order to validate an assay for widespread use, there are a number of steps that are required. Among these are that the assay must have adequate development to be considered robust enough to be used under varying laboratory conditions without failure and the assay must receive acceptance following a peer review which

typically includes a round of interlaboratory comparison testing.

There are many approaches to the assessment (validation) of methods. A modular approach has recently been submitted for publication (6,7). The concept prescribes validating a single *in vitro* assay independent of other *in vitro* assays. This modular concept evaluates the results obtained with a specific group or class of chemicals in an *in vivo* assay (validation standard). The same group of chemicals are evaluated for their response in an *in vitro* assay. The results of the *in vivo* and *in vitro* **assays** are compared to assess whether the *in vitro* assays predict the *in vivo* response. A module consists of the chemical group, the validation standard, and an *in vitro* assay. A validation study may consist of several modules. In this case, one evaluates each module separately and therefore, an *in vitro* assay is not compared to another *in vitro* assay. Validation is an important step in the development of acceptance of alternative methods for testing. Without broad acceptance by the neurotoxicology testing community, new screening methods are unlikely to receive widespread use (6,7).

CHAPTER 2 REFERENCES

1. Auer, C., Chemical use inventory update meeting for stakeholders, Washington, DC, April 13, 1995.
2. Chang, L. W., and Slikker, Jr., W. (eds.), *Neurotoxicology: Approaches and Methods* (Orlando, FL: Academic Press, Inc, 1995).
3. Consumer Product Safety Commission, "Labelling Requirements for Art Materials Presenting Chronic Hazards; Guidelines for Determining Chronic Toxicity of Products Subject to the FHSA; Supplemental Definition of Toxic under the Federal Hazardous Substances Act," final rule, *Federal Register* 57:46626-46674, *1992.*
4. Eisenbrandt, D. L., et al., "Evaluation of the Neurotoxic Potential of Chemicals in Animals," *Food and Chemical Toxicology* 32:655-669, *1994.*
5. European Centre for Ecotoxicology and Toxicology of Chemicals, *Evaluation of the Neurotoxic Potential of Chemicals,* Monograph No. 18 (Brussels, Belgium: September 1992).
6. Goldberg, A. M., et al., "Framework for Validation and Implementation of *in vitro* Toxicity Tests," *Journal of the American College of Toxicology* 12:23-30, *1993.*
7. Goldberg, A.M., Epstein, L.D., and Zurlo, J., "A Modular Approach to Validation," *In Vitro Toxicology,* forthcoming, 1995.
8. Krasavage, W., et al., "The Relative Neurotoxicity of Methyl-n-butyl ketone, n-hexane, and Their Metabolizes," *Toxicology and Applied Pharmacology* 52:433-441, *1980.*
9. National Research Council, Commission on Life Sciences, Committee on Neurotoxicology and Models for Assessing Risk, *Environmental Neurotoxicology* (Washington, DC: National Academy Press, 1992).
10. Nordic Council of Ministers, *Criteria Document for Evaluation of Existing Data: Occupational Neurotoxicity* (Denmark: 1992)
11. Shahar, A., and Goldberg, A.M. (eds.), "Model Systems in Neurotoxicology: Alternative Approaches to Animal Testing," *Progress in Clinical and Biological Research (New* York, NY: Alan R. Liss, Inc., 1987).
12. Tilson, H., "Neurotoxicology in the 1990' s," *Neurotoxicology and Teratology* 12:293-300, *1990.*
13. Tilson, H. A., MacPhail, R. C., and Crofton, K. M., "Defining Neurotoxicity in a Decision-making Context, " *Neurotoxicology 16;363-376, 1995.*
14. Tsukamoto, T., and Yonezawa, T., "Neurotoxic Effect of 2,5-hexanediol *in vitro,"* *Clinical Neurology* 18:697, 1978.
15. U.S. Congress, Office of Technology Assessment, *Neurotoxicity: Identifying and Controlling Poisons of the Nervous System,* OTA-BA-436 (Washington, DC: U.S. Government Printing Office, April 1990).

16. U.S. Environmental Protection Agency, "Neurotoxicity Functional Observational Battery; Motor Activity; Neuropathology; Schedule-controlled Operant Behavior," *Federal Register* 50:39458-39466, *1985.*
17. U.S. Environmental Protection Agency, *Pesticide Assessment Guidelines, Subdivision F: Neurotoxicity* (Washington, DC: U.S. Government Printing Office, March 1991).
18. U.S. Environmental Protection Agency, Interagency Committee on Neurotoxicology, "Principles of Neurotoxicology Risk Assessment," final report, *Federal Register* 59:42360-42404, *1994.*

5

Immunotoxicity

Michael L Luster, Nancy Kerkvliet, and Gary J. Rosenthal

As evidenced by recent documents prepared by the Office of Technology Assessment (9) and the National Research Council in 1992 (8) focusing on immunotoxicology, there has been growing interest and concern within the scientific and public communities on the capacity of environmental agents to perturb normal immune processes. The types of effects that may occur are often agent-specific as well as species-specific and include immuno-suppression in which either systemic or local immunity is targeted (e.g., lung or skin), hypersensitivity disease, manifested as respiratory tract allergies or contact allergic dermatitis, and in certain instances autoimmunity or increased autoantibodies without evidence of disease. In addition to environmental pollutants, other agents of concern have included certain therapeutics, consumer products and biological (e.g., the therapeutic use of recombinant materials). More recently, interest has also focused on potential immunological effects by such diverse agents as excessive UV-B light, electromagnetic fields and pollutants found in the indoor environment. Common indoor pollutants can include not only chemical agents but also bioaerosols such as viruses, bacteria, fungi, algae and protozoa which have the potential to act as either sensitizing agents or mediators of infectious disease.

IMMUNOSUPPRESSION - EXPERIMENTAL TESTING

The sensitivity of the immune system to suppression by exogenous agents is due as much to the general properties of the agent as to the complex nature of the immune system. Because of this complexity, the initial strategies devised by immunologists working in toxicology and safety assessment have been to select and apply a tiered panel of assays to identify immuno-suppressive or, in rare instances, immuno-stimulatory agents in laboratory animals. Although the configurations of these testing panels vary depending on the laboratory conducting the test and the animal species employed, they usually include measures for one or more of the following:

- altered lymphoid organ weights and histology;
- quantitative changes in cellularity of lymphoid tissue, peripheral blood leukocytes and/or bone marrow;
- impairment of cell function at the effecter or regulatory level; and/or
- increased susceptibility to infectious agents or transplantable tumors.

Some of the test panels that have been proposed for evaluating the immune system in experimental animals by various government agencies are shown in tables 5-1 and 5-2. Additional test panel proposals (3, 5, 10, 13) are described in the IPCS Environmental Health Criteria "Principles and Methods for Assessing Direct Immunotoxicity Associated with Exposure to Chemicals", which is in the final stages of publication. The tier testing approaches employed by these agencies are similar in design in that the first tier is a screen for immunotoxicity

This **chapter was adapted,** in part, from the International Programme on Chemical Safety, "Principles and **Methods** for **Assessing Direct Immunotoxicity Associated with Exposure to Chemicals," UNEP, ILO, WHO, September 1994.**

Table 5-1: Panel for Detecting Immune Alterations Following Chemical or Drug Exposure in Rodents[a] Adopted by NTP

Parameters	Procedures
Screen (Tier 1)	
Immunopathology	• Hematology: complete blood count and differential • Weights: body, spleen, thymus, kidney, liver • Cellularity: spleen • Histology: spleen, thymus, lymph node
Humoral immunity	• Enumerate IgM antibody plaque-forming cells to T-dependent antigen (sRBC) • LPS mitogen response
Cell-mediated immunity	• Lymphocyte blastogenesis to mitogens (Con A) • Mixed leukocyte response against allogeneic leukocytes (MLR)
Nonspecific immunity	• Natural killer (NK) cell activity
Comprehensive (Tier 2)	
Immunopathology	• Quantitation of splenic B and T cell numbers
Humoral-mediated immunity	• Enumeration of IgG antibody responses to sRBCs
Cell-mediated immunity	• Cytotoxic T lymphocyte (CTL) cytolysis • Delayed hypersensitivity response (DHR)
Nonspecific immunity	• Microphage function quantitation of resident peritoneal cells and phagocytic ability (basal and activated by MAF)
Host resistance challenge models (endpoints)[b]	• Syngeneic tumor cells • PYB6 sarcoma (tumor incidence) • B16F1O melanoma (lung burden) • *Bacterial models: Listeria monocytogenes (mortality); Streptococcus species (mortality)* • Viral models: Influenza (mortality) • Parasite models: Plasmodium yoelii (Parasitaemia)

[a]The test panel was developed using B6C3F1 female mice.

[b]For any particular chemical tested only two or three host resistance models are selected for examination.

SOURCE: Adapted from International Programme on Chemical Safety," Principles and Methods for Assessing Direct Immunotoxicity Associated with Exposure to Chemicals," UNEP, ILO, WHO, September 1994.

with the second tier consisting of more specific or confirmatory studies, host resistance studies, or in-depth mechanistic studies. At present, most information regarding these models comes from the U.S. National Institute of Environmental Health Sciences, National Toxicology Program (NIEHS/NTP) followed by the model developed at the National Institute of Public Health and Environmental Protection (RIVM) in Bilthoven, The Netherlands. These models are described in more detail, while the others are not since little, if any, data have been published on their performance. The first-tier screening at RIVM consists only of test for general parameters of specific and nonspecific immunity. In contrast, Tier I of the NIEHS-NTP panel includes "functional" tests in which an immune response is measured following *in vivo* antigenic challenge. These are considered the most sensitive indicator of immune integrity but not routinely conducted as part of subchronic toxicology studies as there is concern that immunization may compromise toxicity interpretation. At present, when the NIEHS-NTP protocol is used, functional tests are

performed in a separate groups of animals. In the RIVM screening battery (see table 5-l), histopathology of lymphoid organs is pivotal. Routine histopathology of lymphoid organs has been shown to be useful in assessing the potential immunotoxicity of a chemical, in particular when these results are combined with the effects observed on the weight of the lymphoid organs and sufficiently high doses of the chemical are tested. In the RIVM panel, if the results in tier I suggest immunotoxicity, tier II function studies can be performed to confirm and further investigate the nature of the immunotoxic effect. Information on structure-activity relationships of immunotoxic chemicals can also lead to the decision to initiate function testing. The choice for further studies depends on the type of immune abnormality observed. The second tier consists of a panel of *in vivo* and *ex vivo/in vitro* assays including cell-mediated immunity, humoral immunity, microphage and natural killer (NK) cell function, as well as host resistance assays. Recently, it was suggested that the NK cell assay be added to RIVM/s tier I since it does not require animal sensitization.

The RIVM approach is based on the Organization for Economic Cooperation and Development (OECD) proposed guidelines for testing of chemicals - ##407, Repeated Dose Oral Toxicity - Rodent: 28-day or 14-day Study - which suggests the maximum tolerated dose (MTD), to be used as the high dose level for studies. The standard exposure period is 28 days and the animal species routinely used is the rat. These tests can be performed in the context of studies aimed at determining the toxicologic profile of the compound. Testing is conducted on at least three dose levels, the highest dose being the MTD and the lowest producing no evidence of toxicity.

The most employed screening battery and presumably more sensitive than RIVM since it includes function tests in Tier I is that developed by the NIEHS-NTP (5; see table 5-2). Recently, the database generated from these studies, which consists of over 50 compounds, has been collected and analyzed in an attempt to improve the accuracy and efficiency of screening chemicals for immunosuppression and to better identify those tests that predict immune-mediated diseases (6, 7). While a number of limitations exist in the analyses, several conclusions were drawn:

1) Examination of only two or three immune parameters may be used to successfully predict immunotoxicants in mice. In particular, lymphocyte enumeration and quantitation of the T-cell dependent antibody response appear particularly beneficial. Furthermore, commonly employed apical measures (e.g., leukocyte counts, lymphoid organ weights) appear fairly insensitive;
2) A good correlation existed between changes in the immune tests and altered host resistance in that there were no instances when host resistance was altered without affecting an immune test. However, in many instances immune changes were observed in the absence of detectable changes in host resistance. This can be interpreted to reflect that immune tests are, in general, more sensitive than the host resistance assays;
3) No single immune test was identified which could be considered highly predictive for altered host resistance. However, several assays such as the PFC response, surface markers, thymic weights and DHRs, were good indicators and others, such as proliferative response to LPS and leukocyte counts, were relatively poor indicators for host resistance changes. Combining several immune tests increased the ability to predict host resistance deficits, in some cases to about 80%;
4) Considering that there exists a “background” level of infectious diseases in the population, it is possible that subtle changes in immune function may translate into a significant change in host resistance given that the population exposed is large enough. This can be demonstrated experimentally, but would be difficult to establish in a clinical study where neither the virulence nor dose of infectious agent can be controlled;

Table 5-2: Methods for Detecting Immunotoxic Alterations in the Rat Currently Being Evaluated at RIVM, Bilthoven, The Netherlands

Parameters	Procedures
Tier 1	
Non-functional	• Routine hematology, including differential cell counting • Serum IgM, G, A, and E determination; lymphoid organ weights (spleen, thymus, local and distant lymph nodes) • Histopathology of lymphoid tissue • Bone marrow cellularity • Analysis of lymphocyte subpopulations in spleen by flow cytometry
Tier 2	
Cell-mediated immunity	• Sensitization to T-cell dependent antigens (e.g., ovalbumin, tuberculin, *Listeria),* and skin test challenge • Lymphoproliferative responses to specific antigens *(Listeria)* • Mitogen responses (Con-A, PHA)
Humoral immunity	• Serum titration of IgM, IgG, IgA, IgE responses to T-dependent antigens (ovalbumin, tetanus toxoid, *Trichinella spiralis,* sheep red blood cells) by ELISA • Serum titration of T-cell independent IgM response to LPS by ELISA • Mitogen response to LPS
Microphage function	• *In vitro* phagocytosis and killing of Listeria *monocytogenes* by adherent spleen and peritoneal cells • Cytolysis of YAC-1 lymphoma cells by adherent spleen and peritoneal cells
Natural killer function	• Cytolysis of YAC-1 lymphoma cells by non-adherent spleen and peritoneal cells
Host resistance	• *Trichinella spiralis* challenge (muscle larvae counts and worm expulsion) • *Listeria monocytogenes* challenge (spleen and lung clearance) • Rat cytomegalovirus challenge (clearance from salivary gland) • Endotoxin hypersensitivity • Autoimmune models (Adjuvant arthritis, experimental allergic encephalomyelitis)

SOURCE: Adapted from International Programme on Chemical Safety,'' Principles and Methods for Assessing Direct Immunotoxicity Associated with Exposure to Chemicals," UNEP, ILO, WHO, September 1994.

5) Logistic and standard modeling, using a single large dataset indicated most immune function-host resistance relationships follow a linear rather than linear-quadratic (threshold) models suggesting that even the smallest change in immune function translates into some change in host resistance (table 5-3). However, because of the variability in the responses, it was not possible to establish linear or threshold models for most of the chemicals studied when the datasets were combined;
6) Finally, using one dataset methods were developed for modeling quantitative relationships between changes in selected immune assays and host resistance tests. It is impossible to determine at present how applicable these values will be for immunotoxic compounds with different immune profiles. However, as more analyses become available, our ability to estimate accurately potential clinical effects from immunological tests should increase.

There are, of course, a number of limitations in using such test panels. For example, some endpoints are currently not included (e.g., PMN activity, cytokine production). Furthermore, such test panels seldom examine the effects of chronic exposure, or whether tolerance or reversibility can result from the treatment. In humans, assays that involve *in vivo* antigenic challenge, which are

Table 5-3: Most Appropriate Relationships Describing the Host Resistance and Immune Test[a]

Host Resistance Test	Immune Test									
	PFC	CTL	MLR	Con A	LPS	sig+	Thy 1.2+	CD4+	CD8+	Thy/ BW
L. monocytogenes	L	L-Q	L-Q	L	L	L-Q	L-Q	L-Q	L	L-Q
PYBC Tumor	L	L	L	L	L	L-Q	L	L	L	L
S. pneumonia	L-Q	L	L-Q	L	L	L-Q	L	L	L	N

[a]L = linear; L-Q = linear-quadratic; N = neither linear nor linear-quadratic

SOURCE: Luster, et al., 1993.

usually accepted as the most sensitive and informative of immune tests, are considered "invasive" procedures since they involve immunization and, as such, are not usually feasible or practical for inclusion in human studies.

A variety of factors need to be considered when evaluating the potential of an environmental agent or drug to adversely influence the immune system of experimental animals. These include appropriate selection of animal models and exposure variables, inclusion of general toxicological parameters, and an understanding of the biologic relevance of the endpoints to be measured. Treatment conditions should take into account the potential route and level of human exposure, biophysical properties of the agent such as half-life and any available information on the mechanism of action. Dose levels should be selected which attempt to establish clear dose-response curves as well as a no-observable-effect-level (NOEL). Although in some instances it is beneficial to include a dose level which induces overt toxicity, any immune change observed at such a dose level should not be considered biologically significant since severe stress and malnutrition are known to impair immune responsiveness. If studies are being designed specifically to establish reference doses for toxic chemicals, additional exposure levels are advisable. In addition, inclusion of a positive control group with an agent that shares characteristics of the test compound may be advantageous under certain circumstances when experimental and fiscal constraints permit.

The selection of the exposure route should parallel the most probable route of human exposure, which is most frequently oral, respiratory or dermal. A requirement for accurate delivered dose may require the use of a parenteral exposure route. However, this may significantly alter the metabolism or distribution of the agent from that which would occur following natural exposure and prevents any evaluation of effects on local immune responses at the site of entry.

The selection of the most appropriate animal model for immunotoxicology studies has been a matter of great concern. Ideally, toxicity testing should be performed with a species that will respond to a test chemical in a pharmacologic and toxicologic manner similar to that anticipated in humans (i.e., the test animals and humans will metabolize the chemical similarly and will have identical target organ responses and toxicity). Toxicologic studies are often conducted in several animal species, since it is assumed that the more species showing a specific toxic response, the more likely that the response will occur in humans. For most immunosuppressive therapeutics, rodent data on target organ toxicities and the comparability of immunosuppressive doses have generally been predictive of later observations in the clinic. Exceptions to the predictive value of rodent toxicological data are infrequent but have occurred, such as in studies of glucocorticoids, which are lympholytic in rodents, but not in primates. Although certain compounds may exhibit different pharmacokinetic properties in rodents than in humans, rodents still appear to be

the most appropriate animal model for examining the immunotoxicity of non-species-specific compounds, based on established similarities of toxicological profiles as well as the relative ease of generating host resistance and immune function data. Comparative toxicologic studies should be continued and expanded, particularly for novel recombinant biological compounds and natural products, since their safety assessment will likely present species-specific host interactions and toxicological profiles.

In summary, it is clear that the current OECD guideline #407 or "standard" subchronic toxicological studies are not suited to adequately assess potential adverse effects of exposure to the test chemical on the immune system, since, with respect to immunologic parameters, it is restricted to total and differential leukocyte counting and histopathology of the spleen. An evaluation of this test scheme (12, 13) indicated that in a series of almost 20 chemicals over 50°/0 of the immunotoxic chemicals would not have been identified as such if the tests would have strictly adhered to the guideline. In fact, it is even doubtful if those chemicals that were indicated to be immunotoxic only on the basis of guideline #407 would have been identified as such. For example, in a toxicological experiment a small, but significant, change in the percentage of basophilic leukocytes, without any other parameter to suggest that an effect on the immune system might have been present, would of itself probably not be considered biologically relevant.

▮ IMMUNOSUPPRESSION / HUMANS

Establishing immune changes in humans is considerably more complex than in animals considering non-invasive tests are limited, exposure levels to the agent (i.e., dose) are difficult to establish and responses in the population are extremely heterogeneous. With respect to the latter, the variation in immune responses (genetic or environmental) can exceed a coefficient of variation greater than 20 to 30%. Because many immune changes in humans following chemical exposure may be sporadic and subtle, it is essential that recently exposed populations be studied and sensitive tests for assessing the immune system be performed. Since many of the immune tests performed in humans have a certain degree of overlap (redundancy), it is also important that a positive diagnosis be based not on a change in one test, but only if a profile (pattern) of changes occur, similar to that observed in primary or secondary immunodeficiency diseases. For example, low CD4:CD8 ratios are often accompanied by changes in skin tests to recall antigens. The World Health Organization (WHO) has recently prepared a monograph (4) which provides testing methods and pitfalls for examining immune system changes in humans. However, it should be noted that the selection of most of these tests were derived from observations in patients with primary immunodeficiency diseases. Such individuals suffer from severe recurring infections and the degree of immunosuppression would likely be considerably more severe than that induced by chemicals. Thus, the document may be of limited value for examining potential chemical-induced immuno-suppression, although it should provide a focus for further methods evaluation.

In lieu of the difficulties that exist in identifying chemical-induced immunosuppression in humans, establishment of exposure levels (e.g., blood or tissue levels) of the suspected chemical(s) would not only be useful but in many instances essential to determine a cause-effect relationship. It should not be necessary to observe clinical diseases in order for immunosuppression to be meaningful for several reasons. First, uncertainties exist regarding whether the relationship between immune function and clinical disease follows linear or threshold responses. For instance, in a linear relationship even minor changes in immune function would relate to increased disease, given that the population examined is large enough. While the relationship at the low end of the dose-response curve is unclear, obviously, at the high end of the curve (i.e., severe immunosuppression), clinical disease is readily apparent. This is exemplified

by increases in opportunistic infections that occur in AIDS patients. Secondly, clinical disease may be difficult to establish considering neoplastic diseases may involve a 10-20 year latency before tumor detection and increases in infections are difficult to ascertain in epidemiological surveys (e.g., increased numbers or severity of colds).

The Agency for Toxic Substances and Disease Registry with the CDC (ATSDR/CDC) and National Research Council's subcommittee on "Biologic Markers in Immunotoxicology" have proposed testing batteries which attempts to address many of the above described problems and pit-falls by implementing a comprehensive state-of-the-art immunological evaluation in conjunction with more traditional tests (8, 11). Many of these tests are similar to those used to identify chemical-induced immunosuppression in lab-oratory animals and should help to predict the probability of developing suppressed host resistance or clinical disease in humans. These tests are also recommended in a tiered approach.

HYPERSENSITIVITY

Chemicals that induce hypersensitivity response are often small, highly reactive molecules (haptens) or protein products and produce an antigen-specific immune response. The clinical characteristic that sets these responses apart from immune mechanisms involved in host defense is that the reaction is excessive and often leads to tissue damage. Clinical differentiation of allergic responses from non-immune irritant responses is their persistence and severity. Chemical-induced hypersensitivities fall into two categories distinguished not only mechanistically but temporally; 1) delayed-type hypersensitivity which is a cell-mediated response that occurs within 24-48 hours after challenge; and 2) immediate hypersensitivity which is mediated by immunoglobulin, most commonly lgE, and manifests within minutes following exposure to an allergen. The type of immediate hypersensitivity response elicited (i.e., anaphylactic, cytotocix, Arthus or immune complex) depends upon the interaction of the sensitizing antigen or structurally related compound with antibody. In contrast, delayed-type hypersensitivity responses are characterized by T lymphocytes, bearing antigen-specific receptors which, on contact with cell-associated antigen, respond by secreting cytokines. Hypersensitivity responses usually occur at potential xenobiotic portals of entry, such as the skin and respiratory tract. Mononuclear phagocytic cells (e.g., alveolar macrophages in the lung, Kupffer cells in the liver, and Langerhans cells in the skin) have a major role in mediating local responses initially via antigen processing and later via the release of reactive oxygen species and cytokines that modulate the recruitment and activation of additional cell types including PMNs and lymphocytes. In addition to leukocytes, other cell types are involved including keratinocytes in the skin, epithelial cells and fibroblasts.

Historically, the guinea pig has been used to test for potential sensitizers. In the primary exposure (induction phase), the guinea pigs are treated with the test agent intradermally and/or topically, followed by re-exposure(s) (challenge phase) to the same test compound, normally after a period of 10-14 days. Redness and swelling are measured at the site of the challenge exposure with a non-irritant concentration of test compound. Because guinea pigs are large, several graded doses of antigen may be tested and an entire dose-response curve can be generated by comparing skin reactions in individual animals. However, it is expensive to purchase as well as maintain guinea pigs, there are few inbred strains and immunological reagents are not widely available.

Many variations in procedures for guinea pig hypersensitivity assays exist (e.g., Buehler occluded, guinea pig maximization, split adjuvant); details of which can be found elsewhere (l). These guinea pig models are very sensitive and it has been suggested "too sensitive" in that false positives may occur. This argument may not be valid, however, as there are "sensitive" human populations which need to be considered.

Efforts are presently underway to replace the guinea pig assays with mouse models. Gad *et al.*

(3) have proposed a mouse ear swelling test (MEST). This procedure is similar technically to the guinea pig assay in that both induction and challenge phases are required but the response is quantitated by measuring an increase in ear thickness when the material for challenge is applied. At present, the most promising new assay is the local lymph node assay (LLNA) (2), In this procedure, the test material is applied topically in three successive daily applications to both ears. Control mice are treated with the vehicle alone. After 5 days of exposure, mice are injected with radioisotopically labelled DNA precursors (e.g., ^{3}H-thymidine), and single-cell suspensions are prepared from the lymph nodes draining the ears. At least one concentration of the test chemical must produce a three-fold increase or greater in lymphocyte proliferation in the draining lymph nodes of test animals compared with vehicle-treated control mice to be considered a positive. The primary advantage of this assay is that it minimizes the manipulation of animals. There is some question regarding its sensitivity which can be approved by pretreating the animals with vitamin A. The LLNA would represent a distinct improvement over conventional guinea pig tests. Presently, interlaboratory validation using the LLNA are underway.

SUMMARY

Adverse effects on the immune system that may occur from exposure to chemical agents include autoimmunity, hypersensitivity and immuno-suppression. The diverse pathogenesis of these diseases necessitates that different testing strategies be employed for their assessment. For autoimmunity, there are no models presently available for rapid screening. Autoimmune-prone and hyperimmune rodent models have been used to establish that certain compounds (e.g., lead) contribute to the etiology of autoimmune diseases but their utility in screening is unknown. For hypersensitivity, guinea pig models have been historically used. More recently, a mouse assay has been developed which appear to have similar sensitivity to the guinea pig but is neither more rapid nor reduces the number of animals required. The local lymph node assay (LLNA), which is undergoing extensive validation, should represent a marked improvement for screening purposes. Rapid tests for immunosuppression are currently available. The "gold-standard" test is quantitation of the antibody response following immunization with a T-dependent antigen such as sheep erythrocytes in rodents. Antibody responses can be determined in sera by ELISA or by the plaque forming cell response. In studies where groups of animals are not available for immunization, "non-functional" tests can be used such as described by the RIVM although sensitivity will be lost. Because of the complexities of the immune system, at present *in vitro* test models are not suited for screening. Screening tests need to take into account potential sensitive populations such as the developing immune system as well as wild-life. Regarding the latter, such studies are often hampered by a lack of suitable test reagents.

REFERENCES

1. Anderson, K. E., and Maibach, H. I., "Guinea Pig Sensitization Assays," *Current Problems in Dermatology* 14:263-290, *1985.*
2. Basketter, D. A., et al., "Interlaboratory Evaluation of the Local Lymph Node Assay with 25 Chemicals and Comparison with Guinea Pig Test Data," *Toxicology Methods* 1:30-43, *1991.*
3. Gad, S. C., et al., "Development and Validation of an Alternative Dermal Sensitization Test: Mouse Ear Swelling Test (MEST)," *Toxicology and Applied Pharmacology.* 84:93-114,1986.
4. Hinton, D. M., "Testing Guidelines for Evaluation of the Immunotoxic Potential of Direct Food Additives," *Critical Review of Food Science and Nutrition.* 32: *173-190, 1992.*
5. IUIS/WHO Working Group, "Laboratory Investigations," *Clinical Immunology: Methods, Pitfalls, and Clinical Indications* 49:478-497, *1988.*

6. Luster, M. I., et al., "Development of a Testing Battery to Assess Chemical-induced Immunotoxicity: National Toxicology Program's Guidelines for Immunotoxicity Evaluation in Mice," *Fundamental Applied Toxicology* 10:2-19, *1988.*
7. Luster, M.I., Portier, C., and Pait, D.G., "Risk Assessment in Immunotoxicology: Sensitivity and Predictability of Immune Tests," *Fundamental Applied Toxicology* 18:200-210, *1992.*
8. Luster, M. I., et al., "'Risk Assessment in Immunotoxicology: Relationships Between Immune and Host Resistance Tests," *Fund. Appl. Toxicol.* 21:71-82, *1993.*
9. National Research Council, *Biologic Markers in Immunotoxicology* (Washington, DC: National Academy Press, 1992).
10. U.S. Congress, Office of Technology Assessment, *Identifying and Controlling Immunotoxic Substance,* OTA-BP-BA-75 (Washington, DC: U.S. Government Printing Office, 1991).
11. Sjoblad, R.D., "Potential Future Requirements for Immunotoxicology Testing of Pesticides," *Toxicology and Applied Pharmacology* 4:391-395, 1989.
12. Straight, J. M., et al., *Immune Function Test Batteries for Use in Environmental Health Field Studies* (Washington, DC: U.S. Department of Health and Human Services, Public Health Service, 1994).
13. Van Loveren, H., and Vos, J. G., *Evacuation of OECD Guideline #407 for Assessment of Toxicity of Chemicals with Respect to Potential Adverse Effects to the Immune System* (Bilthoven, The Netherlands: National Institute of Public Health and Environmental Protection, 1992).
14. Vos, J. G., and Krajnc-Franken, M. A.M., "Toxic Effects on the Immune System/Rat," *Hemopoietic System,* T.C. Jones et al. (eds.) (New York, NY: Springer-Verlag, 1990).

6

Reproductive and Developmental

Robert E. Chapin, Sally Perreault-Darney, and George P. Daston

ABSTRACT': Apical tests for reproductive or developmental toxicity assess the potential for a compound to affect any of the thousands of steps involved in making gametes and in the successful development of a fully functional offspring. Conventionally, this is thought to require at least 21 days for rodent female reproduction and for development, and close to 70 days for spermatogenesis. Short-term tests can evaluate some subset of these processes, so multiple tests must be used. The best use of in vitro tests currently is for evaluating a series of structurally-related molecules with an endpoint which reports a specific type of toxicity known to affect at least some members of that class. Because no in vitro tests have been found to correlate well with the breadth of reproductive and developmental toxicity observed in vivo, test-tube or culture-based tests should not be used as a first-pass, general screen for these effects. Even though short-term (21 or 28 day) in vivo studies will miss a variety of transgenerational effects, they remain the best means of identifying the more potent developmental and reproductive toxicants.

We will review the rationale for the current versions of definitive tests for reproductive and developmental toxicity, the approaches taken in reducing the duration of these tests and documenting what is gained and lost by such alternatives. Finally, we will address in vitro and genotoxicity tests, and review briefly their advantages and shortcomings, and their relationship to in vivo developmental/reproductive toxicity results. To be explicit, this consideration moves from the best to the worst, in terms of confidence in the information generated.

It is the feeling of this group that good screens for toxicity evaluate as much of a process at once as possible. This is the standard against which we will judge the value of a potential screen.

DEFINITIVE TESTS FOR REPRODUCTIVE TOXICITY

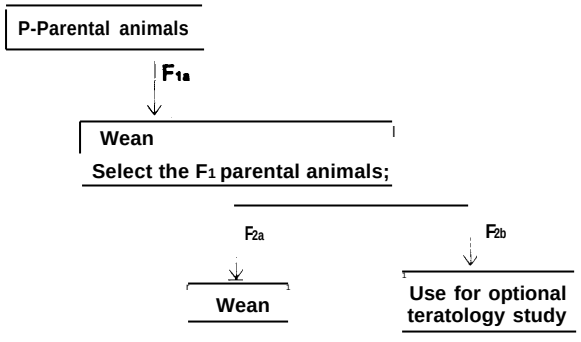
Definitive tests for reproduction and development are, essentially, set up to maximize confidence in a negative result. The definitive *in vivo* tests are apical, that is, they evaluate the integrated function of the entire system in one test. The benefit: if the results are negative, then one has reasonable assurance that there has been no effect anywhere in the process. The down-side is that identifying the location of a lesion or adverse effect can be slightly more time-consuming when starting from apical data.

As apical tests, these designs expose the entire process to the toxicant in question. Although this also depends on the pharmacokinetics of the compound in question, some default durations have been evolved, based on biology: For female rodent reproduction, the adult females should be exposed for 3-4 weeks prior to conception, as this exposes 4 or 5 estrous cycles of 5 or 4 days, respectively. In practice, the females may be exposed for 2-3 times this length of time, but 21-28 days is generally considered the minimum.

For spermatogenesis, the concept is to expose the gamete from a spermatogonium until it is ejaculated (again, the concept of exposing all stages of the process to the toxicant). In the rat, this is approximately 60-70 days. In practice, this often winds up being a 90 day exposure.

The field of developmental toxicology is in transition. In the past, the concept held that most terata resulted from exposures during organogenesis. In the rat, this begins about 6 days after mating, and continues until gdl5 (gestation day 15, out of a 22 day gestation period). Recent evidence shows that significant effects on the fetus can occur shortly after fertilization, and well before implantation (which occurs approximately on gd6), so many newer studies begin exposure of the pregnant female the day after mating, and continue until the day before delivery, when she

Figure 6-1: Sequence of a Two Generation Reproduction and Teratology Study



is killed and her fetuses examined for structural abnormalities. Alternatively (or additionally), some countries require testing of the offspring for behavioral abnormalities ("behavioral teratology"), which requires that in utero exposure be followed by 1-3 months of testing post-partum. Excepting these behavioral tests, this means a maximum of a 3 week exposure period for structural developmental toxicity.

Additionally, we should acknowledge that, while scientists have divided up the process into fields of specialization based on gender or process, Mother Nature is not so cut-and-dried. There is a considerable (and increasing) body of evidence that adverse reproductive effects on the offspring can be produced by a single in utero exposure to some compounds, or even by treating the adult parent before pregnancy is initiated. That is, treating a pregnant female with hormonally active compounds can produce permanent changes in her offspring. In the case of the reproductive system, these changes will not, of course, be visible until those offspring start to reproduce, a time lag of 2.5 months (rodents) to 16-20 years (humans). In these cases, the division between reproductive tox and developmental tox is well and truly blurred. (Note that this also can occur with systems other than reproduction; the postnatal manifestations of pre-natal exposure can be delayed until well into that offspring's life.)

This occurs because of the concept of "critical periods". Each organ system, as it develops, passes through a period (or multiple periods) where the signals received from (or through) the mother determine the long-term status of that system. This set-point is only adjustable for a short period (the "window" opens only briefly). The animal is vastly more sensitive to exposures while the window is open than at any other time in its life. For example, short exposures to TCDD at a specific point in gestation will permanently reduce the size of the gonads or the number of ovarian follicles. Slightly too much thyroid hormone (or a toxicant that mimics thyroid hormone) will have a similar effect, while too little thyroid hormone at a critical period will remove the signal to stop dividing, and testes in the adult will be permanently enlarged (perfectly normal, producing functional sperm, just bigger). This occurs for other systems as well: limited exposure to PCBs will permanently reduce the levels of circulating vitamin A in the kids, an effect with unknown consequences. The important concepts here are that: 1) the developing organism passes through some windows of vulnerability that do not exist in adults; and 2) changes made during these times can have permanent consequences for the offspring. The implications: 1) apical tests will (by definition) continue exposure during these times, and 2) short-term tests that ignore these windows increase the likelihood of missing a potentially significant toxic effect.

In practice, all of the previous considerations are folded together into a multigenerational test (figure 6-1) that starts off with either adult or pubertal animals (generally rats, and 20-30 of each sex per dose level), and exposes them to the toxicant in question for approximately 70-90 days, and then mates them within a treatment group (the high dose males mated to the high dose females, etc). Treatment continues while the dams are pregnant, after they have delivered, and then after weaning, the offspring are treated with the same dose their parents received. The pups are treated until they are about 70-80 days of age, when they are mated (again, within treatment levels), and another generation is produced. This second round of pups is killed either shortly after birth, or at weaning. In theory, this strategy should allow a compound to be identified as toxic

no matter where in the reproductive process it works (Recognize that senescence of reproductive function is not being examined in this scenario, and it is probable that a compound that reduces the number of ovarian follicles will not show up functionally, because the reproductive lifespan of the animal is not being assessed, only the beginning of the process is tested. It is possible that counting ovarian follicles may identify premature follicle loss, but this is rarely done.) The test is apical: it evaluates the entire process of reproduction, from stem cell gamete through finished pup, to the reproductive capabilities of that pup as an adult. It identifies heritable damage (to the gamete's DNA), as well as effects on lactation, parturition, etc.

Variations on this theme are common: two litters can be produced per generation, and one can be reared to evaluate second-generational effects, while the second can be assessed for structural abnormalities. The National Toxicology Program's (NTP) Reproductive Assessment by Continuous Breeding protocol (RACB) is more of a forced-breeding design, generating 4-5 litters in the first generation. The idea is that if the system is "pushed", adverse effects are more likely to be identified. Additionally, the extra litters take no more time, and produce vast increases in the statistical power to identify toxic effects. The Alternative Reproductive Test, developed by the U.S. Environmental Protection Agency's (EPA) Health Effects Research Lab, starts dosing the first generation at weaning, and generates several litters from the second generation. This maximizes the exposure of juveniles to the toxicants, a time period when hormonally-sensitive windows are known to be open.

Inherent in all these protocols are cell- and tissue-based assessments of the reproductive system at necropsy. This is necessary because fertility can be normal even though there are measurable reductions in, for example, gamete number: sperm count must be reduced significantly (by 50°/0-900/0) to reduce fertility in a male, while fewer follicles in a female rodent will not show up as reduced fertility until 4-7 months of breeding. So, conjoint with the in vivo fertility assessments are specific evaluations of the systems at necropsy (sperm measures, ovarian follicle counts, histopathology, etc). This disassembles the system some, providing preliminary information on the site of effect.

It is also important to note that these necropsy endpoints (organ weights, sperm assessments, estrous cyclicity) can all be added to the end of a 90 day subchronic test. This strategy is used routinely by the NTP to identify probable reproductive toxicants, and those compounds that deserve more definitive testing for reproductive toxicity.

To summarize: definitive tests for reproductive toxicity strive to expose all parts of the reproductive process to the putative toxicant. If no adverse effects are seen, there is some confidence that human risk from exposure to such a compound will likely be low.

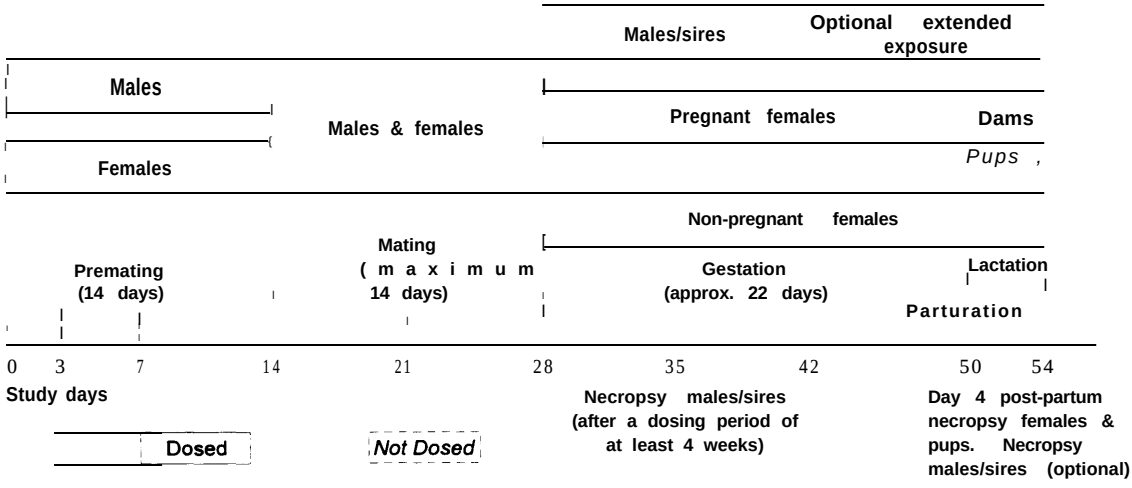
SHORT TERM TESTS

The greatest gains in reducing the duration of testing come from reducing the duration of exposure of the male, since the female and developmental toxicity portions of the definitive test are only about 3 weeks long. Thus, the short-term tests described below tend to truncate the male portion of the process the most, and make less drastic changes in identifying effects in females or fetuses.

An additional theme will become evident below: if a single apical test is going to be replaced, those replacements must be multiple, wherein each examines an individual part of the system (be it female reproduction or development or male reproduction). That is, a group of shorter tests can be acceptable if each component of the process is evaluated individually.

There have been a few designs evaluated for short-term assessments. The Organization for Economic Cooperation and Development (OECD) has recently sponsored a workshop which generated a shortened test for developmental and reproductive toxicity (figure 6-2). In this test, pairs of rats are treated for the duration of the test, which is approximately 54 days. Fourteen days after the start of treatment, males and

Figure 6-2: Experimental Schedule Indicating the Maximum Study Duration, Based on 14-Day Mating Period



females are cohabited for up to 2 weeks, and the females are allowed to gestate and deliver their young. The young are evaluated for a few days after birth, and are then killed, and the parents necropsied and examined. Given the different lengths of time to become pregnant, this test can be as short as 5 weeks, or longer than 7. While the published database for this design is small, businesses and companies around the world are using this design currently. A large database should be available in a few years.

The NTP took an even more stringent approach, and reduced the time further, to 21 or 28 days, depending on which version is being discussed. At each dose level, this design uses one group of males, and two groups of females. One group provides information on developmental toxicity, and is dosed only during gestation; the pups are evaluated after birth for survival and growth to pnd 4, since most severe structural/functional problems become evident by that time. The second group of females is used to assess female reproductive function. They are dosed for the entire duration of the study, evaluating the ability of the females to ovulate, mate, and implant. After impregnating the first group of females, the males are dosed for the remainder of the study, and necropsied at the end. Because only a fraction of spermatogenesis is exposed to the toxicant, this design relies heavily on histopathology of the male reproductive system to correctly identify male reproductive toxicants. While this sounds straightforward, correct and informed histopathologic interpretation of the testis is still relatively uncommon. The shortcomings of this test are that it cannot detect occult genetic or functional damage in the cells that does not manifest as structural damage. In regards to developmental toxicity for this design, notice that exposure continues for organogenesis, but the pups are not evaluated for structural abnormalities per se. Instead, the emphasis is on alterations that threaten the animal's ability to grow.

Finally, the Chernoff-Kavlock test doses the pregnant female during gestation, and evaluates the weight and number of pups for the first 4 days post-partum. This test is short, and quickly identifies life-threatening malformations, or reductions in lactational ability.

Note that all designs are in active use: the OECD design is being used worldwide to generate some data on compounds that currently have no data available; the NTP design is being used on various projects (including one to prioritize drinking water disinfection byproducts for EPA),

by the broad scientific community to help prioritize compounds for further evaluation.

TRADE-OFFS IN USING SHORT TERM TESTS

1. The duration of such studies reduces from, say, 21 weeks (for the in vivo portion of the test) to 7 weeks, or even 4 weeks. This is a significant time savings.
2. What we gain in time, we lose in our confidence in a negative answer. That is, we are not sure that a compound that tests negative is really non-toxic. This was demonstrated with the NTP design. The authors of this design knew that it would be capable of identifying positives, but that the key question was: how sensitive would the test be in identifying a slightly toxic compound? They tested four chemicals of varying known reproductive toxicity (tested in the Continuous Breeding design), and found that, as expected, this short-term test missed one (the least toxic); there were no adverse effects seen for one of the chemicals. Additionally, by their nature, these short tests will not identify adverse functional effects on the second generation, or premature reproductive senescence.
3. When the "system is disassembled", each component needs to be evaluated separately. That is, short tests need to consider each component of female reproductive function individually (ovulation, fertilization, implantation, gestation, delivery, nursing). A corollary of this is that, for males, time limits on short term tests preclude the proper evaluation of germ cell mutagenesis, or spermatogonial renewal.

In essence, this strategy divides chemicals into two categories: known positives, and unknowns. Put another way: there will be compounds that have been shown to produce toxicity, and those that were not toxic in the short-term test, but that may produce toxicity when evaluated for longer durations in more thorough designs. This toxicity may be slight, but it may also work through a window of vulnerability that was not evaluated by the short design.

These tests can have other endpoints "piggybacked" onto them. The NTP uses the males from the 28 day study to provide hematology and clinical chemistry data, as well as histopathology on somatic organs of interest (liver, kidney, etc). Incorporating these designs into a short-term strategy that evaluates a wide variety of endpoints and systems should pose no problem.

IN VITRO TESTS

In vitro tests are excellent for examining specific components of a process in isolation. For example, one can examine limb development in vitro and not worry about dispositional or detoxification processes interfering with the evaluation. They are also very appropriate for screening a group of compounds for a specific activity (for example, the ability of putative antibiotics to inhibit a bacterial cell wall synthetic enzyme). This use will be discussed further in the "New Strategies" section.

This very isolation is detrimental to a screening process. Good screens evaluate as much of a process at once as possible. Again, to cover in vitro what would be covered in vivo, multiple tests are needed.

Using male reproduction as an example, there are short (24-48 hr) in vitro methods for finding effects on spermatogenesis in vitro. However, to keep the cells alive, these methods are too short to correctly identify more than 20% of known testicular toxicants, they lack the testosterone-producing interstitial cells, and they lack the rest of the hormonal control systems (pituitary, hypothalamus). Thus, if there were going to be any confidence in the answer, this approach to male reproductive toxicity would require tests to evaluate those components of the system. The same is true for developmental toxicity and female reproduction: in vitro tests exist for some parts of each process, but not for all.

Since the overall process of reproduction and development is so complex, no "test-tube" assays have been evaluated as surrogates for in vivo testing. Receptor binding assays, second

messenger tests, or other molecular endpoints miss so many of the potentially vulnerable processes that this attempt has not even been made, to our knowledge.

In short, this is a two-edged sword. Coupled with the lack of confidence that a negative answer in vitro truly means a lack of toxicity in vivo, we cannot recommend at this time the use of in vitro tests to correctly identify toxicants.

Structure-activity relationships would likely provide some clues, but work in this area in conjunction with developmental and reproductive endpoints is still nascent. Early indications suggest that it can find application to the broader areas of reproduction and development, but it is too soon to tell.

REPRODUCTIVE TOXICITY AND GENOTOXICITY

It is theoretically possible that tests for genotoxicity would also identify reproductive and/or developmental toxicants. This hypothesis was evaluated using the NTP database, comparing the responses for a variety of genetox tests with outcome in the RACB test.

Overall, there are too few compounds tested in both systems to really evaluate the concordance, but generally, the results are not encouraging. Let us take the most promising relationship: if a compound was positive in the in vitro mouse lymphoma test, there was a 75% chance it would be positive in RACB. However, all compounds that were negative in lymphoma were also positive in RACB. So the sensitivity is reasonable, but the specificity is unacceptably low. Similarly, for in vitro cytogenetics, if a compound was toxic there, it stood approximately 70% chance of being toxic in RACB. However, if it was negative in cytogenetics, it still stood an 84°/0 chance of being toxic in RACB.

The preliminary indication is that we cannot hope that tests for genotoxicity will correctly identify reproductive toxicants. Based on more limited and personal evaluations over the years, our feeling is that the same is true for developmental toxicity.

NEW STRATEGIES AND TECHNOLOGIES

To deal with the cascade of new chemical structures, new approaches are needed. Three can be recommended:

1. Use benchmark dose analysis. There is a single reported application to (male) reproduction, but several reports in the recent literature for application to developmental toxicity. The attraction of BMD is that one could use half the animals that are used in a definitive test, and the model would deal appropriately with the consequent (and slight) reduction in certainty, while yielding a approximately 40-50% cost reduction. Halving the animal numbers also halves the animal care time, the dosing time, necropsy time, tissue prep time, pathology time, etc., although it does not reduce overhead or various preparative costs associated with those activities or others. This also would not change the duration of the test. Although still relatively new, BMD holds such promise as to warrant it is being raised as the most likely solid improvement for this field.
2. Use a tiered approach to requiring information. This may involve some preliminary information triggering a request for further specific tests. This is the case with the new EPA Reproductive Toxicity Testing Guidelines: if changes in epididymal sperm count are found, a count of testicular spermatid nuclei is requested, both for confirmation and to identify site of effect.
3. Use SAR. If a previously-registered compound that reduced sperm motility is structurally related to a new candidate, requesting motility information on this candidate is a reasonable and targeted request. There are such huge benefits to be derived from computer-driven SAR methodologies that further work in this area is clearly warranted. The impact is biggest where the costs are greatest (which generally correlate with duration of exposure or numbers of manipulations of animals).

New technologies include the use of the computer for a variety of tasks: counting sperm and

measuring sperm motion, counting and sizing ovarian follicles and stages of spermatogenesis (using image analysis techniques), and collating and producing an overall toxicologic profile. With the awareness that many transgenerational effects appear to result from the binding of xenobiotics to specific hormone receptors, one could imagine a screen of in vitro tests that assess the ability of a new compound to bind to a variety of hormone receptors and stimulate transcription. Such a vision has been proposed by others, with such receptor systems transected into cultured cells, so that the assays become in vitro cell culture systems. These are still in the planning stages, and it should be noted that, while simple in concept, they present significant technical challenges. Finally, transgenic animals are gaining acceptance as interesting model systems with some significant potential for application. While it is too soon to tell whether transgenics will be useful in identifying and ranking developmental/reproductive toxicants, we will note that many transgenics do have significantly reduced gametogenesis/fertility; whether this is a benefit or a drawback would depend on the question being asked, and the way in which it is being asked. This may be worth some additional consideration in the future.

SUMMARY

Several strategies can be employed to significantly reduce the time and expense of preliminarily identifying reproductive and/or developmental toxicants. Each reduction in time and cost brings with it a concomitant reduction in certainty that a lack of toxicity over the short term also means a lack of toxicity over a longer exposure. Such tests are best used to prioritize compounds for further testing and evaluation. Benchmark dose and SAR strategies also can be viewed as valuable tools in the struggle to maintain public health at the least possible expense. If these reductionist strategies are not used to entirely replace longer, more definitive, tests, they can be used with confidence and success.

7

Environmental Toxicology: Testing and Screening

Maurice Zeeman, Anne Fairbrother, and Joseph. W. Gorsuch

ABSTRACT: *In response to Congress, the Office of Technology Assessment (OTA) is preparing a study on the Toxic Substances Control Act (TSCA) to evaluate the Existing Chemicals Program. The purpose of the Chemical Testing and Screening Workshop was to identify the present and future methods of screening and testing of commercial chemicals using nine specific endpoints, one being environmental toxicology (i. e., ecological effects assessment). This chapter addresses the state of the science by responding to several specific questions asked by the OTA (e.g., "what are the best tests available to identify a chemical of concern and to evaluate its toxicity?'?. This chapter concludes that basic screening and testing methods are already being applied by EPA/OPPT, especially by the use of structure-activity relationships (SARs/QSARs) for ecotoxicity screening purposes, and by the use of rapid and inexpensive tests to actually assess ecotoxicity. Areas for improving existing methods include sorting priorities to assess chemical exposure information and SARs/QSARs for avian species, plants, earthworms, and sediment dwelling organisms.*

One of nine specified topics of interest addressed at the Office of Technology Assessment (OTA) Workshop was the testing and screening methods used by the U.S. Environmental Protection Agency (EPA) and others to assess environmental toxicology. Methods for "environmental toxicology" were understood to mean screening and testing methodologies used to assess potential ecological effects on organisms found in the environment from TSCA-regulated chemicals.

The Toxic Substances Control Act of 1976 (TSCA) provided the EPA Office of Pollution Prevention and Toxics (OPPT) with authority to require development of adequate data for assessing the risk to human health and the natural environment from industrial chemicals identified as having risk potential. "Protection of the environment" means different things to different people. To some it means maintaining a place where humans can live and be healthy. To others, it is tied to commodity production or extraction. Still others look for a system that looks and functions as it did prior to the arrival of Europeans in North America and that has the capacity to sustain all native plants and animals. Congress purposefully left this definition vague in almost all environmental legislation in order to allow continued public debate to frame the question. Nevertheless, implementation of TSCA requires the EPA to explicitly describe what "protection of the environment" means within this context, in order to request the proper information to evaluate whether a chemical has the potential to significantly degrade that environment.

Within OPPT, the Environmental Effects Branch (EEB) has provided the scientific and technical evaluation of environmental/ecological hazard of industrial chemicals, and has determined the type and adequacy of data needed to identify and assess their possible adverse effects. Over the past 15 years this group has provided significant direction to, and rationale for, how ecological hazard and risk assessment activities have been addressed under TSCA (26, 29, 30).

Environmental protection can occur at many different levels of ecological organization. Traditionally, wildlife and fisheries managers have protected populations while plant ecologists look for healthy, evolving communities. The Endangered Species Act (ESA) requires protection of the health of individual organisms. Animals,

The contents of this chapter do not necessarily reflect the views and policies of the Environmental Protection Agency.

however, have differing strategies for population maintenance. Small mammals (rodents, etc.) have a large reproductive capacity to balance the high annual mortality (up to 80% of the population in some cases). In this instance, a little additional mortality from environmental contaminants would be inconsequential, while a reproductive inhibitor could have longer-term effects. Conversely, most large animals such as elephants and eagles have a long life span, relatively low annual mortality, and a low reproductive rate. Loss of one or two reproductive seasons would have little effect on these populations as the adults would survive to reproduce another year. However, increased mortality of adults due to an environmental contaminant would severely depress the population.

While species differ in their life history strategies, it is intuitively obvious that increased mortality and decreased reproduction will affect the population over the long-term. The amount of changes in these parameters that is "significant" depends on the species and the community with which it is associated. Compensatory changes in reproduction, predation, competition, etc., all affect the severity of the impact of chemical-induced effects. Thus, TSCA-related ecological risk assessments include measures of lethality (LC_{50}) and reproductive effects with an associated uncertainty ("assessment") factor to accommodate our imprecise knowledge of ecological systems (29). Sublethal effects (immune suppression, endocrine disruption, neurotoxicity, etc.) could potentially influence population demographics, but in more subtle ways that have not yet been clearly established. Therefore, their inclusion as endpoints upon which regulatory decisions can be based is still open for debate.

To assure that adequate ecotoxicity data are developed to assess the possible adverse ecological effects of industrial chemicals, screening methods, test procedures, and guidelines have been established by OPPT (26, 29, 30, 3 1). For example, several hazard assessment structure-activity relationship (SAR/(Q)SAR) screening methodologies have been developed and refined in OPPT specifically y for the data-poor new chemical assessment process (4, 5, 6, 13, 14, 15, 28). These (Q)SAR methodologies are now being applied to the hazard and risk screening of TSCA-regulated existing chemicals (3, 30). It should be noted that (Q)SARs have been used to predict toxicity, biodegradability, and bioaccumulation (18).

BEST AVAILABLE TESTS

The best available tests to identify and evaluate chemical toxicity will depend on the potential risks, the uncertainties, the natural resources at risk, and the resources available for analysis. No single test is "best" for all situations. The best tests to assess the ecological effects of a chemical of concern, i.e., the most ecologically relevant and producing the most accurate results, would most likely be field assessment tests. These should identify where the TSCA existing chemicals are being released or applied in the field and also assess the impacts on the numerous types of organisms that exist in the environments that are being exposed. Depending on the location and size of the area of concern and the level of biological focus, the number of species potentially exposed and impacted could vary from dozens, to hundreds, thousands, or even millions of species.

Although most meaningful ecologically, field testing seldom would be conducted without prior knowledge of the potential toxicity of the chemical to plants or animals, particularly at concentrations expected to be found in the environment. Field tests are very expensive (in the order of several million dollars) and are technically difficult to conduct. In addition, it is difficult to communicate the significance of such study results to chemical industries, regulatory decision-makers, and the public. Multiple stressors: chemical, biological, and physical, are often difficult to differentiate in populations/communities. However, *in situ* effluent biomonitoring frequently is done in aquatic situations as a bioassay for toxicity detection (10).

In fact, most ecological risk assessments of TSCA-regulated chemicals are oriented toward

the aquatic environment. This is because the majority of environmental releases are presumed to be aquatic releases. Air releases are another route of environmental exposure and may influence terrestrial systems as well, but with the exception of smelters and intentional applications to land (e.g., dioxins in sludges applied to forests and pastures), adverse terrestrial effects from this or any other source have not been well documented (17, 25, 26).

Another main reason is that seldom can an adequate regulatory case of significant exposure, hazard, and risk to organisms in the environment be provided to warrant field testing. Typically, the majority of cases where TSCA-regulated existing chemicals are known to be released into the environment and resulted in exposures of organisms, has often focused on chemical production releases into the aquatic environment. As a result of estimated environmental dilution, and adsorption to particulate, this frequently ends up in predictions of very low chemical exposures and risks. For the terrestrial environment, only a few examples of potential exposures and effects have even been assessed, let alone been considered for any form of field testing (17,25, 26).

As we move away from field testing, because they are so complex and expensive, to other more derived test methods, that may be less meaningful ecologically, our ability to accurately predict the overall effects of a chemical may be compromised. One of the more feasible surrogates for testing in the field is mesocosm or microcosm testing of chemicals. However, these tests can also be fairly lengthy, moderately expensive, and their results difficult to interpret and defend, as is the situation with field study results.

The next most ecologically realistic and important level of testing is long-term ecotoxicity testing performed in the lab. If such tests are of sufficient duration, they can be designed to evaluate the potential impacts of a chemical on the mortality, growth, development, and reproduction of field populations (or of appropriate surrogates for these species). Test durations long enough for whole life-cycle testing are preferred, but such test results are seldom available for industrial chemicals. More available, but still relatively uncommon, are industrial chemical results (e.g., maximum acceptable toxicant concentration (MATCs)) from different versions of the 30-90 day fish early life-stage test, or the 14-21 day partial life-cycle test for some aquatic invertebrates, such as *Daphnia*.

Short-term (e.g., 2-4 days) ecotoxicity testing results of acute lethality (i.e., LC_{50} or EC_{50} values) are usually the most readily available (but perhaps less ecologically meaningful) results found for existing industrial chemicals. From such limited test results, estimations of longer-term impacts can be made by using uncertainty factors to set potential exposure levels where ecological risks may occur **(31).**

Practically speaking, short-term testing of fish, aquatic invertebrates, and algae (the three basic trophic levels found in many aquatic food chains) represents most, if not all, of the testing performed for industrial chemicals (29). The primary reasons for this are the rapidity and inexpensiveness of these short-term tests. Performing a 48-hr daphnid EC_{50} test is quick, and if it is only used for internal chemical screening purposes (e.g., does not follow Good Laboratory Practices (GLPs) standards), would cost approximately $1,000 to accomplish. However, without chemical test concentration verifications, some test results might be of little value in predicting what would happen if the chemical were released into the environment. Similarly, inexpensive and quick tests are available for screening chemicals for toxic effects to plants. The germination and root elongation test (5-7 days in length) (11, 21), the vegetative vigor test (14-21 days in length) (20), or seedling growth tests (1, 9) could support the evaluation of the potential impact of a chemical in soils. Even limited acute ecotoxicity test data are preferred to no data at all.

Proposed cellular and molecular toxicity endpoint tests (e.g., promoter gene activation, stress protein induction, Ah receptor binding) may be useful for providing information about modes of action for a chemical and, therefore, direct concern towards particular species that may be most sensitive in this response. For example, a chemi-

cal that is shown to be induced through the Ah receptor would raise concern for mustelids (mink) and songbirds, but not for waterfowl. These suborganismal tests do not provide enough information about ecologically relevant effects (e.g., "significant" change in mortality or reproduction) to form the basis of a regulatory decision without additional information. It will take many years of research to develop relationships between gene induction and changes in population growth rates, and realistically, it may not be possible to do.

PRIORITY SETTING

OPPT SAR/(Q)SAR methodologies were developed for estimating ecotoxicity in order to screen thousands of chemicals per year in a very short time frame. Screening assessments typically occur with little or no ecotoxicity data being provided by chemical sponsors of the industrial chemicals submitted to EPA.

One primary use of (Q)SAR technology has been to set testing priorities by estimating how toxic a chemical may be to aquatic organisms. If this estimate results in a prediction of sufficient risks in the environment, the sponsor is encouraged to consider performing testing to define the actual toxicity of that industrial chemical. As discussed in detail elsewhere, these quick and inexpensive (Q)SAR methods have been used extensively in assessing the over 26,000 new industrial chemicals submitted to OPPT from 1979 through 1994 (1, 2,27,29, 30).

TRADEOFFS IN SCREENING

One potential problem is uncertainty about the accuracy of the ecotoxicity that is predicted by (Q)SAR. The (Q)SAR values themselves are only estimations of toxicity. They are only as good estimates as are possible based upon the ecotoxicity values present in the data set for the chemical class or biological activity being predicted (18). In general, the larger the number of chemical toxicity values that are present in the data set for an appropriate chemical class, the higher the chances are that the ecotoxicity predictions for that class are accurate.

OPPT SAR estimations can vary from simple similarities, such as using test data available for a similar chemical grouping or analogs, to being able to provide quantitative estimates of ecotoxicity. Quantitative estimates are possible when an empirical mathematical relationship has been established for a chemical grouping/class to which the new chemical also belongs. OPPT has developed over 120 (Q)SARs for about 45 classes of industrial chemicals (23, 24, 31).

Except for earthworms, the OPPT SAR database is limited to aquatic organisms. Similar models for terrestrial organisms (other than laboratory animals used for human risk assessments) need to be developed. Sufficient data exist for some classes of chemicals so that this could be done for plants, birds, and mammals. However, the database of toxicity information for reptiles and amphibians is too sparse to allow SAR models to be developed for herptofauna.

The aquatic (Q)SARs for some chemical classes result in a hazard profile of six ecotoxicity values that estimate both the acute and chronic toxicity of such chemicals to fish, daphnids, and algae, respectively (table 7-l). Typically not enough chemical ecotoxicity data exist to construct (Q)SARs for all parts of these hazard profiles (e.g., sometimes only one, two, or all three of the acute ecotoxicity values can be predicted). Some (Q)SARs may also be based upon data for only a few chemicals in the class.

Table 7-1: OPPT Standard Hazard Profile for Aquatic Ecotoxicity

Freshwater Test Descriptions
Fish Acute Toxicity (96hr LC_{50})
Daphnid Acute Toxicity (48hr EC_{50})
Algal Toxicity (96hr EC_{50})
Fish Early Life Stage (28-90 day MATC)
Daphnid Partial Life Cycle (14-21 day MATC)
Algal Toxicity (96hr NOEC)

SOURCE: Zeeman, M., "Ecotoxicity Testing and Estimation Methods Developed Under Section 5 of the Toxic Substances Control Act (TSCA)," *Fundamentals of Aquatic Toxicology: Effects, Environmental Fate, and Risk Assessment,* Chapter 23, G. Rand (ad.) (Washington, DC: Taylor & Francis, 1995)

However, the vast majority (95-98%) of the discrete organic chemicals found on the TSCA inventory come from only 7-10 chemical classes (7, 28, 30). Therefore only a smaller subset of the OPPT (Q)SARs will need to be used. Furthermore, several of the most commonly used (Q)SARs are also for chemical classes with a relatively large amount of ecotoxicity data used to construct the model. Therefore, they are among the more reliable (Q)SAR estimation methods.

VALIDATION/ REPRODUCIBILITY

As discussed in detail elsewhere, the validation of these OPPT (Q)SAR ecotoxicity estimation methods is an ongoing process (3, 7, 14, 15, 30). The validation of the OPPT (Q)SARs used for assessing the aquatic toxicity of new industrial chemicals has been performed and results were published in the peer-reviewed literature (1 5).

In addition, a recent joint EPA/ European Union (EU) project independently assessed the accuracy of a variety of the SARs used by OPPT for estimating the physical/chemical parameters, environmental fate, human health, and ecotoxicity of industrial chemicals. This study compared the (blinded) U.S. predictions with the limited base set of test data received by the EU for their new chemicals (16, 22). (For example, only acute toxicity data for fish and daphnids are required as the ecotoxicity base set by the EU at the time of this study). In this "Structure Activity Relationship/Minimum Premarketing Dataset" (SAR/MPD) study, the European Union experts concluded that the EPA/OPPT ecotoxicity (Q)SAR methodologies "performed extremely well in predicting acute toxicity to fish and Daphnia" (16, 30).

Significant attempts have been made to make these OPPT ecotoxicity screening methods available to the public. The 1988 version of the OPPT (Q)SAR Manual (4) was widely distributed, both nationally and internationally. It has been updated and currently contains about 120 OPPT SAR/QSARs available for assessing the ecotoxicity of about 45 classes of chemicals (23). A computer program was also developed that incorporates the revised OPPT (Q)SAR Manual and it was recently released as a PC Version, called ECOSAR. ECOSAR is publicly available (24), has been widely demonstrated and distributed, e.g., in national and international fora, such as at trade association meetings, scientific meetings, and public meetings.

RECEPTOR AND MECHANISM-BASED ASSAYS AND SAR

Knowing the mechanism(s) by which a chemical impacts specific receptors of organisms and thereby causing adverse effects is a highly desirable scientific goal. In human health risk assessments, extrapolations of toxic effects are made from several species to one species. In ecological risk assessment, on the other hand, extrapolations must be made from data on a few species to many thousands of species, and from individuals to populations. Information on mechanisms of action of new chemicals (e.g., inhibition of the enzyme AChE essential to nerve conduction), coupled with knowledge of the comparative physiology of various plant and animal classes would allow toxicity estimates to be made to a wide variety of species without the need for empirical testing.

However, such mechanistic approaches are not currently feasible as the information on which to base them is lacking. Moreover, the technical expertise required to make such comparative physiological-based toxicology interpretations is scarce. This can prove especially difficult when it is necessary to rapidly screen and assess very large numbers of chemicals that also have widely different structures. Pragmatically speaking, that is why the development and use of chemical class specific (Q)SARs have been such a priority for OPPT in its need to routinely assess the ecotoxicity of the thousands of industrial chemicals reviewed each year from industries.

INTEGRATED SCREENING AND TESTING STRATEGIES

A comprehensive evaluation of an industrial chemical would require not simply (Q)SAR estimations of ecotoxicity, but also data from acute,

subchronic, and chronic toxicity testing in a variety of appropriate environmental species (or their lab surrogates). Where feasible, microcosm and mesocosm studies should be performed and, when significant exposures are anticipated, even field testing should be considered.

The types and utility of several diverse ecotoxicity testing methods that are readily available to assess industrial chemicals have been determined for aquatic and terrestrial environments and have already been implemented by OPPT (19, 26). OPPT has also developed a tier-testing strategy (19, 26, 31) that allows for a sequencing to move from the quick and simple ecotoxicity tests to the test methods that are more long-term in duration and, therefore, more expensive.

It should therefore be relatively simple to integrate new test methods and results into an overall screening and testing strategy for a comprehensive evaluation of an industrial chemical.

NEW DEVELOPMENTS

During the next decade several different types of new test technologies will certainly be developed. It is very important to keep in mind that there needs to be a reality check on the direct utility of such test technologies for ecological effects assessment, including their applications (e.g., relevance, cost, and exposure routes). Because we are, or will be, capable of performing specific tests does not make them useful or meaningful for the purpose of determining their ability to detect the significant effects of a chemical on organisms in the environment. The basic issue is whether a technology will help in determining if the chemical of concern can affect the mortality, growth, development, and/or reproduction of the populations of organisms that exist and interact in the environment.

Because ecotoxicity test data are becoming available for earthworms, OPPT has developed and is starting to use, a (Q)SAR for neutral organic industrial chemicals for these terrestrial species. However, the need for additional ecotoxicity data for other terrestrial and sediment-dwelling species is well documented. The places where such additional terrestrial (Q)SARs could be most useful are for plant and avian species. Also needed are (Q)SARs for sediment-dwelling species, such as burrowing worms and crustaceans. Mammalian SARs should also be broadened beyond the laboratory animal data to integrate information on carnivore and ruminant species, as well as information available on wild rodents. This will broaden the basis for the SAR and may confound the model used for human health risk assessment, but it will become much more helpful for ecological risk assessments. It may be that two mammalian SARs can be developed: one that utilizes all the data and one that uses a subset specifically directed toward making extrapolations for humans only. The EPA/OPPT (Q)SAR program can be used to help direct the current controversy about chemicals that are endocrine disrupters both in humans (e.g., reduced sperm counts) and wildlife (e.g., abnormal breeding behaviors of gulls) (8), and has increased our awareness that the types of adverse impacts that some believe have occurred for many years to several species in the Great Lakes may also be happening to humans.

CONCLUSION

One of the main reasons that society should care more about what happens to organisms in the environment is that these organisms serve as monitors of what chemicals are capable of doing to other living organisms, such as humans (10). It is very easy to think that significant impacts to nonhumans, which may mean nothing at all, will happen to us.

The screening tools and test methods that have been developed by OPPT and other researchers to assess ecotoxicity are reasonable and cost-effective. Society decides how much it is willing to spend to generate ecotoxicity data. The reasonableness to adequately assess the potential impacts of industrial chemicals should be based upon what quality of data science indicates can be reasonably expected and needed to derive a specific level of certainty around risk or safety assessments. Excellent ecotoxicity screening tools

already exist (e.g., (Q)SAR) and are being used by OPPT to screen and prioritize several thousands of the discrete organic existing chemicals on the TSCA inventory for their potential to persist, bioconcentrate, and be highly toxic to organisms in the aquatic environment (6, 28, 30, 31).

Similar screening methods are still needed for organisms in the terrestrial environment. However, it must be recognized that exposure scenarios in terrestrial environments are much more complex than those in aquatic systems and may not be amenable to incorporation into (Q)SAR models in a similar fashion. Furthermore, our knowledge of actual long-term ecosystem effects of chemicals in the environment will remain rudimentary unless well-designed monitoring studies can be put in place. This type of "adaptive management" would allow us to verify our predictions and alter our management strategies accordingly, while allowing chemicals to remain in commerce.

REFERENCES

1. American Society for Testing and Materials, "Standard Practice for Conducting Early Seedling Growth Tests," *Annual Book of ASTM Standards* (Philadelphia, PA: 1994).
2. Auer, C. M., Nabholz, J. V., and Baetke, K.P., "Mode of Action and the Assessment of Chemical Hazards in the Presence of Limited Data: Use of Structure-Activity Relationships (SAR) Under TSCA, Section 5," *Environmental Health Perspectives* 87:183-197, *1990.*
3. Auer, C. M., et al., "SAR – The U.S. Regulatory Perspective," *SAR & QSAR in Environmental Research* 2(1-2):29-38, 1994.
4. Clements, R. G., et al., *Estimating Toxicity of Industrial Chemicals to Aquatic Organisms Using Structure-Activity Relationships,* Environmental Effects Branch, Health & Environmental Review Division, U.S. Environmental Protection Agency, EPA-560-6-88-001 (Washington, DC: 1988).
5. Clements, R.G., et al., "The Use and Application of QSARs in the Office of Toxic Substances for Ecological Hazard Assessment of New Chemicals," *Environmental Toxicology and Risk Assessment: 1st Volume,* W.G. Landis, J.S. Hughes, and M.A. Lewis (eds.) (Philadelphia, PA: American Society for Testing and Materials, 1993).
6. Clements, R. G., et al., "The Use of Quantitative Structure-Activity Relationships (QSARs) as Screening Tools in Environmental Assessment," *Environmental Toxicology and Risk Assessment: 2nd Volume,* J.W. Gorsuch, et al. (eds.) (Philadelphia, PA: American Society for Testing and Materials, 1993).
7. Clements, R. G., et al., "The Application of Structure-Activity Relationships (SARs) in the Aquatic Hazard Evaluation of Discrete Organic Chemicals," *SAR & QSAR Environmental Research,* in press.
8. Colborn, T., and Clement, C. (eds.), *Chemical-Induced Alterations in Asexual and Functional Development: The Wildlife/Human Connection* (Princeton, NJ: Princeton Scientific Publishing Co., 1992).
9. Food and Drug Administration, Center for Food Safety and Applied Nutrition and Center for Veterinary Medicine, Environmental Impact Staff, "Environmental Assessment Technical Assistant Document 4.07: Seedling Growth," *Environmental Assessment Technical Handbook* (Washington, DC: 1987).
10. Glickman, L., et al., *The Use of Animals as Sentinels of Environmental Health Hazards* (Washington, DC: National Academy Press, 1991).
11. Gorsuch, J. W., Kringle, R. O., and Robillard, K. A., "Chemical Effects on the Germination and Early Growth of Terrestrial Plants," *Plants for Toxicity Assessment,* W. Wang, J.W. Gorsuch, and W.R. Lower (eds.) (Philadelphia, PA: American Society for Testing and Materials, 1990).
12. INFORM: Toxics Watch 1995 (New York, NY: INFORM, Inc., 1995).
13. Nabholz, J. V., "Environmental Hazard and Risk Assessment Under the United States

Toxic Substances Control Act," *Sci. Total Environ.* 109/110:649-665, *1991.*

14. Nabholz, J.V., Miller, P., and Zeeman, M., "Environmental Risk Assessment of New Chemicals Under the Toxic Substances Control Act TSCA Section Five," *Environmental Toxicology and Risk Assessment: 1st Volume,* W.G. Landis, J.S. Hughes, and M.A. Lewis (eds.) (Philadelphia, PA: American Society for Testing and Materials).
15. Nabholz, J.V., "Validation of Structure-Activity Relationships Used by the USEPA's Office of Pollution Prevention and Toxics for the Environmental Hazard Assessment of Industrial Chemicals," *Environmental Toxicology and Risk Assessment: 2nd Volume,* J.W. Gorsuch, et al. (eds.) (Philadelphia, PA: American Society for Testing and Materials, 1993).
16. Organization for Economic Cooperation and Development, *US EPA/EC Joint Project on the Evaluation of (Quantitative) Structure-Activity Relationships (QSARS),* OECD Environment Monographs No. 88 (Paris, France: 1994).
17. Rabert, W., and Zeeman, M., "Dioxins/Furans: U.S. EPA Ecological Risk Assessment for Land Application and Disposal Methods for Paper Pulp Sludge," *Chemosphere 25: 1499-1504, 1992.*
18. Richard, A. M., et al., "Testing and Screening Technologies for Review of Chemicals in Commerce: SAR/MODELING", Proceedings of OTA *Workshop on Chemical Testing and Screening Technologies for Review of Chemicals in Commerce* (Washington, DC: Office of Technology Assessment, April 24-25, 1995).
19. Smrchek, J., et al., "Assessing Ecological Hazard Under TSCA: Methods and Evaluation of Data", *Environmental Toxicology and Risk Assessment: 1st Volume,* ASTM STP 1179, W.G. Landis, et al. (eds.) (Philadelphia, PA: American Society for Testing and Materials, 1993).
20. U.S. Environmental Protection Agency, Office of Pesticides and Toxic Substances, *Pesticide Assessment Guidelines, Subdivision J. Hazard Evaluation Nontarget Plants,* EPA540/9-82-0020 (Washington, DC: 1982).
21. U.S. Environmental Protection Agency, Office of Research and Development, *Protocols for Short-Term Toxicity Screening of Hazardous Waste Sites, A.8.6.* Lettuce Seed Germination *(Lactuca sativa),* and A.8.7. Lettuce Root Elongation *(Lactuca sativa)* (Corvallis, OR: 1988).
22. U. S. Environmental Protection Agency, Office of Pollution Prevention & Toxics, Chemical Control Division, US *EPA/EC Joint Project on the Evaluation of (Quantitative) Structure-Activity Relationships* (Washington, DC: July 1993).
23. U.S. Environmental Protection Agency, Office of Pollution Prevention and Toxics, Health and Environmental Review Division, Environmental Effects Branch, *Estimating Toxicity of Industrial Chemicals to Aquatic Organisms Using Structure-Activity Relationships,* EPA-748-R-93-001 (Washington, DC: 1994).
24. U.S. Environmental Protection Agency, Office of Pollution Prevention and Toxics, Health and Environmental Review Division, Environmental Effects Branch, *ECOSAR: Computer Program and User's Guide for Estimating the Ecotoxicity of Industrial Chemicals Based on Structure-Activity Relationships,* EPA-748-R-93-O02 (Washington, DC: 1994).
25. Zeeman, M., "Case Study 3B: Ecological Risk Assessment of TCDD and TCDF," *Issues in Risk Assessment,* National Academy of Sciences Committee on Risk Assessment Methodology (Washington, DC: National Academy Press, 1993).
26. Zeeman, M., and Gilford, J., "Ecological Hazard Evaluation and Risk Assessment Under EPA's Toxic Substances Control Act (TSCA): An Introduction," *Environmental*

Toxicology and Risk Assessment: 1st Volume, ASTM STP 1179, W.G. Landis, et al., (eds.) (Philadelphia, PA: American Society for Testing and Materials, 1993).

27. Zeeman, M., Nabholz, J. V., and Clements, R. G., "The Development of SAR/QSAR for Use Under EPA's Toxic Substances Control Act (TSCA): An Introduction," *Environmental Toxicology and Risk Assessment: 2nd Volume,* ASTM STP 1216, J.W. Gorsuch, et al. (eds.), (Philadelphia, PA: American Society for Testing and Materials, 1993).
28. Zeeman, M., et al., "SAR/QSAR Ecological Assessment at EPA/OPPT: Ecotoxicity Screening of the TSCA Inventory," *SETAC Abstract Book for the 14th Annual Meeting at Houston, TX, Abst. P312A* (Pensacola, FL: Society of Environmental Toxicology and Chemistry, 1993).
29. Zeeman, M., "Ecotoxicity Testing and Estimation Methods Developed Under Section 5 of the Toxic Substances Control Act (TSCA)," *Fundamentals of Aquatic Toxicology: Effects, Environmental Fate, and Risk Assessment, G.* Rand (cd.) (Washington, DC: Taylor & Francis, 1995).
30. Zeeman, M., et al., "U.S. EPA Regulatory Perspectives on the Use of QSAR for New and Existing Chemical Evaluations," *SAR & QSAR in Environmental. Research,* in press.
31. Zeeman, M., "Environmental Toxicology/Ecological Effects Assessment by the U.S. EPA OPPT Under TSCA: Chemical Testing and Screening Technologies," *Proceedings of the OTA Workshop on Chemical Testing and Screening Technologies for Review of Chemicals in Commerce* (Washington, DC: Office of

8

EPA's Framework for Ecological Effects Assessment

Maurice Zeeman

ABSTRACT: The Office of Technology Assessment (OTA) is studying the implementation of the Toxic Substances Control Act (TSCA). One focus of the OTA study has been the TSCA existing chemical review program, which is administered by the U.S. Environmental Protection Agency's (EPA) Office of Pollution Prevention and Toxics (OPPT formerly the Office of Toxic Substances). The level and pace of EPA evaluation of the over 72,000 chemical substances on the TSCA Inventoty of existing chemicals lead the OTA to consider the adequacy of the testing and screening methods and technologies that are, or could be, used to assess such industrial chemicals in commerce.

One of the nine specified topics of interest to be addressed at the OTA Workshop was the testing and screening methods used by EPA (and others) to assess environmental toxicology, i.e., the testing and screening methodologies used to assess the potential ecological effects of TSCA regulated industrial chemicals. This chapter provides a review of OPPT's Environmental Effects Branch (EEB) efforts over the last 15 years in screening and assessing the potential ecological effects of industrial chemicals.

BACKGROUND

The Toxic Substances Control Act of 1976 (TSCA) provided the EPA (OPPT) with the authority to require development of adequate data for assessing the risk to human health and the natural environment from industrial chemicals identified as having risk potential. Within OPPT, the Environmental Effects Branch (EEB) has provided the expert scientific and technical evaluation of the environmental/ecological hazard of industrial chemicals, and has determined the type and adequacy of data needed to identify and assess their possible adverse effects. Over the past 15 years this group has provided significant direction to and rationale for how ecological hazard and risk assessment activities have been addressed under TSCA (32, 35, 36).

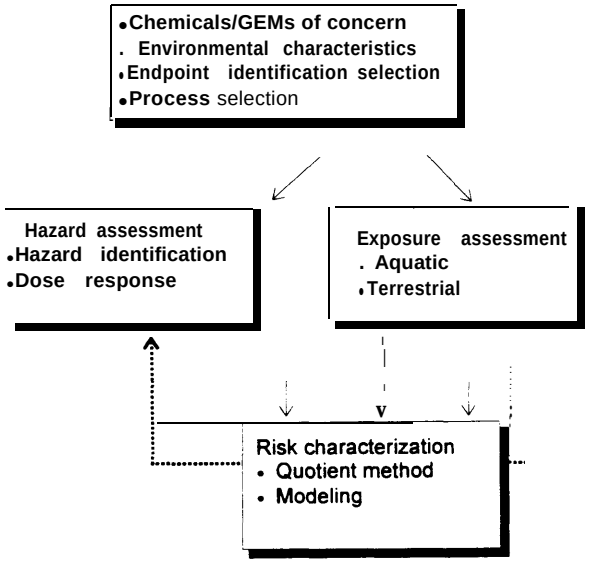
For example, from 1979 through 1994, EEB staff have been responsible for the screening and assessment of the potential ecotoxicity of over 26,000 new industrial chemicals (36). Since 1979, over 10,000 of the new chemicals that have been assessed as acceptable have been placed on the TSCA Inventory of existing chemicals because industry has commenced production and/or importation of them into the U.S.

To assure that adequate ecotoxicity data are developed to assess the possible adverse ecological effects of industrial chemicals, procedures and guidelines were established for developing data that are appropriate and adequate for assessing ecological hazard and risk. For industrial chemicals, the OPPT approach to ecological risk assessment (figure 8-1) is analogous to the risk assessment paradigm of the National Academy of Sciences (16) and is also consistent with the recently developed EPA Framework for Ecological Risk Assessment (12).

This approach ultimately required the active development of six specific areas: 1) defining appropriate ecological endpoints, 2) a tier-testing scheme for estimating impacts on these endpoints, 3) ecotoxicological testing guidelines, 4) structure-activity relationship technologies (SAR/QSAR) for estimating ecotoxicity from chemical structure, 5) hazard "assessment factors" for estimating chemical concentrations of concern, and 6) risk assessment methodologies that characterize risks by contrasting the ecotoxicity

The contents of this chapter do not necessarily reflect the views and policies of the Environmental Protection Agency.

Figure 8-1: The OPPT Ecological Risk Assessment Process



SOURCE: Zeeman, M., and Gilford, J., "Ecological Hazard Evaluation and Risk Assessment Under EPA's Toxic Substances Control Act (TSCA)," *Environmental Toxicology and Risk Assessment: 1st Volume,* W.G. Landis, et al. (ads.) (Philadelphia, PA: ASTM, 1993).

and exposure data. These several sets of developments allows OPPT to estimate the adverse effects of industrial chemicals on ecological endpoints of concern.

ECOLOGICAL ENDPOINTS

Ecological endpoints of concern are those adverse effects on the environment of sufficient importance to warrant regulatory action under TSCA (e.g., high fish toxicity). Ecological endpoints were a basic consideration in determining the kind and amount of ecotoxicological data needed to evaluate the potential hazard and risk posed by a chemical.

U.S. environmental legislation was examined to determine what ecological endpoints have been perceived by the U.S. Congress to be of sufficient importance to be protected by legislation. Resources such as wildlife, water, land, and air were to be protected from reduction, degradation, or loss in quality, quantity, or utility (5).

Also a search of the scientific literature on toxic effects of chemicals in the field identified occurrences of adverse environmental effects that resulted in some form of regulatory action. This search revealed nine cases of adverse environmental effects under field conditions, in which toxic chemicals reduced, or led to a loss of quality, quantity, or utility of the above valued resources. The adverse effects caused by the toxic chemicals were the result of: a) undesirable changes in the rates of mortality, growth, or reproduction of organisms; or b) through bioaccumulation of the chemical within a food chain to a level hazardous to other organisms in the environment (32).

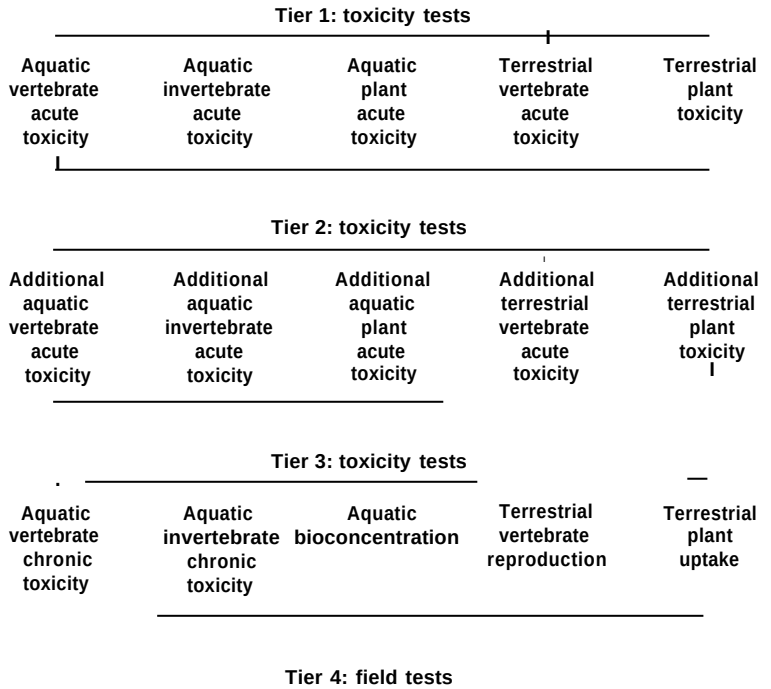
Therefore **mortality, growth and development,** and **reproduction, and their potential impacts at the population level were selected as critical features to be considered when assessing the ecological impacts posed by industrial chemi**cals. These endpoints are still being used as the primary focus in OPPT in assessing the potential for industrial chemicals to cause adverse environmental effects that may be of regulatory significance.

TIER-TESTING SCHEME AND SURROGATE SPECIES

Next to be determined was the kind and amount of testing needed to develop data adequate to measure the potential hazard of a chemical and be useful in assessing its potential risk to the environment. That effort resulted in the development of a testing scheme (figure *8-2)* that identifies the kind and amount of ecotoxicological testing required for ecological hazard and risk assessment (23, 32).

This scheme provides for sequencing (tiering) testing so that quick and inexpensive screening tests are performed first. Criteria or "triggers" for additional testing (e.g., for acute results to trigger chronic testing) and the logic for moving from one tier to another are provided and this limits testing to data essential for measuring hazard and assessing potential ecological risk (figure *8-3).*

Figure 8-2: The OPPT Ecological Testing Scheme



SOURCE: Zeeman, M., and Gilford, J., "Ecological Hazard Evaluation and Risk Assessment Under EPA's 'Toxic Substances Control Act (TSCA)," *Environmental Toxicology and Risk Assessment: 1st Volume*, W.G. Landis, et al. (eds.) (Philadelphia, PA: ASTM, 1993)

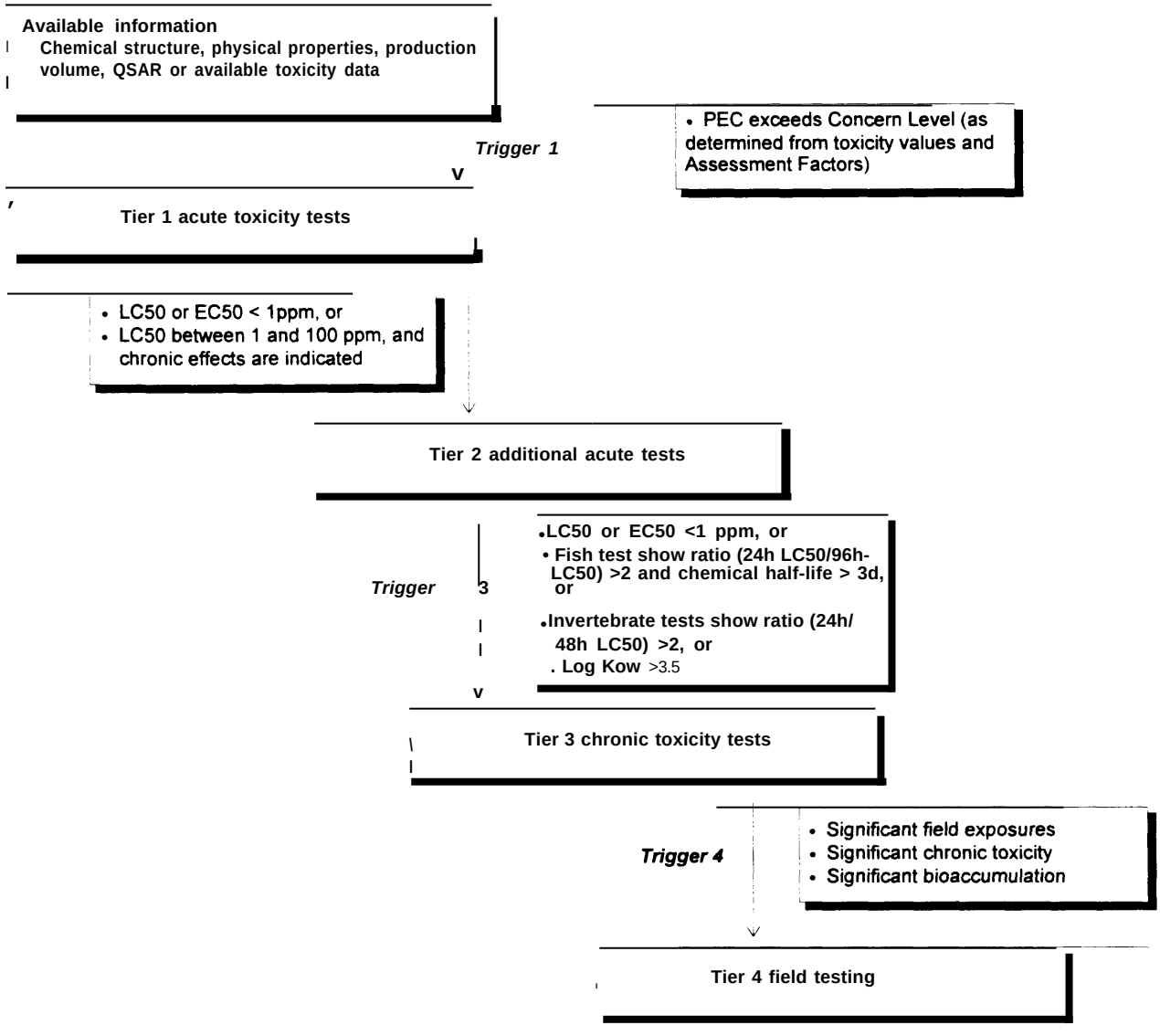
The question of using suitable surrogate test species under laboratory conditions for evaluating the ecotoxicity of chemicals in the field was also addressed. Representative organisms were selected for laboratory testing that would be acceptable as surrogate species (for example, fish suitably represent a species of aquatic vertebrates found in the water column). The number and variety of organisms that could serve as appropriate surrogates for evaluating the ecotoxicity of industrial chemicals were incorporated into the testing scheme (23).

The importance of the testing scheme is that it provides a reasonable and consistent approach to developing those test data that are needed to assess the potential environmental hazard of an industrial chemical. It also develops ecotoxicity data in a manner that does not unduly impede or create unnecessary economic barriers to technological innovation while providing adequate information to protect the environment. An additional benefit of the testing scheme is that manufacturers and testing laboratories know in advance how much testing will be needed to meet OPPT concerns, thus alleviating industry's concern about open-ended testing requirements.

TESTING GUIDELINES

After settling on appropriate testing sequences, OPPT next developed and published guidelines for performing acute and chronic ecotoxicity tests, and for determining the capacity of chemicals to bioconcentrate (26). Included in that set of ecological test guidelines are procedures for conducting acute and chronic toxicity tests using invertebrates, fish, and birds. Also included are bioconcentration tests using fish and oysters, bioassays using freshwater and marine algae, and plant toxicity tests. Tests conducted using these guidelines result in ecotoxicological data that estimates the significant endpoints of mortality or impairment (i.e., LC50, LD50, or EC50), and effects on growth and development, and/or on

Figure 8-3: Triggers for Aquatic Tier Testing System



SOURCE:Zeeman, M., and Gilford, J., "Ecological Hazard Evaluation and Risk Assessment Under EPA's Toxic Substances Control Act (TSCA): An Introduction," *Environmental Toxicology and Risk Assessment:1st Volume,* Landis, et al. (eds.) (Philadelphia, PA. ASTM, 1993).

reproduction (i.e., LOEC, NOEC, and MATC). New ecological guidelines have been added since the initial set was published (23).

These standardized test guidelines provide the means for OPPT to assure that ecotoxicological test data developed for existing chemicals (those already on the TSCA Inventory and subject to testing under TSCA Section 4) are suitable for performing adequate and reliable ecological hazard and risk assessments. These guidelines also provide a means for OPPT to assist manufacturers and testing laboratories in developing ecotoxicological test data suitable for evaluating the hazard and risk of the thousands of new chemicals subject to OPPT review under TSCA Section 5. The above testing scheme and guidelines are used routinely by industry and testing laboratories in developing ecotoxicological data for OPPT.

These test guidelines provide an additional benefit. Data based upon the use of standard test guidelines are of great value to OPPT for comparative purposes, in developing new structure-activity relationships (SAR), and for developing valid data suitable for inclusion in ecotoxicological data bases that can also be used to help develop quantitative SAR (QSAR). Reliable test data developed through these TSCA chemical testing requirements have provided OPPT with valuable information on chemical analogs and on chemical structure-activity relationships. These data have proven essential in evaluating the potential ecotoxicity of similar industrial chemicals for which test data are not available and in the ongoing validation efforts of the OPPT (Q)SAR technologies developed byEEB(15).

SAR/QSAR FOR ASSESSING ECOTOXICITY

The development and use of structure-activity relationships (SAR) and quantitative SAR (QSAR) to assess ecotoxicity has been an essential and active area of interest in OPPT/EEB for over a decade (1, 7, 8, 11, 13, 14, 35). This SAR/QSAR [or (Q)SAR] development became essential because estimations of ecotoxicity have to be provided in a very short time-frame for the risk assessments required for the thousands of new industrial chemicals that are submitted by industry to OPPT for evaluation every year.

Over the last decade, the new chemicals program has required the rapid assessment of about 2,300 chemicals/yr – almost 50/week, typically with these numerous ecotoxicity estimates needing to be available for preliminary risk assessment purposes within 2-3 weeks after industry submits the chemical for evaluation by OPPT. As up-front testing is not required for these submissions, the vast majority of OPPT new chemical notices (ea. 95°/0) contain no ecotoxicity data and, therefore, our (Q)SAR methodologies have been developed to fill these data gaps (34, 36).

SAR estimations can vary from simple similarities, such as using test data available for a similar chemical grouping or analogs, to being able to provide quantitative estimates of ecotoxicity (QSAR) because an empirical mathematical relationship has been established for a chemical grouping/class to which the new chemicals also belong. The 1988 version of the OPPT (Q)SAR Manual (6) has been updated and currently contains about 120 OPPT SAR/QSARs available for assessing the ecotoxicity of about 45 classes of chemicals (29). A computer program was also developed for the OPPT (Q)SAR Manual and was recently released as a PC Version, called ECOSAR, which is publically available (30).

The OPPT aquatic toxicity (Q)SARs used for estimating the acute toxicity of industrial chemicals to fish, daphnia, and algae have generally been proven to be quite reliable. The validation of these OPPT SAR/QSAR relationships is an ongoing effort (15, 34, 36). In addition, a joint EPA/European Union evaluation of the accuracy of the OPPT SARs was undertaken from 1991-93 and it was termed the “Structure Activity Relationship/Minimum Premarketing Dataset” (SAR/MPD) study (17, 23). For the EPA ecotoxicity (Q)SAR methodology, the European Union experts concluded that these OPPT QSAR methods “performed extremely well in predicting acute toxicity to fish and daphnia” (23, 36).

The structure-activity relationship (SAR) and quantitative SAR (QSAR) technologies that have been actively developed by EEB for the formidable new chemical endeavor were also recently applied to the ecotoxicity screening and assessment of over 8,000 discrete organic industrial chemicals on the TSCA Inventory of existing chemical substances. This technology was useful in assessing the hazard distributions for different chemical classes and the proportions of such chemicals displaying high acute or chronic ecotoxicity (8, 9,34,35, 36).

HAZARD “ASSESSMENT FACTORS”

As so little up-front ecotoxicity test data was provided to assess new or existing chemicals, EEB dealt with the several levels of uncertainty created due to this lack of data by developing “assessment factors” (25). Pragmatically, these

Table 8-1: OPPT Assessment Factors Used in Setting "Concern Levels"

Available data on chemical or analogue	Assessment factor
Limited (e.g., only One Acute LC50 via SAR/QSAR)	1000
Base Set Acute Toxicity (e.g., Fish and Daphnid LC50's, and Algal EC50)	100
Chronic Toxicity MATC's	10
Field Test Data for Chemical	1

SOURCE: Zeeman, M., "Ecotoxicity Testing and Estimation Methods Developed Under Section 5 of the Toxic Substances Control Act (TSCA)," *Fundamentals of Aquatic Toxicology: Effects, Environmental Fate, and Risk Assessment,* Chapter 23, G. Rand (cd.) (Washington, DC: Taylor & Francis, 1995).

assessment factors are used in a fashion somewhat akin to uncertainty factors to provide a consistent regulatory basis for assessing the potential for ecological risks.

These factors vary by three orders of magnitude (see table 8-1) in an attempt to account for uncertainties such as a) estimating chronic toxicity from acute toxicity, b) accounting for species-to-species differences, and c) extrapolating from laboratory toxicity tests to field toxic effect levels. The hazard estimate, even a (Q)SAR estimate, is divided by an appropriate assessment factor, and this results in predictions of concentrations of concern (concern level) in the environment.

Environmental exposures (often PECs, predicted environmental concentrations) below this concern level are not presumed to be safe or without risk. However, for practical purposes, if this level is not exceeded, it has typically been assumed by OPPT that the likelihood of a significant environmental risk is probably too low to warrant taking any regulatory action (13).

For regulatory purposes, the concern level is the environmental concentration above which risk in the aquatic environment could be inferred. When risk to organisms in the environment is estimated to be likely (due to sufficient potential exposures), that forms one possible basis for requesting the development of ecotoxicity test data to further refine the potential hazards of the chemical.

Although some would argue that these concern level estimates may be too conservative because of what may seem large assessment factors, in reality there are additional factors that need to be considered. First, the assessment factor of 1,000 is seldom used because we can often estimate through (Q)SARs the acute toxicity to fish, daphnia, and algae, and/or chronic toxicity to one or more of these aquatic species. Therefore assessment factors of 10 or 100 are more typical.

Next, the uncertainty factor approach used for the estimation of risk to organisms in the environment would seem to be substantially less rigorous than that used for estimating risk to humans. For example, results of toxicity testing in a few rodent and other mammalian species are all typically used, along with appropriate uncertainty factors, to estimate the chemical risks to only one species, humans. The extrapolations that are used here for environmental species should often seem more uncertain. How many would argue that a short-term laboratory toxicity test result for only one species of fish is suitable enough to extrapolate such limited data as being representative of, let alone being protective for, the other 20,000 species of fish found in the real world?

Other more refined ecological protection methods, based on fairly extensive ecotoxicity test data and on sets of complex statistical methodologies that may protect 95% of the species are being developed by the Dutch and have been proposed in the ecological assessment schemes of the OECD (18). Independent evaluations of the simple EPA/OPPT assessment factor approach and these complex statistical schemes seem to find that often there is not that much difference in predicting the respective levels of concern in the environment (4, 10, 35).

ECOLOGICAL RISK ASSESSMENT

Unlike human health risk assessment, ecological risk assessment must consider adverse effects of chemicals on many species, not just one. As is seen in the previous discussion on assessment

factors, one way that this has been done is to use the hazard estimates for several species and to apply such "uncertainty" factors to estimate the environmental concentrations of concern.

A slightly more refined method also used by OPPT for its chemical assessments is to contrast these hazard estimates with predictions of exposure concentrations (PECs) expected in the environment. This is called the quotient method and it is very widely used as a measure of potential ecological risk. As predicted exposures approach the lowest hazard estimate (for toxicity, reproductive effect, etc.) more of a risk for such impacts in the environment is inferred (21, 28, 32).

The quotient method is most typically used in the ecological assessment of existing chemicals. An example would be the assessment of the potential impacts on terrestrial organisms in the environment from the dioxins and furans (TCDD/TCDF) found as contaminants in the sludges of the paper and pulp industries that are applied to lands and forests as soil conditioners (12, 19,20, 31). The quotient method allows for a simple comparison of the best estimates of toxic thresholds and the no-observed-adverse-effect levels (NOAELs) for terrestrial birds and mammals with the TCDD and TCDF concentrations likely to be in their food (soil organisms) that had become contaminated from the soils amended with such sludges.

Through the expanded development of population and ecosystem models, the typical ecotoxicity test data on mortality, growth, and reproduction may be extended to more adequately deal with population- and ecosystem-level effects (24). Such models are starting to be used to augment the typical existing chemical assessments, such as in the assessment of the potential adverse aquatic effects of the chlorinated paraffins, a widely used industrial chemical (3, 21). The continued development of pragmatic and user-friendly population and ecosystem models is essential for these powerful methods to be useful to regulators needing quick and simple responses to the difficult issues posed in the ecological risk assessment of industrial chemicals.

The goal of an OPPT ecological risk assessment is to be as realistic as is reasonable with the data available. When data are scarce, as typically occurs in new chemical assessment, the estimates made of hazard, exposure, and risk must be viewed as being somewhat preliminary. Ecological data provided for existing chemical assessments may be almost as scanty as for new chemi--cals, however, it is more common that additional ecotoxicological data will be provided for these major chemicals in commerce. Hopefully this additional data allows for more reasoned decisions on hazard, exposure, and risk to be made.

CONCLUSION

The many years of experience by OPPT/EEB in screening and assessing the ecotoxicity and risks of the thousands of new industrial chemicals submitted for evaluation each year has resulted in procedures and technologies, i.e., (Q)SARs and concern concentrations, that are extremely versatile and efficient in assessing chemicals. These efficient ecotoxicity assessment procedures and technologies are now also being applied to the discrete organic substances in the existing chemical arena, especially to those existing chemicals for which little or no reliable ecotoxicity data is readily available.

The use of the chemical class (Q)SAR methodology by OPPT should continue to expand. As new test data for terrestrial organisms on specific chemical classes have become available, OPPT/EEB has already expanded the use of (Q)SAR into that environment (e.g., earthworm QSAR for neutral organics). Also, as additional targeted ecotoxicity test data are provided, there are many areas into which this screening and assessment technology could be expanded further, e.g., the development of (Q)SARs for sediment-dwelling organisms and for avian species should be of high priority.

REFERENCES

1. Auer, C. M., Nabholz, J. V., and Baetke, K.P., "Mode of Action and the Assessment

of Chemical Hazards in the Presence of Limited Data: Use of Structure-Activity Relationships (SAR) Under TSCA, Section 5," *Environmental Health Perspectives* 87:183-197, *1990.*

2. Auer, C. M., et al., "SAR - The U.S. Regulatory Perspective," *SAR & QSAR in Environmental. Research.* 2(1-2):29-38, *1994.*
3. Bartell, S.M., "Ecosystem Context for Estimating Stress-Induced Reductions in Fish Populations," *American Fisheries Society Symposium* 8:167-182, *1990.*
4. Calabrese, E.J., and Baldwin, L.A., *Performing Ecological Risk Assessments* (Boca Raton, FL: Lewis Publishers, 1993).
5. Clements, R. G., *Environmental Effects of Regulatory Concern Under TSCA - A Position Paper,* (Washington, DC: U.S. Environmental Protection Agency, Office of Toxic Substances, Environmental Effects Branch, 1983).
6. Clements, R.G. et al. (eds.), *Estimating Toxicity of Industrial Chemicals to Aquatic Organisms Using Structure Activity Relationships,* Environmental Effects Branch, Health & Environmental Review Division, U.S. Environmental Protection Agency, EPA-560-6-88-001 (Washington, DC: 1988).
7. Clements, R. G., et al., "The Use and Application of QSARs in the Office of Toxic Substances for Ecological Hazard Assessment of New Chemicals," *Environmental Toxicology and Risk Assessment: 1st Volume,* W.G. Landis, J.S. Hughes, and M.A. Lewis (eds.) (Philadelphia, PA: American Society for Testing and Materials, 1993).
8. Clements, R. G., et al., "The Use of Quantitative Structure-Activity Relationships (QSARs) as Screening Tools in Environmental Assessment," *Environmental Toxicology and Risk Assessment: 2nd Volume,* J.W. Gorsuch et al. (eds.) (Philadelphia, PA: American Society for Testing and Materials, 1993).
9. Clements, R. G., et al., "The Application of Structure-Activity Relationships (SAR's) in the Aquatic Hazard Evaluation of Discrete Organic Chemicals," *SAR & QSAR Environmental Research* 3(3):In press, 1995.
10. Forbes, V. E., and Forbes, T. L., *Ecotoxicology in Theory and Practice (New York, NY:* Chapman and Hall, 1994).
11. Lipnick, R. L., et al., "Comparison of QSAR Predictions with Fish Toxicity Screening Data for 110 Phenols," *Aquatic Toxicity and Hazard Assessment: 8th Symposium,* R.C. Bahner and D.J Hansen (eds.) (Philadelphia, PA: American Society for Testing and Materials, 1985).
12. Meyn, O., et al., "Landspreading of Sludge from Pulp and Paper Mills: Potential Risks from Dioxins and Furans to Terrestrial Wildlife," *SETAC Abstract Book for the 15th Annual Meeting at Denver, CO, Abst. TD25,* (Pensacola, FL: Society of Environmental Toxicology and Chemistry, 1994).
13. Nabholz, J.V., "Environmental Hazard and Risk Assessment Under the United States Toxic Substances Control Act," *Sci. Total Environment* 109/110:649-665, *1991.*
14. Nabholz, J.V., Miller, P., and Zeeman, M., "Environmental Risk Assessment of New Chemicals Under the Toxic Substances Control Act TSCA Section Five," *Environmental Toxicology and Risk Assessment: 1st Volume,* W.G. Landis, J.S. Hughes, and M.A. Lewis (eds.) (Philadelphia, PA: American Society for Testing and Materials, 1993).
15. Nabholz, J. V., et al., "Validation of Structure Activity Relationships Used by the USEPA's Office of Pollution Prevention and Toxics for the Environmental Hazard Assessment of Industrial Chemicals, " *Environmental Toxicology and Risk Assessment.. 2nd Volume,* J.W. Gorsuch et al. (eds.) (Philadelphia, PA: American Society for Testing and Materials, 1993).
16. National Academy of Sciences, National Research Council, *Risk Assessment in the Federal Government: Managing the Process* (Washington, DC: National Academy Press, 1983).

17. Organization for Economic Cooperation and Development, *US EPA/EC Joint Project on the Evaluation of (Quantitative) Structure Activity Relationships (QSARS),* OECD Environment Monographs No. 88 (Paris, France: 1994).
18. Organization for Economic Cooperation and Development, *Guidance Document for Aquatic Effects Assessment,* OECD Environment Monographs No. 92 (Paris, France: OECD, 1995).
19. Rabert, W., Morcock, R., and Zeeman, M., "U.S. EPA Dioxin in Paper Project: Risk Assessments for Aquatic and Terrestrial Environments," *Dioxins '90: 3rd Volume, O.* Hutzinger and H. Fiedler (eds.) (Bayreuth, Germany: Eco-Inform Press, 1990).
20. Rabert, W., and Zeeman, M., "Dioxins/Furans: U.S. EPA Ecological Risk Assessment for Land Application and Disposal Methods for Paper Pulp Sludge," *Chemosphere* 25:1499-1504, *1992.*
21. Rodier, D.J., and Mauriello, D. A., "The Quotient Method of Ecological Risk Assessment and Modeling Under TSCA: A Review," *Environmental Toxicology and Risk Assessment: 1st Volume,* W.G. Landis, J.S. Hughes, and M.A. Lewis (eds.) (Philadelphia, PA: American Society for Testing and Materials, 1993).
22. Rodier, D.J., and Zeeman, M., "Ecological Risk Assessment," *Basic Environmental Toxicology,* L.G. Cockerham and B.S. Shane (eds.) (Boca Raton, FL: CRC Press, Inc., 1994).
23. Smrchek, J., et al., "Assessing Ecological Hazard Under TSCA: Methods and Evaluation of Data," *Environmental Toxicology and Risk Assessment: 1st Volume,* W.G. Landis, J.S. Hughes, and M.A. Lewis (eds.) (Philadelphia, PA: American Society for Testing and Materials, 1993).
24. Suter, G.W. (cd.), *Ecological Risk Assessment* (Boca Raton, FL: Lewis Publishers, 1993).
25. Environmental Protection Agency, Office of Toxic Substances, Environmental Effects Branch, *Estimating "Concern Levels" for Concentrations of Chemical Substances in the Environment* (Washington, DC: 1984).
26. U.S. Environmental Protection Agency, "Toxic Substances Control Act Test Guidelines, Part 797," final rule, *Federal Register* 50:39252-39516, *1985.*
27. U.S. Environmental Protection Agency, Office of Research and Development, *Framework for Ecological Risk Assessment, EPA/630/R-92/001* (Washington, DC: 1992).
28. U.S. Environmental Protection Agency, Office of Pollution Prevention & Toxics, Chemical Control Division, US *EPA/EC Joint Project on the Evaluation of (Quantitative) Structure Activity Relationships,* Final Report (Washington, DC: 1993).
29. U.S. Environmental Protection Agency, Office of Pollution Prevention and Toxics, Health and Environmental Review Division, Environmental Effects Branch, *Estimating Toxicity of Industrial Chemicals to Aquatic Organisms Using Structure Activity Relationships,* EPA-748-R-93-001 (Washington, DC: 1994).
30. U.S. Environmental Protection Agency, Office of Pollution Prevention and Toxics, Health and Environmental Review Division, Environmental Effects Branch, *ECOSAR: Computer Program and User's Guide for Estimating the Ecotoxicity of Industrial Chemicals Based on Structure Activity Relationships,* EPA-748-R-93-002 (Washington, DC: 1994).
31. Zeeman, M., "Case Study 3B: Ecological Risk Assessment of TCDD and TCDF," *Issues in Risk Assessment* (Washington, DC: National Academy Press, 1993).
32. Zeeman, M., and Gilford, J., "Ecological Hazard Evaluation and Risk Assessment Under EPA's Toxic Substances Control Act (TSCA): An Introduction," *Environmental Toxicology and Risk Assessment: 1st Volume,* W.G. Landis, J.S. Hughes, and M.A.

Lewis (eds.) (Philadelphia, PA: American Society for Testing and Materials, 1993).

33. Zeeman, M., Nabholz, J. V., and Clements, R. G., "The Development of SAR/QSAR for Use Under EPA's Toxic Substances Control Act (TSCA): An Introduction," *Environmental Toxicology and Risk Assessment: 2nd Volume,* J.W. Gorsuch et al. (eds.) (Philadelphia, PA: American Society for Testing and Materials, 1993).
34. Zeeman, M., et al., "SAR/QSAR Ecological Assessment at EPA/OPPT: Ecotoxicity Screening of the TSCA Inventory," *SETAC Abstract Book for the 14th Annual Meeting at Houston, TX, Abst. P312A* (Pensacola, FL: Society of Environmental Toxicology and Chemistry, 1993).
35. Zeeman, M., "Ecotoxicity Testing and Estimation Methods Developed Under Section 5 of the Toxic Substances Control Act (TSCA)," *Fundamentals of Aquatic Toxicology: Effects, Environmental Fate, and Risk Assessment, G.* Rand (cd.) (Washington, DC: Taylor & Francis, 1995).
36. Zeeman, M., et al., "U.S. EPA Regulatory Perspectives on the Use of QSAR for New and Existing Chemical Evaluations," *SAR & QSAR in Environmental Research 3(3):* (In Press), 1995.

9

Molecular Assays for Environmental Endpoints

Spencer B. Farr

Without precise estimates of toxicity, any method will lack the accuracy necessary to identify the cost/benefit ratio of proposed remedial actions intended to identify true chemicals of concern and reduce environmental hazards. This lack of precision results in agreement between environmentalists, government agencies, and industry that we are currently wasting large sums of taxpayer money on present methods of environmental hazard analyses. I believe we can do the job more accurately and less expensively.

A great deal of energy is being devoted to identification and remediation of sites containing potentially hazardous materials. Environmental engineers are developing remarkable technologies for finding hidden waste sites, including the use of LANSAT satellites and ground penetrating radar probes. Similarly, there are numerous technologies being developed for disposal, incineration, or encapsulation of that material once it has been found. Unfortunately, there has been little progress in developing rapid testing procedures to determine if the material is toxic, and therefore, in need of remedial attention in the first place. Even a rudimentary economic analysis shows that the cost of remediating every site known to contain environmentally hazardous material is astronomical, and in fact, prohibitive. There is no economic or environmental justification for remedial actions at sites that pose no real biological threat. On the other hand, we must find ways of prioritizing which sites shall receive attention.

One of the most important factors to be considered in such prioritization is whether or not a site actually contains bioactive/toxic materials which pose risks to human health and the environment. Few methods are currently available that can monitor the degree of toxicity, or determine the mechanisms whereby mixtures of chemicals may be toxic beyond that of the naturally occurring bioactive/toxic materials. Predictions of toxicity based on a subset of identified chemicals occurring in a sample fall short of the goal of protecting the environment and saving money. Determination of the true toxicity can best be achieved by monitoring molecular responses to environmental mixtures or pure compounds in living organisms.

The method currently employed to assess human health and environmental risks associated with contaminants usually relies on physical and chemical analysis of soil, water, and air samples. Samples are analyzed for the presence of approximately 400 chemicals that have been declared "toxic" based upon toxicity tests in whole animals using high concentrations of the pure form of a compound. If the sample analysis indicates the presence of compounds above a certain threshold limit, the site is then considered to pose a human health or environmental hazard. The cost of remediation usually depends upon the concentration of the contaminants found at the site, the contaminated area, and relative toxicity of those contaminants as determined by testing pure chemicals in whole animals. Therefore, several critical sources of error can lead to high uncertainty in predictions based on chemical analyses which include: (1) the presence of chemicals that were not identified, (2) the presence of chemicals that lack toxicity data, (3) the effects of synergy or antagonism in mixtures of contaminants, (4) bioavailability, and (5) the effectiveness

of treatment methods which may generate toxic substances as a result of remediation.

While this analytical chemistry approach to environmental monitoring was the best available technology twenty years ago, there are five shortcomings. These shortcomings are briefly described below:

1) There are approximately 120,000 chemicals manufactured world-wide. The toxic potential of which is largely unknown. If a given environmental sample does not contain any of the 400-plus toxins on the Priority List of Hazardous Substances published by the U.S. Environmental Protection Agency (EPA) above the allowable level, the sample receives a clean bill of health. It is impossible to monitor all 120,000 compounds by current physical chemical analysis and thus the search is conducted for only a small percentage of known toxin compounds. Such analysis may vastly underestimate the toxic potential of a site because it only looks for only 0.003 of the potential "man-made" toxins. If cost were not an issue, physical-chemical analysis could still only detect and identify approximately 5% of the known man-made compounds.
2) The physical-chemical approach to hazard analysis ignores most naturally occurring chemical hazards such as heavy metals and organic toxins. Therefore, it may underestimate toxic potential of any particular site. In addition, the turn-around time between collecting samples and receiving analytical data may be several months.
3) The physical-chemical approach to hazard analysis cannot detect novel compounds formed by the interaction of manmade compounds with each other or with naturally occurring compounds. This is likely to result in novel compounds that are impossible, not just costly, in determining their toxic potential.
4) The means by which environmental "toxins" have been designated as such is questionable. Pure compounds suspected of being toxic are tested in a limited number of whole animals at high concentrations. For economic reasons, classical toxicologists have had to assume that high concentrations in a small cohort of test animals give the same results as low concentrations in a large cohort! Furthermore, they generally extrapolate to expected responses at low doses using a linear dose response curve, when in fact, most compounds show a threshold level below which there is no detectable effect. Thus, if the compound is found to be toxic in test animals, then with the appropriate safety factor, it is assumed to be toxic in humans. This represents a vast assumption.
5) Finally, previous animal studies as well as the few animal studies used today in environmental analysis focus almost exclusively on cancer potential while ignoring most of the other noncarcinogenic toxic endpoints.

There are a number of ways in which the tools of modem molecular biology can aid in the assessment of risk posed by chemicals in pure form *and* in mixtures. Physical-chemical analysis of site samples is useful for detecting the presence of only a limited number of known toxic agents. Because such an analysis overlooks so many potential toxins, it may underestimate the true toxicity of a site. Conversely, because these analyses base toxicity analysis on whole animal exposures to pure compounds at extremely high doses, it may also dramatically overestimate the health hazards of a site. As you can see, the room for error using current techniques and models is so great that its value is highly questionable. We must identify new methods to correctly identify those sites that pose a legitimate toxic threat to humans versus those that contain biologically insignificant levels of compounds found to be toxic only in test animals at high doses. At the risk of being redundant, there is simply no justification for remedial actions at sites that pose no biological threat.

How do we improve our ability to accurately estimate the health hazard potential of an environmental site? First, we do not attempt to ascertain toxicity by physical-chemical means alone. Rather, we measure more direct end points, namely, the toxic effects on living organisms. If our end goal is to determine the health effect that a certain environment poses on living organisms,

the most direct and accurate method is to expose living organisms to that environment (or a sample thereof), and ask if there are observable toxic manifestations. Unfortunately, while whole animal assays would certainly improve our ability to predict human health impacts over physical-chemical analyses, whole animal (mammal) tests are extremely expensive and time-consuming, not to mention politically unpopular, and ethically suspect. Furthermore, they do not generally provide mechanistic information about the biochemical event that causes harm to the cell.

Leading molecular toxicologists have developed a battery of *in vitro* and transgenic assays for the rapid, inexpensive, and technically-simple collection of toxicological information. This technology utilizes a panel of bacterial, yeast, insect, and mammalian (including human) cell assays. Unlike existing *in vitro* toxicity assays, this panel of assays provides results which are integrated, and thus allow a thorough and mutually confirming analysis of relative toxicity. In addition, these assays are *directly relevant to humans* because the tests are performed on human cells.

One example of how we are employing the power of modem molecular biology toward assessing environmental toxicity is as follows. In order to rapidly assess the bioactivity/toxicity to humans of a complex mixture, realistic of most environmental samples, we have taken advantage of the fact that individual human cells respond to toxic stimuli *in vitro* in most cases identically to the way they respond *in vivo*. Part of that response is an induction in the transcriptional activity of specific genes with well defined functions. The genes that can be directly monitored encode proteins that can detoxify the toxic chemical, repair the damage that the toxic chemical causes to cell components (a toxin is toxic because it damages one or more cell components), or reduce the bioavailability by binding or excretion. The stress/damage genes that are induced are highly specific for the type of stress/damage caused by a given class of environmental toxins, and any given class of toxins induces a "signature" subset of stress genes.

Several published papers indicate that this assay system can provide the most accurate assessment of both the degree and mechanism of toxicity available in an *in vitro* assay. The advantages of such an assay are as follows: 1) the cost of this assay is in the range of hundreds of dollars versus tens of thousands of dollars using traditional whole animals, 2) the time required to run this assay is hours versus months for traditional assays, 3) this assay provides useful data about the level and mechanisms of toxicity; information that is rarely provided by whole animal tests, and 4) this assay dramatically decreases the reliance on whole animal tests. Thus, these assays represent the broadest range of molecular biological approaches to human toxicology available.

Another example of a molecular toxicity assay is the utilization of transgenic nematodes for the detection of mutagenic potential contained in soil samples. Analysis of mutagenic potential is facilitated by the insertion of a mutation-reporter gene inserted into the genome of every cell in the nematode in the same location, as well as a facile means of monitoring mutations in that gene. These are but a few examples of several molecular toxicology assay systems available today that can dramatically reduce the cost and time of analysis while simultaneously increasing the quality and value of information available for risk assessment.

Pertaining to environmental toxicological endpoints, a review of current screening technologies relevant to the needs of the TSCA existing chemical program on the use of a battery of molecular toxicology methods as a prescreening technique to complement and guide chemical and whole animal tests.

BEST TESTS TO IDENTIFY CHEMICALS OF CONCERN

Along with the limitations of chemical and whole animal testing discussed in the previous section, no test will provide accurate results in all cases, thereby supporting the validity of using a battery of assays for scientifically sound weight-of-evidence predictions.

- *Chemical tests* - most widely used, Confirms the presence of potential toxicants but may identify a subset of all substances in a sample and misrepresent the bioavailable components. Important to identify sources of toxicity predicted with animal and molecular toxicology tests so that remedial treatment technologies can be tested
- *Animal tests* - widely used. Useful to determine lethal effects and some gross sublethal effects such as weight loss, fertility, and behavioral changes.
- *Molecular toxicology tests* - Major technological advancement of recent development. Useful to quickly determine lethal effects and sublethal mechanisms that may explain why substance is toxic. Have also been used to determine if remedial treatment technologies are cost effective and fine tune environmental studies so that scientists can predict where to look for effects from contaminants.

WHAT FASTER AND CHEAPER SCREENS ARE AVAILABLE

Two types of molecular assays are currently available. Immunoassay detect the presence of specific chemicals or specific effects from chemicals; and *in vitro* tests using genetically engineered organisms.

The tests with the highest cost/benefit ratio are stress gene assays that can quickly monitor many of the known primary and secondary mechanisms produced by toxins in a single test. Weight-of-evidence data from multiple species reduce the probability of false negatives and false positives. There, assays also provide information about many non-genotoxic endpoints.

A list of commercially available immunoassay include:

- ENSYS, INC.: produces immunoassay for the rapid detection of certain classes of chemicals such as PAHs.

Commercially available *in vitro* assays using genetically engineered cells and organisms include:

- *Pro-Tox Bacterial Stress Gene Assay:* prediets 16 mechanisms of toxicity and lethality.
- *CAT-Tox Mammalian Stress Gene Assay:* predicts 14 mechanisms of toxicity and lethality in human HepG2 liver cell lines.
- *Ames II Genotoxicity Assay:* predicts 8 specific genotoxic point mutation and frame shift types of DNA damage, as well as lethality in bacterial cell lines with improvements over the widely used Ames Assay.
- *Yeast DEL Genotoxicity Assay:* predicts DNA damages in a eukaryotic cell line that responds by a global recombination repair pathway more similar to that found in mammals than may be predicted by the Ames Assay.
- *E. coli TRP Assay:* predicts several types of genotoxic damage in a bacteria that has evolved closely with humans.
- *Mutametrix Nematode Assay:* predicts mutagenicity in a transgenic nematode for determination of mutagenicity.

TRADEOFFS: CONFIDENCE, VALIDATION, AND REPRODUCIBILITY

Currently screening of samples can be achieved with a battery of molecular toxicology assays (using a minimum of two species) with capabilities to monitor both cytotoxic and genotoxic effects at the subcellular level. The assays evaluated were selected based on requirements that they can provide high precision at low cost, provide rapid turn-around or can be adaptable to field use, and predictive of potential mechanisms of toxicity. The rationale is that if the presence of bioactive/toxic materials cannot be demonstrated on the total sample then the hazard is minimal, whereas indications with multiple species that a sample can produce DNA damage or subcellular damage is a “red flag” warning that addition testing may be required. This approach allows fine-tuned site evaluations instead of the current “shotgun” technique that is costly, time consuming, and inaccurate.

A major advantage of molecular assays is the ability to control variables. Therefore, reproducibility of results is improved. The interpretation of results comes from comparisons to data on test chemicals that have known mechanisms of toxicity, as well as a thorough understanding of the causal relationship between the damage and the endpoint measured. Comparisons of gene inductions from chemicals used to validate the assays allow prediction of mechanisms of toxicity with mixtures of chemicals – the most difficult class of samples to evaluate. Using good laboratory procedures in the conduct of molecular assays provides high agreement in data. Currently, interlaboratory validation studies are ongoing.

RECEPTOR-BASED, MECHANISM-BASED, AND SAR APPROACHES

Taken together, the three types of tests – chemical, whole animal, and molecular toxicology give scientifically strong, mutually confirming, weight-of-evidence evaluations. However, the cost of such an extensive test sequence is not justified at most contamination sites nor to evaluate most chemicals unless the risk to human health and the environment could be substantial and exposure widespread. The best application of these approaches is in a tiered battery starting with simple biological endpoints. One factor that may be overlooked in the evaluation of potential hazards from chemical tests is the time delays in reaching a decision on remedial actions. Since chemical tests are the most indirect approach to determine toxicity then it is logical to conclude that data from ambiguous chemical analyses may be the least precise of the three types of tests. As the complexity of the contaminant mixture increases the accuracy of using chemical tests to predict actual toxicity decreases. The *use* of chemical analyses in a SAR, a common practice, may benefit most from the additional use of molecular toxicology data.

The use of animal testing is the only method to detect systemic and chronic effects in multiple species. Many of these effects can be accurately predicted with chemical or molecular assays. Additionally, the expense, ethical concerns, and time interval to conduct many animal tests limits the utility of these options, and requires that they be used only if chemical and molecular toxicity assays fail to produce clear, unambiguous results. Much of the current animal testing may be replaced with molecular assays in the future.

The use of molecular testing is gaining widespread support due to qualities such as high precision, low cost, rapid analyses that indicate why substances may be expected to cause adverse effects. Classical dose-response curves using multiple species can be generated simultaneously for a chemical or mixture of chemicals in several hours to one or two days. The use of up to dozens of different genetically engineered cells, each monitoring the activity of a different gene with characterized functions, in a single assay is a powerful tool not previously available for prescreening chemicals to predict the probability of adverse effects.

INTEGRATION INTO AN OVERALL SCREENING AND TEST STRATEGY

Much is already known of the tests available using chemical and whole animal tests. Less well known in the environmental community are the commercially available tests from several suppliers of molecular assays currently used by government agencies and laboratories, chemical and pharmaceutical industries, and research institutions to rapidly and quickly screen substances for toxicity. Molecular tests can be readily integrated with current test strategies to provide first-tier evidence indicating the potential of toxicity. These assays should be used as a prescreen prior to expensive animal tests or chemical tests that may indicate only a subset of contaminants in a mixture. The common endpoints in whole animal and molecular assays based in genetically engineered organisms are the lethal concentrations. When lethal concentrations indicate similar sensitivity between the whole animal test and molecular toxicology organisms to the test substance then the probability of predicting applicable mechanisms of toxicity may be improved. The

precision of molecular assays and results from validation tests may allow calculation of confidence limits with data. Therefore, the uncertainty of evaluations using molecular, whole animal, and chemical test is likely to within limits allowing meaningful predictions of risk to human health and the environment.

NEW DEVELOPMENTS

It would not seem to be an overstatement to predict that major advancements in molecular toxicology will overshadow state-of-the-art advances in animal testing and chemical testing aimed at predicting toxicity. Since organisms are more accurate predictors of toxicity than indirect chemical tests we expect use of transgenic animal models to be the greatest contribution to the field of toxicology over the next decade.

10

Exposure Biomarkers

Larry L. Needham, James Bond, and Steve Tannenbaum

ABSTRACT: This workshop was designed primarily to examine available technologies for screening all or a selected portion of the approximately 72,000 chemicals that are included in the Toxic Substances Control Act (TSCA) for health effects in humans and, to a lesser degree, their effects on the ecological system. In terms of risk assessment, such screening procedures would yield data for use in hazard identification. Obviously, in order to screen this immense number of chemicals, the chemicals must be prioritized. One such focus could be on the 14,000 non polymeric TSCA Inventory chemicals produced in amounts greater than 10,000 pounds per year. Nonetheless, screening all of these 14,000 chemicals for various health endpoints still requires that they be further prioritized. No doubt, quantitative structure activity relationships will be used to set priorities. However, we submit that priority setting could also be based, at least in part, on another aspect of risk assessment - human exposure assessment, for without human exposure, there are ***no*** *adverse health effects, and no need would exist for further risk characterization. Human exposure has been assessed by a variety of means. We believe that the most accurate means of assessing human exposure is the measurement of biomarkers of exposure in human specimens.*

In this presentation, we give examples of how using such biomarkers provided qualitative and quantitative exposure information that proved useful in conducting epidemiological studies. We also present how reference range levels of exposure biomarkers in humans (as acquired by biomonitoring programs) have been extremely beneficial in conducting exposure assessments and how expansion of such programs would directly benefit TSCA. Programs, such as the National Health and Nutrition Examination Survey (NHANES) and the National Human Exposure Assessment Survey (NHEXAS), are available to collect and bank the needed specimens. Analytical methods would then be used in these programs to determine whether, and to what extent, humans were being exposed to particular TSCA-related substances. If so, more extensive "effect screening" methodologies would be used for these substances; if no, or little exposure, was detected, these substances may be given a low priority for "effect screening", and further risk characterization.

Humans are exposed daily to a variety of chemicals that are present in the environment as pollutants or that are in commercial products. For this presentation, we shall assume that these chemicals are included in the Toxic Substances Control Act (TSCA) list of some 72,000 chemicals. Humans are exposed when they come into contact for an interval of time with such chemicals in an environmental medium–soil, water, air, or food or in another medium, such as a commercial product, or in an occupational setting (16, 21, 23, 26). Epidemiologists, risk assessors, and others often classify the degree of exposure or potential exposure, by using the concentration of a given chemical in the media that humans contact, integrated over the time of contact; this is then the basis of an exposure index (18). When humans have contact with these environmental media, the chemical may enter the body via inhalation, ingestion, and/or skin absorption. Once in the body, the chemical may distribute to tissues, and adverse health effects may result.

The amount of chemical absorbed in body tissue is called the internal dose. Common measures used to determine internal dose are the blood and urine levels of chemicals or their metabolizes (23). A portion of this internal dose may reach and interact with a target site over a given period so as to alter physiologic function; this portion is called the biologically effective dose (23). All exposure and dose terms are further defined by Sexton et al. (26). Various methods have been used to assess human exposure to xenobiotics. Our charge is to concentrate on the use of biomarkers in exposure assessment. We do this by presenting case studies which demonstrate the benefits of using biomarkers as opposed to those exposure indices that do not use biomarkers of exposure. We then describe biomonitoring programs and analytical methods that may be beneficial to priority setting in TSCA. In the

following paragraphs, we critique various methods that have been used to assess human exposure.

EXPOSURE INDEX (WITHOUT BIOMARKERS)

Traditionally, exposure has been assessed by estimating the individual's or population's potential for exposure. If the concentration of a given chemical in various media is known, then the total concentration of that chemical in the environmental media that humans contact, integrated over the time of contact, forms the basis of an exposure index. The concentration of the chemical in the environmental media is sometimes based on analytical measurements of environmental samples – water, air, soil, food – collected at the exposure site near or as to close to the time of exposure as possible. Depending on the pathway of exposure, all of these environmental media, and perhaps multiple samples of each, may have to be analyzed at a high cost and yet may not be representative of the concentration of the chemical in the media at the time of human exposure. For example, is the average level of a pollutant in fish caught in a river representative of all such fish in that river? Perhaps the "best" environmental sample for an airborne chemical would be a personal air sample collected at the time of exposure by an organic vapor badge; in one experiment this technique correlated more highly with blood levels of selected volatile organic compounds (VOCs) than did VOC levels in breathing zone air that was collected by charcoal tubes (12). Clearly, such techniques are only available in designed experiments. The estimated time of contact, including frequency and duration, with the environmental medium containing the chemical is generally collected by questionnaire or information obtained during history taking. This combination of questionnaire/history information and environmental measurements are then weighted into an exposure model, which is used as an estimate of exposure for each person. We call this the "environmental approach" for assessing exposure.

This approach may be useful in human exposure assessment as a preliminary screen to help ascertain the potential for human exposure. Various models have been used for both qualitative and quantitative predictions. However, they are based on a plethora of assumptions that may contain several potential problems, such as the inability to adjust for individual factors that relate to how much chemical enters the body and how much is absorbed (individual metabolism differences, individual nutritional status during exposure, individual differences in surface area or body mass, and personal habits such as hand-to-mouth activities). In addition, the frequency and duration of contact with the environment that contains the chemical are difficult to estimate because of uncertainty of recall or bias in administering and answering the questionnaire. This bias may arise whenever non comparable information is obtained from the different study groups, a factor that may be the result of the interviewer eliciting or interpreting the information differently (interviewer bias) or of the participants either intentionally or unintentionally reporting the events in a non comparable manner (recall bias). For example, participants may have problems recalling the frequency of playing on contaminated soil or consuming a certain food. Thus, we believe that such exposure indices are useful but are not the best means to assess human exposure to environmental chemicals.

By definition, the best measure of exposure for assessing dose-response relationships is the biologically effective dose. Ideally, environmental health scientists would like to have sensitive and specific measurements of the biologically effective dose. However, identifying the target site(s) of the chemical is a major impediment to using measures of the biologically effective dose to quantify exposure. Even when the target site is known, an invasive procedure may be required to sample that site (e.g., liver, brain). Some organic toxicants or their metabolizes covalently bond to DNA, thus forming a DNA adduct; most notably, carcinogens and mutagens form such adducts. The measurement of such adducts is called the biologically effective dose, but the levels of these

adducts may reflect only recent exposure because of DNA repair. Likewise, measurements of adducts with hemoglobin and other proteins, such as albumin, have also been considered measurements of the biologically effective dose, and as exemplified by 4-amino biphenyl, the hemoglobin adduct has been shown to be significantly associated with DNA adduct concentration in the human bladder epithelial cells (21). Some of these adducts are specific markers for a toxicant (e.g., benzo(a)pyrene in lymphocytes), whereas others are much less specific (e.g., DNA adducts with alkyl groups). The measurement of adducts in humans is still in the developmental stage, and for most chemicals, much more information is needed before the biologically effective dose can be used as a quantitative measurement of exposure (28). Nonetheless, it can be used as a marker of exposure. Other disadvantages to be considered in these measurements are that sample throughput may be too low for moderate-size epidemiological studies, and many adducts may arise from a single chemical.

The next most useful exposure measures are those of internal dose. The direct measurement of a chemical or one of its metabolizes in blood or urine has significantly improved human exposure assessment and thus has improved assessing the risk to humans of many important chemicals. For example, it is fair to say that without blood lead measurements, most of the central nervous system effects of low-level lead exposure could not have been detected.

To interpret blood or urine chemical levels accurately, analysts must know the pharmacokinetics of the chemical and also must have a knowledge of the background levels found in the general population. For example, some chemicals, such as VOCs, are rapidly eliminated, whereas others, such as the chlorinated hydrocarbon pesticides, may have a half-life in humans of greater than 5 years. Thus, such information is critical for interpreting whether the measured concentration of a chemical reflects recent exposures, long-term (chronic) exposures, or both. Of course, to the extent possible, it is still of great importance for the epidemiologist to collect, non-biased information from study participants regarding their potential exposure.

Additional biomarkers that have been monitored in humans include biomarkers of susceptibility and effect. Biomarkers of response, such as cytogenetic markers, stress proteins, and enzyme induction, are sometimes classified as exposure biomarkers and sometimes as effect biomarkers. We will not consider them in this presentation because of space limitations but more importantly because these biomarkers are very nonspecific; i.e., abnormalities of these biomarkers would not specify to what TSCA chemical one may have been overtly exposed, if any.

EXAMPLES OF USE OF BIOMARKERS IN EXPOSURE ASSESSMENT

We will now demonstrate how biomarkers of exposure have been used in exposure assessment in epidemiological studies and how this approach is preferred over the "environmental approach" for assessing human exposure. In so doing, we do not wish to imply that the environmental approach is meaningless, but that the biologic approach is preferred as a marker of human exposure. Certainly, in risk management when the objective is to reduce the potential exposure, the "environmental approach" is useful for identifying where the pollution is taking place.

Dioxin: Operation Ranch Hand Study

From 1962 through 1970 during the Vietnam Conflict, the main mission of the U.S. Air Force's Operation Ranch hand was to spray defoliants, such as Agent Orange, over densely vegetated areas of South Vietnam. Agent Orange consisted of an equal mixture of 2,4-D and 2,4,5-T in diesel oil; the 2,4,5-T was contaminated with 2,3,7,8-TCDD (dioxin) in the parts-per-million range. Dioxin is lipid soluble and thus tends to be stored in the lipid-rich depots of the human body. Dioxin has a long half-life-more than 7 years in humans (20, 25). In 1982, the Air Force began a prospective cohort study, specifically looking at health, reproductive, and mortality outcomes that might be associated with exposure to Agent Or-

ange and other herbicides containing dioxin. These health studies will examine the veterans of Operation Ranch Hand every 5 years through the year 2002. One of the first tasks was to develop an exposure index in order to classify each veteran's exposure; this index would then be used as the basis for exposure and for correlating with any health effects.

This exposure scenario was similar to that of exposure in an occupational setting in that the primary exposure was thought to be direct exposure to the herbicide itself, rather than indirect exposure through an environmental pathway. The exposure index consisted of the average concentration of dioxin in the Agent Orange during one's tour of duty multiplied by the number of gallons of Agent Orange sprayed during one's tour divided by the number of men in one's specialty during that period. The total number of eligible men in the study was limited to the 1200 to 1300 survivors of the Operation. The U.S. Air Force and various review boards believed that this index not only could serve as a reliable basis for assessing exposure to dioxin but that any noted adverse health effects could be related to this index.

In 1987, the U.S. Air Force contracted with our laboratory to analyze 150 serum samples from Operation Ranch Hand veterans in order to compare the Air Force's exposure index with the measured internal dose of the veterans. There was essentially no correlation between the exposure index and the serum dioxin level (14). Because of this finding, the Air Force further contracted with CDC to analyze the serum of all surviving members of Operation Ranch Hand, and this serum-dioxin level became the exposure index used to correlate with any adverse health effects (33). Had the Air Force used its original exposure index for the Operation Ranch Hand study, a great deal of misclassification would have resulted, and any health effect conclusions of the study would have been invalid.

Thus, the use of the serum dioxin measurement, the biomarker, was preferred over the exposure index that was derived without the biomarker.

Dioxin: U.S. Army Ground Troops in Vietnam

The chemical of concern was again the dioxin in Agent Orange. The potential environmental pathways were skin contact with and inhalation of the spray containing the herbicide, skin contact with sprayed vegetation and soil, and ingestion of water and food that had been sprayed. The amount of dioxin in the Agent Orange from 1966-1969 was known. The duration of contact was gathered from questionnaires given to the veterans and from U.S. military records containing the locations of military units, the locations where herbicide was sprayed, and the dates when the herbicide was sprayed.

Six exposure indices were generated from this information; four of the indices were based on a soldier's potential for exposure from direct spray or on his being located in an area that had been sprayed within the previous 6 days; the other two exposure indices used self-reported data and included an index that was based on the veteran's perception of how much herbicide he has been exposed to. To test the validity of these exposure indices, CDC measured serum dioxin levels in 646 enlisted ground troop veterans who had served in III Corps a heavily sprayed area, for an average of 300 days during 1966 to 1969. For comparison, serum-dioxin levels in 97 non-Vietnam U.S. Army veterans who served during the same time were also measured (30).

The results showed no meaningful association between dioxin levels and any of the exposure indices. The mean, median, and frequency distributions for both the Vietnam and non-Vietnam veterans were remarkably similar, indicating that there was little, if any, increased exposure to dioxin in this population. The study had a 95°/0 statistical power to detect a difference of only 0.6 ppt in the medians, but this difference was not found. This finding exemplifies the value of measurements of internal dose in exposure assessment. It also points out the need to develop specific and sensitive methods, for if the detection limit for dioxin had been 20 ppt (lipid adjusted), then most all the results would have been non detectable. Furthermore, because elevated expo-

sures could not be documented, plans for a prospective cohort health study were dropped.

Dioxin: Occupational Setting

CDC National Institute of Occupational Safety and Health (NIOSH) conducted a retrospective study to evaluate health outcomes, including mortality from cancer, among more than 5000 workers who may have been occupationally exposed to dioxin, as a result, for example, of the production of 2,4,5 -trichlorophenol (9). Many of these workers were deceased. Because many were deceased and because of the large number of potentially exposed men, NIOSH epidemiologists had to develop an exposure index for use in correlating the health outcomes (the effect). Serum dioxin measurements were performed on 253 workers; the results were compared to various exposure indices. From this analysis, epidemiologists determined that the best exposure index was years of work in a job with potential exposure. Since this exposure index had been validated to and calibrated with serum dioxin levels, it could be used as the exposure index in this study and exposure and effects could be compared directly with those found in other studies. This process again demonstrates the need for measuring the internal dose in exposure assessment or health effect studies.

Lead

Toxicity associated with high levels of lead in humans has long been recognized. However, biochemical and epidemiological studies have noted hematological and neurological damage among children with relatively low levels of lead in their blood and teeth. The second National Health and Nutrition Examination Survey (NHANES II), conducted by the National Center for Health Statistics (NCHS), provided blood lead measurements, which were the basis for estimating the degree of exposure of the general U.S. population to lead (1). As a result of federal regulations requiring the removal of lead from gasoline, the amount of lead in gasoline decreased about 55°/0 from early 1976 to early 1980. The population-based NHANES II Study showed that the predicted mean blood level in the U.S. population had decreased 37% during that same period, from 14.6 µg/dL to 9.2 µg/dL. Environmental modeling did not accurately predict the magnitude of the impact of decreasing the amount of lead in gasoline because the contribution of lead from gasoline to humans via the soil was not well characterized. These data were a major factor in the Environmental Protection Agency's (EPA's) decision to implement a more rapid removal of lead from gasoline. This implementation and the banning of the lead-soldering of cans produced in the U.S. have been major factors in the NHANES III predicted mean blood level decreasing to 2.8 µg/dL in the U.S. population in 1990 (24).

Thus, exposure assessment by measuring blood lead levels has been a public health success story. It helped identify lead in gasoline as a major preventable source and showed that removing lead from gasoline was an effective prevention strategy. However, the latest data indicate that 8.9%, or approximately 1.7 million children, aged 1-5 years, have blood lead levels equal to or greater than 10 µg/dL, which is the level of concern under the 1991 CDC guidelines. The population at risk for excessive lead exposure comprises primarily black, inner-city children and has been targeted for more extensive lead poisoning prevention efforts (6). This example again shows the need for biomarkers of exposure for relating exposure to health effects.

Volatile Organic Compounds (VOCs)

Many volatile organic compounds (VOCs) are ubiquitous in the environment. They have been shown to exist in higher concentrations in indoor air than in outdoor air (32). Reported health effects from exposure to VOCs have included eye irritation, sick-building syndrome, neurological effects, and cancer. CDC developed an isotope-dilution purge and trap gas chromatography/mass spectrometry method to quantity 32 VOCs (see table 10-1) in 10 mL of blood with detection limits in the parts-per-trillion range (3). This method is a full-scan method at 3000 resolving power, so that in addition to acquiring quantitative data on these 32 VOCs, many additional VOCs can be

Table 10-1 Biological Monitoring at CDC's National Center for Environmental Health

Metals (typical urine or blood sample -3 mL; typical limit of detection - low parts per billion (ppb)

Lead	Beryllium	Arsenic
Mercury	Chromium	Thallium
Cadmium	Nickel	Vanadium

Polychlorinated dibenzo-dioxins, polychlorinated dibenzo-furans, coplanar polychlorinated biphenyls (PCBs) **(measured in serum from one 25 mL blood sample if exposure is near background levels - smaller samples are adequate** for higher exposures; typical limit of detection - low parts-per-trillion (ppt) on a lipid-weight basis, low parts-per-quadrillion on a whole-weight basis)

2,3,7,8-Tetrachlorodibenzo-p-dioxin	(TCDD)	1,2,3,4,7,8-Hexachlorodibenzofuran	(H_xCDF)
1,2,3,7,8-Pentachlorodibenzo-p-dioxin	(P_nCDD)	1,2,3,6,7,8-Hexachlorodibenzofuran	(H_xCDF)
1,2,3,4,7,8-Hexachlorodibenzo-p-dioxin	(H_xCDD)	1,2,3,7,8,9-Hexachlorodibenzofuran	(H_xCDF)
1,2,3,6,7,8-Hexachlorodibenzo-p-dioxin	(H_xCDD)	2,3,4,6,7,8-Hexachlorodibenzofuran	(H_xCDF)
1,2,3,7,8,9-Hexachlorodibenzo-p-dioxin	(H_xCDD)	1,2,3,4,6,7,8-Heptachlorodibenzofuran	(H_pCDF)
1,2,3,4,6,7,8-Heptachlorodibenzo-pdioxin	(H_pCDD)	1,2,3,4,7,8,9-Heptachlorodibenzofuran	(H_pCDF)
1,2,3,4,6,7,9-Heptachlorodibenzo-p-dioxin	(H_pCDD)	1,2,3,4,6,7,8,9-Octachlorodibenzofuran	(OCDF)
1,2,3,4,6,7,8,9-Octachlorodibenzo-p-dioxin	(OCDD)	3,3',4,4'-Tetrachlorobiphenyl	(TCB)
2,3,7,8-Tetrachlorodibenzofuran	(TCDF)	3,4,4',5-Tetrachlorobiphenyl	(TCB)
1,2,3,7,8-Pentachlorodibenzofuran	(P_nCDF)	3,3',4,4',5-Pentachlorobiphenyl	(P_nCB)
2,3,4,7,8-Pentachlorodibenzofuran	(P_nCDF)	3,3',4,4',5,5'-Hexachlorobiphenyl	(H_xCB)

Volatile organic compounds (VOCs) (measured in one 10 mL blood sample; typical limit of detection - low ppt)

1,1,1 -Trichloroethane	Acetone	Ethylbenzene
1,1,2,2-Tetrachloroethane	Benzene	Hexachloroethane
1,1,2-Trichloroethane	Bromodichloromethane	m-/p-Xylene
1,1-Dichloroethane	Bromoform	Methylene chloride
1,1-Dichloroethene	Carbon Tetrachloride	o-Xylene
1,2-Dichlorobenzene	Chlorobenzene	Styrene
1,2-Dichloroethane	Chloroform	Tetrachloroethene
1,2,-Dichloropropane	cis-1,2-Dichloroethene	Toluene
1,3-Dichlorobenzene	Dibromochloromethane	trans-1,2-dichloroethene
1,4-Dichlorobenzene	Dibromomethane	Trichloroethene
2-Butanone		

Chlorinated pesticides and non-coplanar polychlorinated biphenyls (measured in serum from one 5 mL blood sample; typical limits of detection - low ppb)

Aldrin	Biphenyls, Polychlorinated (congeners)	Endrin
Chlordane, alpha	DDD	Heptachlor
Chlordane, gamma	Trans-Nonachlor	Heptachlor epoxide
beta-Hexachlorocyclohexane	DDE	Hexachlorobenzene
gamma-hexachlorocyclohexane	DDT	Mirex
Biphenyls, Polychlorinated (total)	Dieldrin	Oxychlorodane

Non-persistent pesticides (measured in one 10 mL urine sample typical limits of detection - low ppb)

Urine metabolites	Parent Pesticides
2-Isopropoxyphenol (IPP)	Propoxur
2,5-Dichlorophenol (25DCP)	1,4-Dichlorobenzene
2,4-Dichlorophenol (24DCP)	1,3-Dichlorobenzene, Dichlofention, Prothiofos, Phosdiphen
Carbofuranphenol	Carbonfuran, Benfuracarb, Carbosulfan, Furanthiocarb
2,4,6-Trichlorophenol (246TCP)	1,3,5-Trichlorobenzene, Hexachlorobenzene, Lindane
3,5,6-Trichloro-2-pyridinol (TCPY)	Chloropyrifos, Chlorpyrifos-methyl
4-Nitrophenol (NP)	Parathion, Methyl parathion, Nitrobenzene, EPN
2,4,5-Trichlorophenol (245TCP)	1,2,4-Trichlorobenzene, Fenchlorphos, Trichloronate, Lindane
1-Naphthol (1 NAP)	Naphthalene, Carbaryl
2-Naphthol (2NAP)	Naphthalene
2,4-Dichlorophenoxyacetic acid (24D)	2,4-D
Pentachlorophenol (PCP)	Pentachlorophenol

qualitatively identified and in many cases, quantified (5).

CDC, with financial support from the Agency for Toxic Substances and Disease Registry (ATSDR), selected a 1,000 person subset of the NHANES III population to determine reference ranges for these 32 VOCs. The 1,000 people were chosen from both sexes, all regions of the contiguous U. S., urban and rural residents, and were adults between 20 and 59 years of age (19). The data showed that 11 of these VOCs were measured in more than 75°/0 of the people with the non chlorinated aromatics being the most prevalent. These VOCs included styrene, toluene, ethyl benzene, o-xylene, m,p-xylene, and benzene, which is a known human carcinogen. The primary sources of these compounds are tobacco smoke and exhaust from internal combustion engines. The non endogenous compound found at the highest concentration and highest frequency was 1,4-dichlorobenzene (4). The blood exposure data for this moth repellent and room deodorizer correlated highly with urinary levels of its primary metabolize, 2,5-dichlorophenol (1 1). This high correlation indicated that either blood 1,4-dichlorobenzene or urinary 2,5-dichlorophenol levels could be used as a biomarker of exposure to 1,4-dichlorobenzene.

Five of the VOCs were found in 10%-75% of the selected population, whereas the remainder of the VOCs were in less than 10% (4). Thus, this latter group would be of low priority for inclusion in human effect studies. These analytical methods and reference range studies have been applied to a wide variety of case studies and population studies. These include exposure assessment studies of toxic waste sites, oil-well fires (7), sick building syndrome (4), multiple chemical sensitivity, and oxygenated fuels involving methyl tertiary-butyl ether (MTBE) (15). In each of these examples, the blood concentrations of VOCs were compared with the reference-range population data. However, pharmacokinetic data are needed to properly interpret blood levels of VOCs. Scientists from CDC and EPA have collaborated in determining the half-lives of many VOCs in humans subjected to low level mixtures of VOCs in well-controlled chamber studies (2). The blood half-lives were less than one-half hour, but the elimination time courses were multiexponential, thereby suggesting multiple storage sites within the body. The blood uptake portion of the 4-hour exposure curve exhibited a rapid uptake that reached a plateau after about 50 minutes; the uptake rate was not concentration dependent, but the blood concentration was directly dependent on the air concentration. When exposure ceased after 4 hours, the decay was rapid, but the decay rate also reached a plateau after about 1 hour; however, the VOC levels remained elevated even 24 hours after exposure as compared with the pre-exposure blood levels. Thus, like those compounds with long biologic half-lives, such as dioxin, VOCs also can be the focus of exposure assessment studies, if the blood samples are collected within 1 day following exposure.

COLLECTING AND BANKING OF HUMAN SPECIMENS

We have presented examples of the benefits of biomarkers of exposure; now we focus on the mechanisms of collecting and banking human specimens for such biomonitoring. The first U.S. program of biological monitoring tissue specimens for environmental pollutants and also for human tissue specimen banking was the National Human Monitoring Program (N-IMP), which began in 1967 and was conducted by the U.S. Public Health Service. When the U.S. Environmental Protection Agency (EPA) was created in 1970, the NHMP was transferred to it. One of the major activities of the NHMP was the National Human Adipose Tissue Survey (NHATS), which was designed to be a continuously operating survey that would collect, store, and analyze autopsy and surgical specimens of human adipose tissue from the major U.S. metropolitan areas. However, during the 1980s, budget cuts restricted the NHMP to a reduced and modified NHATS, which continued until 1990. In 1991, the National Research Council published its findings that programs that provide more useful data based on

probability samples for the entire U.S. population should be designed and properly funded (1 7).

One program that is based on a national probability sample is the National Health and Nutrition Examination Survey (NHANES), which is conducted by the Centers for Disease Control and Prevention's National Center for Health Statistics. Data from NHANES I, II, and 111 have provided important information on the prevalence of various health conditions and distributions of physical and biochemical characteristics of the U.S. population. As previously mentioned, data on blood lead levels in NHANES II and III provided longitudinal trend data on human levels and the effect of legislation on that trend. The data also pinpointed a sub population still at risk for excessive lead exposure. Serum levels of cotinine, the major metabolize of nicotine, are being measured in NHANES III in order to ascertain exposure levels as a result of both active and passive smoking (29). As mentioned previously, CDC measured blood VOCs and selected urinary pesticide residues in a subset of the NHANES population in order to assess human exposure to these compounds. In addition, blood, urine, and DNA have been banked from NHANS III.

Phase I of the National Human Exposure Assessment Survey (NHEXAS), which is conducted under cooperative agreements with the EPA, began in 1995. These Phase I studies are population based surveys for exposure assessment to selected environmental pollutants in the state of Arizona and in EPA's Region V (29).

Designing and implementing national probability sampling surveys for human exposure assessment must consider many issues (8). However, certainly NHANES, and now NHEXAS, have addressed these issues. Therefore, the mechanism is in place to collect and bank specimens needed to assess biomarkers of exposure in human specimens for many of the chemicals included in TSCA.

PRIORITIZING CHEMICALS

We have presented examples of the benefits of biomarkers of exposure and the ability of programs, like NHANES, to collect and bank the needed biologic specimens for assessing human exposure to many of the chemicals included in TSCA. This does not argue that the entire number of probability based samples have to be analyzed but that mechanisms are in place to collect such samples. Assuming the needed biologic samples are available, the list of TSCA chemicals must be prioritized for the effective application of biomarkers for human exposure assessment. The following factors would be included in such prioritization:

- potential for human exposure
  - degree of exposure
    - pounds produced per year
    - physical/chemical characteristics of chemical
    - how the chemical is made, used, fate
  - number of people potentially exposed
  - susceptible population
- hazard identification/severity of effect information
- dose/response information in both animals and humans
- possibility of measuring biomarkers

Such prioritization of this chemical list would therefore involve development of a model that would include the following factors: the potential for human exposure (degree and number), severity of adverse effects in a dose response manner, and the possibility of the biomarkers existing and ability of the laboratory to develop the needed analytical methods. For those chemicals that lack the needed information, quantitative structure activity relationship data, if available, would also be used. Exposure databases (1 O) would be used in this process.

ANALYTICAL METHODS

As mentioned previously, one of the criteria for prioritizing the list of chemicals for the development of biomarkers is the possibility of measuring biomarkers of exposure; i.e., does a biomarker exist and can the laboratory develop the needed analytical methods to measure the biomarker? Unless the biomarker exists, there is

no need for the analytical method. Assuming the biomarker exists, the analytical methods should have the following characteristics:

- Multianalyte (several biomarkers)
- Compatible with sample matrix
- Demonstrated acceptable sensitivity
- Demonstrated acceptable specificity
- Demonstrated acceptable precision
- Demonstrated acceptable accuracy
- Cost effective
- Rapidity

These characteristics, except for cost effective and rapidity, can be defined in objective terms. Certainly, the methods used in our examples meet the needed objective criteria. For measuring organic biomarkers of exposure, the analytical methods that are atop the method hierarchy include high resolution mass spectrometry and tandem mass spectrometry using the isotope dilution technique for quantification. Whether a particular analyses is cost effective is more subjective. For example, the cost for the measurement of 32 VOCs in 10 mL of Blood is about $500 per sample or less than $20 per analyte. Commercial prices for measuring the 17 polychlorinated dibenzo-p-dioxins and furans plus 4 co-planar polychlorinated biphenyls that are in human serum are about $1000 per sample or less than $50 per analyte. One can decide if this is too costly for the intended purpose.

Historically, mass spectrometric methods have suffered in the area of rapidity or high throughput, but this is not always the case. For example, for the measurement of cotinine in NHANES III, serum extracts are analyzed at the rate of 1 every 2 minutes by using high performance liquid chromatography/atmospheric pressure ionization tandem mass spectrometry. This technique also requires less sample preparation than traditional methods although sample preparation is the rate limiting step because of the speed of mass spectrometric analysis.

Other methods which may appear to be more amenable to screening methodologies; i.e., low cost and rapid, have been developed for many chemicals, primarily pesticides, in the environmental area (15, 31). To expand this list to many of the TSCA chemicals in biological specimens would require much work in both developing the antiserum and the methods. Many of the current immunoassay have high levels of false positives (because of cross-reactivity or matrix effects) and false negatives (because of matrix effects unless sufficient sample preparation procedures are followed). Therefore, frequently to meet the objective requirements of the desired analytical methods, one must employ methods of higher specificity for many of the samples. One new technique that employs many of the advantages of immunoassay with the specificity and multianalyte capability of the mass spectrometer is a mass spectrometric immunoassay (15). Such combinations of techniques will be used increasingly for biomonitoring.

We believe that the bottom line is that following some prioritization of the chemicals, if the biomarker of exposure exists in a readily accessible biologic specimen, such as blood or urine, this biomarker can be measured effectively to assess human exposure and thus be used to help prioritize TSCA chemicals for health effect screening. The converse that biomarkers of effect can help prioritize chemicals for exposure studies is also true.

SUMMARY

We have attempted to show that a biomonitoring program would be beneficial in assessing human exposure to many of the chemicals on the TSCA list. Such a program might be also a way to 1) establish reference ranges in the general population; 2) identify sub populations potentially at risk; 3) establish trends in exposure and, hence, judge the effectiveness of pollution prevention practices and regulations; 4) provide dose assessment over total exposure; 5) and provide a data base for comparison with other data sets such as ecological data sets. The needed sample collection programs and analytical procedures are now available for conducting such a program. These procedures incorporate the benefits of having the required sensitivity, specificity, and

multi-chemical measurements and are cost effective. National population-based programs such as the National Health and Nutrition Examination Survey (NHANES) or the National Human Exposure Assessment Survey (NHEXAS) could be used to collect the specimens. Each of these would offer certain advantages. The TSCA list would have to prioritized by using an algorithm consisting of the potential for human exposure, severity of adverse human effects and the possibility of measuring the required biomarker. Once this model is formed, it could be validated by the biomonitoring program.

We have also included a list of the chemicals (table 10-1), for which CDC has national human internal dose data, the biologic specimen needed and amount, and the lower detection limits; these data are from various sources and are of varying quality for predicting national mean and ranges of human levels. Nonetheless, they do show whether exposure is common for particular chemicals. In addition, many of these chemicals, such as the pesticides, are not on the TSCA list.

We believe that there is a hierarchy of means to assess human exposure. This hierarchy includes self reports, professionally-developed exposure questionnaires, measurements of external dose, and modeling of all or portions of these data. All of this information may be useful, but we believe that the "gold standard" is the measurement of a biomarker of exposure in human specimens. Thus, if exposure data and classification from any of the other techniques are to be used, they should be both validated and calibrated to human biomonitoring data. However, programs such as NHANES or NHEXAS and many of the analytical methods are available to gather exposure information on many TSCA chemicals. This exposure information would then be used to determine which chemicals should be examined for health effects, for without a receptor population, there would be little need to study associated health effects.

REFERENCES

1. Annest, J. L., et al., "Chronological Trend in Blood Lead Levels Between 1976 and 1980," *New England Journal of Medicine 308:373-1377, 1983.*
2. Ashley, D. L., and Prah, J.D. "Time Dependence of Blood Concentrations During and After Exposure to a Mixture of Volatile Organic Compounds," personal communication,
3. Ashley, D.L., et al., "Determining Volatile Organic Compounds in Human Blood from a Large Sample Population by Using Purge and Trap Gas Chromatography/Mass Spectrometry," *Analytical Chemistry* 64:1021-1029, *1992.*
4. Ashley, D. L., et al., "Blood Levels of Volatile Organic Compounds in a Nonoccupationally Exposed U.S. Population and In Groups of Suspected Exposure," *Clinical Chemistry* 40:1401-1404, *1994.*
5. Benin, M. A., et al., "Measurement of Methyl Tert-Butyl Ether and Tert-Butyl Alcohol in Human Blood by Purge-and-Trap Gas Chromatography-Mass Spectrometry Using an Isotope-Dilution Method," *Journal of Analytical Toxicology* 19:187-191, *1995.*
6. Brody, D.J., et al., "Blood Lead Levels in the U.S. Population," *Journal of the American Medical Association* 272:277-283, 1994.
7. Etzel, R. A., and Ashley, D. L., "Volatile Organic Compounds in the Blood of Persons in Kuwait During the Oil Fires," *Int. Archives in Environmental Health* 47:1-5, 1994.
8. Ezzoti-Rice, T. M., and Murphy, R. S., "Issues Associated With the Design of a National Probability Sample for Human Exposure Assessment," *Environmental Health Perspectives* 103(3):55-59.
9. Fingerhut, M.A., et al., "Cancer Mortality in Workers Exposed to 2,3,7,8 -Tetrachlorodibenzo-p-dioxin," New *England Journal of Medicine* 324:212-218, 1991.
10. Graham, J., et al., "Role of Exposure Databases in Risk Assessment," *Archives of Environ. Health* 47:408-420, 1992.

11. Hill, Jr., R. H., et al., "p-Dichlorobenzene Exposure Among 1,000 Adults in the United States," *Archives of Environmental Health 1995.*
12. Mannino, D. M., et al., "Human Exposure to Volatile Organic Compounds: A Comparison of Organic Vapor Monitoring Badge Levels with Blood Levels," *International Archives of Occupational Environmental Health 67:59-64, 1995.*
13. Meulenbert, E.P., Mulder, E. H., and Stoks, P. G., "Immunoassay for Pesticides," *Environment Science and Tech.* 29:553-561, 1995.
14. Michalek, J. E., "The Value of Epidemiological Studies, " *Applied Industrial Hygiene* 12:68-72, 1989.
15. Moolenarr, R. L., et al., "Methyl Tertiary Butyl Ether in Human Blood After Exposure to Oxygenated Fuel in Fairbanks, Alaska," *Archives in Environmental Health* 49:402-409, *1994.*
16. National Research Council, Committee on Advances in Assessing Human Exposure to Airborne Pollutants, "Principles of Exposure Assessment," *Human Exposure Assessment for Airborne Pollutants: Advances and Opportunities* (Washington, DC: National Academy Press, 1991).
17. National Research Council, *Monitoring Human Tissues for Toxic Substances* (Washington, DC: National Academy Press, 1991).
18. Needham, L. L., "Examples of Measuring Internal Dose for Assessing Exposure in Epidemiological Studies," *Environmental Epidemiology: Effects of Environmental Chemicals on Human Health, Advances in Chemistry Series,* W.M. Draper (cd.) (Washington, DC: American Chemical Society,)
19. Needham, L. L., et al., "The Priority Toxicant Reference Range Study: Interim Report, *Environ Health Perspectives* 103(3):89-94, *1995.*
20. Needham L. L., et al., "Half-Life of 2,3,7,8-TCDD in Serum of Seveso Adults: Interim Report," *Organohalogen Compounds 21:81-86, 1994.*
21. Needham, L. L., et al., "Case Studies of Relationship Between External Dose and Internal Dose," *Journal of Exposure. Analysis in Environmental Epidemiology.* l(Supplement):209-221, 1992.
22. Nelson, R. W., et al., "Mass Spectrometric Immunoassay," *Analytical Chemistry* 67: 1153-1158, 1995.
23. Lioy, P.J., "Assessing Total Human Exposure to Contaminants," *Environmental Science and Technology* 24:938-945, 1990.
24. Pirkle, J. L., et al., "The Decline in Blood Lead Levels in the United States," *Journal of the American Medical Association 272:284-291, 1994.*
25. Pirkle, J. L., et al., "Estimates of the Half-Life of 2,3,7,8-TCDD in Ranch Hand Veterans," *Journal of Toxicology and Environmental Health* 27:165-171, *1989.*
26. Sexton, K., Callahan, M.A., and Bryan, E. F., "Estimating Exposure and Dose to Characterize Health Risks: The Role of Human Tissue Monitoring in Exposure Assessment," *Environmental Health Perspectives.* 103(3): 13-*29, 1995.*
27. Skipper, P.L, and Tannenbaum, S. R., "Molecular Dosimetry of Aromatic Amines in Human Populations," *Environmental Health Perspectives* 102:17-21, *1994.*
28. Skipper, P.L, and Tannenbaum, S. R., "Protein Adducts in the Molecular Dosimetry of Chemical Carcinogens," *Carcinogenesis* 11:507-518, 1990.
29. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, "Preliminary Data: Exposure of Persons Aged Greater Than or Equal for 4 Years - United States, 1988 -1991," *Mortality and Morbidity Weekly Report* 42:37-39, 1993.
30. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, "Veterans Health Studies: Serum 2,3,7,8-Tetrahlorodibenzo-p-dioxin Levels in U.S. Army Vietnam-Era Veterans," *Journal of the American Medical Association* 260:1249-1254, *1988.*

31. Van Emon, J.M., and Lopex-Avilla, V., "Immunochemical Methods for Environmental Analysis," *Analytical Chemistry* 64:79-88, 1992.
32. Wallace, L.A., et al., "The TEAM Study: Personal Exposures to Toxic Substances in Air, Drinking Water, and Breath of 400 Residents of New Jersey, North Carolina, and North Dakota," *Environmental Research* 43:290-307, 1987.
33. Wolfe, W. E., et al., "Health Status of Air Force Veterans Occupationally Exposed to Herbicides in Vietnam, Physical Health," *Journal of the American Medical Association* 264:1824-1831, 1990.

11

Exposure Evaluations Through Biomonitoring

P. J. (Herb) Hakkinen and Elin J. Eysenbach

ABSTRACT: *Testing and screening are conducted, in most cases, to support risk assessment. Therefore, any discussion of testing methodologies should take into account not only testing or hazard evaluation, but also exposure evaluation. Although the workshop exposure assessment discussion focused on a biomonitoring approach, an environmental factors approach to exposure assessment is much more broadly used for Toxic Substances Control Act (TSCA) decision-making. The level of detail needed for exposure assessment depends on the type of decisions involved. Exposure assessment may be at a screening level or may involve detailed data collection and analysis, including biomonitoring when the costs are warranted.*

Humans are exposed to chemicals via breathing air, drinking water or bathing, coming in contact with soil, eating food, using consumer products, etc. Exposure assessment is the determination or estimation (qualitative or quantitative) of the magnitude, frequency, duration, and route of exposure, either for a particular exposure scenario, or for up to all known potential exposures to a given chemical. These comments provide an overview of the status of human exposure assessment, how improving exposure screening and assessment and helping to ensure quality and consistency among exposure assessors are key to improving how human risk assessments are performed, other ways to improve the science of exposure assessment, a description of the tiered approach and the use of monitoring to validate modeling.

OVERVIEW OF HUMAN EXPOSURE ASSESSMENT

A 1991 National Academy of Sciences, National Research Council report (10) stated that "Exposure assessment is an integral and essential component of . . . risk assessment.... Exposure assessment is an equal partner with toxicology." Other documents and scientists have noted that exposure assessment is still perhaps the overall weakest link in risk assessment, and has the greatest opportunities for improvement. Therefore, it is important to incorporate its status into any testing and screening review.

Humans are exposed to chemicals via breathing air, drinking water or bathing, coming in contact with soil, eating food, using consumer products, etc. Exposure assessments attempt to assess the degree or magnitude of contact a person has with a chemical, either from a particular route of exposure or exposure scenario, or as a summation of up to all known potential exposures. Qualitative exposure screening tools estimate the likelihood and magnitude of exposure and the nature of potentially exposed populations.

Factors affecting the degree of exposure include the duration and frequency of exposure, the route of exposure (e.g., oral, skin, and inhalation), and the degree of uptake of the chemical from a given route and location on the body (i.e., there can be large differences in body area (e.g., hands versus forehead) skin permeability). Other factors affecting the degree of exposure include human characteristics such as differences in metabolic activation and deactivation of a chemical, differences in age (e.g., adults and children have large differences in the amount of air breathed per minute, the amount of food or liquid consumed per day, body weights, and body surface areas). Even among adults or children as individual

groups, there can be large differences in physiological parameters such as body weight and consumption of food.

The degree of contact with a chemical is driven by the known (i.e., measured) or suspected concentration(s) of a chemical in the media being assessed. Analytical measurements of the medium of interest yield direct values for use in exposure assessments. Biomonitoring of body fluids (e.g., urine, blood, or exhaled air) or body parts (e.g., hair or fingernails) can be used in the exposure assessment (see below for further discussion). Various modeling approaches can be used to factor-in a chemical's stability, home air changeover rates, weather conditions, the distance from the source of exposure to the potentially exposed subject, or other information.

Improving exposure screening and assessment, and helping to ensure quality and consistency among exposure assessors, are key to improving how human risk assessments are performed:

In recent years, great strides have been made toward improving the science of exposure screening and assessment, both in the values and approaches used, and in helping to ensure consistency and quality among exposure assessors (13, 14, 15). Current US Environmental Protection Agency (EPA) efforts include revising the Exposure Factors Handbook, development of additional resource and guidance documents such as the "Residential Exposure: A Source Book" cooperative effort between EPA, the Society for Risk Analysis, and the International Society of Exposure Analysis, and the "THERdbASE" (Total Human Exposure Risk Database and Advanced Simulation Environment) cooperative agreement between EPA and the University of Nevada, Las Vegas. The latest version of THERdbASE is available to exposure assessors around the world via the Internet's World Wide Web (http: //eeyore.lv-hrc.nevada.edu) as a downloadable set of files. Once the files are downloaded, exposure assessors are able to model a wide variety of possible exposures using data they can input and data available from several THERdbASE databases (e.g., food consumption patterns, physiological parameters, human activity patterns, and chemical properties).

Other recent noteworthy efforts helping to ensure quality and consistency include those of the European Centre for Ecotoxicology and Toxicology of Chemicals (1, 3, 4, 7, 8, 11)

WAYS TO IMPROVE EXPOSURE SCREENING AND ASSESSMENT

Two recent publications in particular have discussed improvements to exposure screening and assessment techniques. Whitmyre et al. (16) noted the following potential improvements:

- use of more appropriate exposure default values;
- incorporation of time-activity data;
- the use of reasonable exposure scenarios;
- the use of stochastic or probabilistic approaches;
- use of bivariate analysis;
- use of less than lifetime exposure; and
- incorporation of physiological considerations relevant to absorbed dose estimation.

Whitmyre et al (16) also discussed other ways to improve the exposure assessment process, and identified key research needs.

Paustenbach (12) presented several "lessons learned" in the United States about how to improve exposure assessments. They include:

- avoid too much emphasis on risk estimates for the maximally exposed individual (MEI);
- evaluate the uptake (absorbed dose) for both the 50% and 95% persons;
- avoid repeated use of conservative or worst-case assumptions. Incorporate Monte Carlo techniques whenever possible;
- ensure a proper statistical analysis of environmental data;
- conduct sensitivity analysis to understand fragility of dose estimates;
- understand the role of environmental fate when estimating exposure;
- validate the reasonableness of the exposure estimates;

- consider using biological monitoring to confirm exposure estimates; and
- consider all indirect pathways of exposure.

A TIERED APPROACH: ESSENTIAL FOR SOUND RESOURCE ALLOCATION

Many companies, regulatory agencies, and others use a tiered approach to risk assessment and its components, such as exposure assessment. Iterations proceed from low effort, inexpensive first-cut evaluations to increasingly complex, costly and data intensive assessments. The appropriate level of exposure assessment, whether preliminary, qualitative, or quantitative, is determined by the nature of the decisions to be made. Iterations would be increasingly detailed with the specific approach selected at each iteration or tier being determined by considering the decision-making needs, available resources, existing data, and other factors.

A preliminary exposure screening is frequently used to set priorities for testing, product development or regulation. This may be simply a volume cut as an exposure surrogate, or it may involve very rough exposure scenarios quantifying the likelihood, of some exposure, but not the magnitude. Although still a preliminary screen, consideration of factors such as likelihood, magnitude, and nature of exposed population can assist in obtaining the most benefit from decisions for testing, regulatory consideration, or other expenditures. Such an initial exposure assessment may be designed to determine whether potential for exposure exists. It may be based on available public, government, or company data to support initial development or risk assessment activity, and to identify key data needs and areas of uncertainty to be addressed later.

Later iterations, i.e., detailed exposure assessments, are generally conducted by one of three approaches: predictive, direct, and reconstructive. The predictive approach estimates exposures based on modeling of a chemical's transport to the receptor and transformations resulting form environmental fate processes, as well as on knowledge of activities that bring the receptor organism into contact with the chemical. The direct approach attempts to quantify exposure while it is taking place by measuring concentration of the agent in the media of contact, e.g., air in the breathing zone. The reconstructive approach back-calculates exposure based on concentrations of a chemical or a chemical's metabolite in biological tissues, fluids, or exhaled breath.

Use of the reconstructive method concurrently with model development for the predictive method enhances future optimization and improves confidence in modeling results. It is useful to compare biomonitoring results to modeling results to validate or confirm the modeling approach, assumptions, and parameter values. Publications discussing biomonitoring and model validation are National Academy of Sciences (9) and US EPA (14). As predictive modeling is less costly and time-consuming than biomonitoring, validation studies increase assurance of effective resource deployment.

These later iterations involve an increase in sophistication in exposure assessment techniques as required to support a particular level of decision-making. Key in this activity is the judgment of the risk assessment experts about how much information is needed at any given time in the product development or risk assessment cycle, along with expert judgment about when enough risk assessment-related work has been done to support, for example, commercialization of the chemical and the resulting potential for human and environmental exposures. References discussing the tiered approach to human risk assessment for chemical exposures include Hakkinen and Leep (6), Jayjock and Hawkins (8), European Centre for Ecotoxicology and Toxicology of Chemicals (3), European Commission (4), and Organization for Economic Cooperation and Development (1 1). From a resource allocation standpoint, it is important to recognize that monitoring, and even modeling techniques for exposure assessment are quite costly and would not be economically supportable, even at an early tier, for more than a small number of chemicals or chemical applications.

REFERENCES

1. American Industrial Health Council, *Exposure Factors Source Book* (Washington, DC: 1994).
2. Chemical Manufacturers Association, *Exposure Glossary* (Washington, DC: 1994).
3. European Centre for Ecotoxicology and Toxicology of Chemicals, *Assessment of Non-Occupational Exposure to Chemicals,* Technical Report No. (Brussels, Belgium: 1994).
4. European Commission, *Risk Assessment of Existing Substances Technical Guidance Document,* XI/919/94-EN (Brussels, Belgium: 1995).
5. Hakkinen, P. J., et al., "Multimedia Exposure and Risk Assessment Models" (letter), *Society of Environmental Toxicolgy and Chemistry's SETAC NEWS* Pages 10-11 (November, 1993).
6. Hakkinen, P.J., and Leep, C., "Industry's Use of Risk, Values, Perceptions, and Ethics in Decision Making," *Handbook for Environmental Risk Decision Making: Values, Perceptions, and Ethics, C.* Richard Cothern (cd.) (CITY, STATE: Lewis Publishers, in press).
7. Hawkins, C. et al., "A Rationale and Framework for Establishing the Quality of Human Exposure Assessments," *American Industrial Hygiene Association Journal* 53:34-41, *1992.*
8. Jayjock, M.A., and Hawkins, N.C., "A proposal for Improving the Role of Exposure Modeling in Risk Assessment," *American Industrial Hygiene Association Journal 54:733-741, 1993.*
9. National Research Council, *Frontiers in Assessing Human Exposure to Environmental Toxicants* (Washington, DC: National Academy Press, 1991).
10. National Research Council, *Human Exposure assessments for Airborne Pollutants: Advances and Opportunities* (Washington, DC: National Academy Press, 1991).
11! Organization for Economic Cooperation and Development, *Environment Monograph No. 70. Report of the OECD Workshop on Occupational and Consumer Exposure Assessments* (draft) (CITY, STATE: PUBLISHER, 1993).
12. Paustenbach, D.J., "The Practice of Health Risk Assessment in the United States (1975-1995): How the U.S. and Other Countries Can Benefit from that Experience," *Human and Ecological Risk Assessment* 1:29-79, 1995.
13. U.S. Environmental Protection Agency, *Exposure Factors Handbook, EPA/60018-89/043* (Washington, DC: 1989).
14. U.S. Environmental Protection Agency, "Guidelines for Exposure Assessment," final rule, *Federal Register 57: 22888-22938, 1992.*
15. U.S. Environmental Protection Agency, *Dermal Exposure Assessment: Principles and Applications,* EPA/600/8-91/011B (Washington, DC: 1992).
16. Whitmyre, G.K., et al., "Human Exposure Assessment II: Quantifying and Reducing the Uncertainties," *Toxicology and Industrial Health* 8:321-342, 1992.

12

SAR and Modeling

Ann M. Richard, Pauline Wagner, Richard Purdy, and Gilman Veith

ABSTRACT: SAR plays a prominent role in TSCA screening of new chemicals and existing chemicals in commerce. SAR models for bioavailability, ecotoxicology and health toxicology endpoints are being used to identlify chemicals with the greatest potential for ecological or health hazard, to set testing priorities, and to provide scientific support for a testing recommendation. SAR models vary considerably in accuracy and utility for screening application depending upon the quality of available data and level of current knowledge for a toxicity endpoint. The main deficiency of current SAR capabilities is inadequate data, and lack of knowledge of mechanisms of toxicity for many chemical classes and toxicity endpoints of potential regulatory concern. Knowledge or inference of a common mechanism of toxic action is crucial for selecting appropriate chemical analogues, guiding SAR model development, establishing model plausibility, and providing the necessary scientific rationale for model acceptance and use in prediction. This paper discusses the present role and capabilities of SAR in TSCA screening, general features and limitations of SAR, current and evolving SAR technologies, and advances most likely to lead to improvements in SAR models for use in TSCA screening. Although SAR has the clear potential to further reduce the need for testing or eliminate testing in some circumstances, the promise of SAR will not be fulfilled without proper application of these methods. This entails clear recognition of the limitations of SAR, and appreciation for the essential roles of research into mechanisms of toxicity, and strategic testing for further SAR model development and refinement.

A structure-activity relationship (SAR) relates features of chemical structure to a property, effect, or biological activity associated with that chemical. The fundamental premise is that the structure of a chemical determines its physical properties and activities. The term "structure-activity relationship" has taken on a wide range of meaning over the years, from heuristic chemical associations and human expert approaches that consider primarily structural features, to formal mathematical relationships that relate specific chemical attributes to a quantitative measure of the property or activity of interest, the latter being commonly referred to as "quantitative structure-activity relationships" (QSARs). In both the pharmaceutical and chemical industries, structure-activity considerations have long been used to design chemicals with commercially desirable properties. In the environmental protection field, SAR is being used to predict adverse ecological and health effects, with applications ranging from the prediction of relevant properties, such as chemical stability, bioavailability and bioaccumulation, to the prediction of various forms of chemical toxicity.

The focus of this workshop is testing and screening strategies for review of the Toxic Substances Control Act (TSCA) inventory of existing chemicals in commerce. This problem poses a significant and immediate challenge, not only in terms of the sheer numbers of chemicals that have undergone little testing or review (>10,000), but also in terms of the multiple exposure routes and ecological and health endpoints of potential concern. The foremost goal is to identify the chemicals that pose the greatest potential ecological and health risks, and to strategically allocate limited testing resources to best characterize these risks. SAR, coupled with exposure and use estimates, represents the top tier in a multiple tier screening approach for assessing chemical hazard, and provides the primary means for setting testing priorities. SAR currently plays a prominent

This manuscript has been reviewed by the U. S. Environmental Protection Agency and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

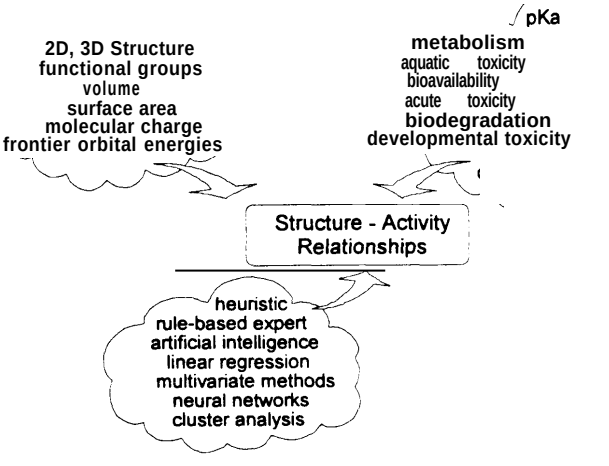
role in testing and screening strategies for TSCA review of new chemicals and existing chemicals in commerce. SAR screening is being used for hazard identification, to set testing priorities, to provide scientific support for a testing recommendation, and in a relatively new U.S. Environmental Protection Agency (EPA) initiative, to aid in the design of safer chemicals by suggesting modifications in structure predicted to minimize toxicity. SAR has the potential to further reduce the need for property measurements and animal testing, generate insight into mechanisms of action, and achieve better environmental protection by providing for more efficient screening of the TSCA inventory for a wide range of toxicity endpoints.

The following will consider some general characteristics of SAR, principles of application to toxicity screening, limitations and guidelines for use of current SAR technologies, current SAR capabilities being applied to the TSCA screening problem, and new technologies and advances that will lead to improvements in SAR capabilities for toxicity screening.

SAR FOR TOXICITY SCREENING: GENERAL CONSIDERATIONS

SAR approaches are extremely general with respect to the possible representations of chemical structure, the types of chemical or biological activity that can be modeled, and the methods for relating the two (figure 12-1). In contrast, an SAR model is highly specific to the particular set of chemicals, attributes, and experimental activities used in its derivation. An SAR model codifies and rationalizes existing data. It follows that the range of application, predictive accuracy, and ultimate relevance of the SAR model is wholly determined by the quality and quantity of existing data and knowledge upon which the SAR model is based. For example, an SAR model developed to predict mutagenicity based on qualitative (+/- activities) experimental data for a series of aromatic amines tested in the TA100 Salmonella reversion assay is not likely to be applicable to other chemical classes (e.g. small haloorganics), other strains of Salmonella (e.g. TA90), or the prediction of quantitative potencies. An obvious corollary is that an SAR model is only as relevant to the ultimate health effect of concern (e.g. carcinogenicity) as the toxicity endpoint that it purports to model (e.g. mutagenicity).

Figure 12-1: Structure-Activity Approaches



There are two complimentary goals of an SAR study: 1) to predict the activities of untested chemicals and prioritize chemicals according to relative activities; and 2) to provide a rational scientific basis for understanding and interpreting existing biological/chemical activity data in terms of chemical structure and mechanism of action. The former is the primary allure of SAR, while the latter provides the foundation and prescription for its success.

SAR methods are optimally applied to the prediction of intrinsic physical properties of a chemical where a single "mechanism" is operative and the property can be considered independent of external interactions. A number of high quality QSAR and computational chemistry models have been developed and are used routinely by industry and regulatory agencies to predict chemical properties such as vapor pressures, melting points, acid dissociation constants, spectral properties, chromatographic retention times, and octanol/water partition coefficients (commonly referred to as "logP"), to name but a few.

Such models have the advantages of significant cost savings over laboratory determinations, speed, ease of use, and no need for the availability or handling of the chemical of concern. Modeled properties also serve as key parameters for use in the development of SAR models for biological fate or effect. For example, an octanol/water partition coefficient, a property used extensively in QSAR studies of biological activity, approximates the ability of a chemical to transport through biological membranes and can be modeled easily and accurately by the computerized CLOGP method (13), yielding cost savings from $10,000 to $30,000 per chemical.

Toxicology provides a more severe SAR modeling challenge. In this case, the extrinsic chemical "property" being modeled is a biological endpoint, i.e. an activity determined by the complex interaction of a chemical within the biological system. Whereas an intrinsic chemical property relates unambiguously to a single physical process or mode-of-action, there are most often many possible mechanisms by which chemicals with different structural characteristics elicit a common biological activity or toxicity endpoint. This complexity coupled with lack of knowledge concerning mechanisms of toxicity introduces greater uncertainty and imposes greater restrictions on the application of SAR concepts to toxicology. The key to ameliorating these concerns is to restrict SAR models, whenever possible, to chemicals that elicit their effect by a common mode-of-action, and to incorporate whatever knowledge is available concerning the mechanism of toxicity into SAR model development. This does not necessarily require full, detailed knowledge of the molecular mechanism, but a common mode-of-action must be indirectly inferred or hypothesized to maximize validity and reliability, and minimize uncertainty in the SAR model. (The terms "mechanism" and "mode-of-action" are used interchangeably in the present text). This explicit linkage between SAR and mechanism of action is crucial to establishing the plausibility of an SAR model and providing the necessary scientific rationale for its acceptance and use (6).

It follows that SAR models will be most successful when applied to mechanistically well-defined toxicity endpoints. Such endpoints are more likely to consist of specific biochemical indicators (e.g. P-450 induction, inhibition of DNA repair), *in vitro* bacterial assays, tissue and organ-specific effects, and *in vivo* assays where a common unifying process, transformation, or event is central to the activity. Examples include the central role of: logP or bioavailability in narcosis mechanisms of acute aquatic toxicity; formation and stability of electrophilic nitrenium ion intermediates in mutagenicity of nitroaromatics; Ah receptor binding in toxic effects of dioxin and PCBs; and dermal penetration (logP) and acidity (pKa) in determining skin corrosivity. The most difficult types of toxicity endpoints to model with SAR are termed "apical" endpoints, i.e. typically whole animal *in vivo* assays of chronic disease or effect that consider as much of the integrated physical and biological process as possible in a single test (e.g. developmental toxicity, neurotoxicity behavioral effects, rodent carcinogenicity). While these assays are often considered most relevant and useful to human health or ecological risk assessment, they are also the most costly, most controversial in terms of animal usage, least likely to be available, and most difficult to interpret mechanistically. In these cases, restriction of the SAR to a narrowly defined chemical class is the best assurance that a common mode-of-action applies.

Another essential element of an SAR model is the data used in its development, i.e. the chemicals and activities or potencies. There are two fundamentally distinct types of SAR models for any toxicity endpoint, those that model the conditions for distinguishing between activity classes, e.g. "actives" and "inactive", and those that model the conditions for modulating potencies among a group of chemicals belonging to a common activity class, i.e. "actives" (5). The SAR requirements for being a member of the "active" class may be quite distinct from those that explain differences in potency among the actives. In addition, data requirements differ for the two objectives: sufficient test data on

negative, or inactive analogues are essential for establishing boundaries of an SAR, while test data on positives, or actives covering a wide range of potencies are required for QSAR development. Often, sufficient and appropriate negative test data for SAR model development are lacking. Testing recommendations are driven by concern for adverse effects and, hence, testing resources are applied to the chemicals considered to pose the greatest hazard, i.e. most likely to be active in a toxicity assay. In addition, negative test data are less likely to be published and available since they are perceived to be of less interest to the scientific community. A legitimate role of strategic testing, particularly in a research setting, should be to challenge and improve the quality of current SAR models and, in some cases, verification of a negative test prediction may be the best use of available resources. Often the most dramatic structure-activity differences among analogues, e.g. where a minor structural change eliminates or imparts an activity, are the most informative and useful in SAR analysis. For example, addition of a single methyl group in the bay region of the PAH, benz(a)anthracene, eliminates its carcinogenic activity due to steric crowding and blocking of metabolic activation to the ultimate carcinogen, i.e. the diol epoxide.

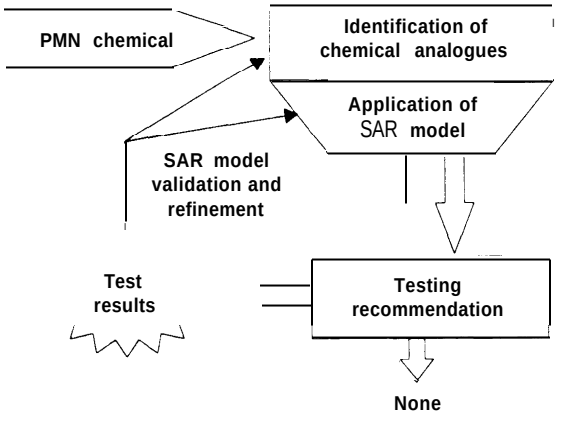
SAR has been most successfully applied to classes of organic chemicals where quantitative, reproducible activity data are available for pure chemicals with known structures. Since SAR requires knowledge of individual chemical structures, it cannot be applied to uncharacterized chemical mixtures. When SAR is applied to polymers, it generally deals with reactivity characteristics of the monomeric units. Also, very little SAR modeling has been done for inorganic chemicals, i.e. metals or metal complexes, due to sparsity of data on chemical analogues, and the greater challenges in characterizing and calculating the pertinent chemical characteristics of these species in biological systems. Subject to these constraints, conditions for optimal SAR model development and application include: restriction of the SAR model to a well-defined toxicity endpoint or single mode-of-action chemical class; availability of test data for a range of chemical structures, attributes, and potencies; use of mechanistically relevant molecular descriptors; a mechanism-based scientific rationale for SAR model and predictions; and prospective validation on test chemicals not used in SAR model derivation. Optimally, some knowledge of a possible or probable biological mechanism of action guides parameter development, provides the basis for determining chemical analogy, defines the region of chemical/activity space where the SAR model is likely to be applicable, i.e. places limits on model extrapolation, and provides scientific rationale for a model prediction (17). Even in the absence of explicit knowledge of mechanisms of toxicity, however, an SAR model developed mindful of the above constraints has the potential to generate insight into possible mechanisms of toxicity and guide further experimentation.

Hence, there is a continuum of SAR modeling tools, biological endpoints, and considerations that impact on the relevance and utility of SAR models for use in toxicity screening. The next section will consider current SAR capabilities being applied to TSCA screening.

USE OF SAR IN TSCA REVIEW OF NEW AND EXISTING CHEMICALS

The bulk of the SAR expertise within EPA currently being brought to bear on the TSCA existing chemicals problem has evolved out of the Pre-Manufacture Notification (PMN) review process (2). Hence, the PMN process, and its strengths and weaknesses warrants some discussion. By law, TSCA requires companies wishing to manufacture a chemical not on the TSCA Inventory to submit a premanufacturing notice (PMN) to EPA. EPA then has 90 days in which to determine if the manufacture, processing or use of that chemical in commerce may present an unreasonable risk to human health or the environment. If this is determined, EPA has the legal authority to request further test data be submitted for the PMN chemical. The Structure-Activity Team (SAT) within the Office of Pollution Pre

Figure 12-2: EPA Screening Procedure



vention and Toxics at EPA was conceived in order to efficiently and systematically screen PMN chemicals for health and/or ecological hazard. The 20 or so members of the SAT represent a wide range of chemical, ecological, and toxicological disciplines. Some characteristics of the PMN process are as follows. Since no toxicity testing is required by law, test data accompany fewer than 5% of the PMN chemicals submitted. Hence, SAR frequently provides the sole means for evaluating these chemicals. In the area of ecotoxicology, a number of computerized, chemical class-based QSAR models have been developed for use in predicting physical/chemical properties, ability to degrade and bioconcentrate, and toxicity to fish, aquatic invertebrates, and algae (8, 10, 23). In the health toxicology area, models and SAR expertise vary considerably depending on the state of knowledge in the particular field of toxicology and, in contrast to ecotoxicology, most models are qualitative and heavily reliant on chemical analogy, rules and expert judgement.

The mandate of the SAT is primarily operational, i.e. to evaluate more than 2000 chemicals/yr within a 90-day deadline from the date of each PMN submission (20). The SAT operates under strict confidential business information (CBI) restrictions with respect to the chemicals it evaluates, which prohibits the sharing of chemical structures used in SAR model development with outside parties. Computers are used for data base searching, to aid the identification and retrieval of chemical analogues, for the calculation of chemical properties required for estimating bioavailability and fate, and for the application of QSAR models for eco-tox endpoints. Finally, there is an emphasis on mechanism-based approaches and interpretations, whenever possible, to reduce uncertainty in the SAR prediction, increase plausibility, and provide the necessary scientific rationale to support a testing recommendation.

The PMN screening process is summarized in figure 12-2. Upon receiving a PMN chemical submission, the SAT reviews the literature and in-house data bases of previously reviewed chemicals to identify possible analogues. Analogues consist of chemicals with similar structures or fictional groups to the PMN chemical, for which test data are available or a previous SAT assessment is on record. In some cases, when very little is known about the chemical or a toxicity endpoint, this is the extent of the SAR, i.e. available data for the analogue are assumed to apply to the PMN chemical. This information could flag the PMN chemical as a potential fish toxicant, developmental toxicant, carcinogen, etc. In other cases, additional SAR considerations or models apply to the chemical class, of which the PMN chemical is assumed to be a member, and are used in making a testing recommendation. This SAR hazard assessment is considered along with exposure data in making the “may present an unreasonable risk” determination, which may trigger a testing requirement under Section 5 of TSCA. The importance of the analogue selection step as the top-most tier in this overall process should be stressed. If suitable analogues are unavailable, or if inappropriate analogues are chosen with respect to toxic mode-of-action, inappropriate SAR considerations and incorrect judgement could be applied to the PMN chemical under review. Finally, an extremely important element of the PMN process occurs subsequent to the issuance of a testing requirement. Comparison of the toxicity test result to the SAT prediction provides the primary mechanism for the continual validation and refinement of the SAR models and assumptions used in the PMN process.

By virtue of their legal mandate, the EPA/SAT has evolved into a unique and valuable resource that plays a pivotal role in environmental and health protection. Nowhere else in the world has such a concentration and wide range of expertise been focused solely on the task of SAR model development and ecological fate and toxicity screening. The SAT is also unique in terms of its unparalleled access to unpublished, proprietary, and internally generated toxicity test data from a wide range of sources and for a wide variety of chemicals, data which are essential for the development and refinement of predictive SAR models. More than 24,000 chemicals have been screened through the PMN process since 1979, contributing greatly to the evolution and improvement of SAR expertise and models in current use. In addition, the SAT has engaged in a number of outside collaborations to further verify and improve upon existing models. An example is a recent collaboration with their European Community (EC) counterparts in which the SAT blindly evaluated 144 chemicals concurrently undergoing toxicity testing (19). This and other exercises have provided support for many of the SAR models and assumptions in current use by the EPA/SAT, while pointing to deficiencies in others. The overall performance results of the European Community exercise are available in the form of a joint EPA/EC summary report from either the European Union in Brussels, or the EPA as Document Number EPA 743-R-94-001. However, confidentiality restrictions required that all of the individual chemical identities associated with the exercise be destroyed, a loss that limits the potential benefit of the study to both the EPA and the outside SAR community.

The current expertise of the EPA/SAT is being used in the development of a "Use Cluster Scoring System" for evaluation of the TSCA existing chemicals inventory. A tiered strategy has been implemented, the first step involving the identification of "use clusters", i.e. categories of common use chemicals, such as paints, rust inhibitors, plasticizers, etc., that are likely to have similar exposure scenarios (e.g. paints - occupational inhalation exposure). The second stage involves prioritization of these use clusters based on SAR, exposure, and available toxicity data. A more complete SAR toxicity evaluation using the models and expertise of the EPA/SAT, is then applied within the use cluster to establish testing priorities among the individual chemicals. All available test data, which may include possible occupational exposure health data for existing chemicals in commerce, are considered in the preliminary toxicity screening assessment.

The main deficiencies of the EPA/SAT approach for TSCA review of PMN chemicals or existing chemicals in commerce, shared by the SAR community at large, are inadequate data and lack of knowledge of mechanisms of toxicity for many of the chemical classes and toxicity endpoints of potential regulatory concern. In addition, the SAT has neither the mandate, nor the time or resources to evaluate new technologies or to carry out research to improve existing SAR models. Hence, a deliberate outreach effort must be made by the SAT to communicate and interact with industry and research groups with the potential to impact on the process. While the final SAR models and expertise developed by the EPA/SAT can and are being made available to the public (two examples being the ECOSAR program for eco-tox screening, and the ONCOLOGIC expert system for predicting chemical carcinogenicity), CBI confidentiality constraints prohibit the release of the primary data and chemical structures used in model development. CBI constraints are designed to safeguard the rights of industry, yet are in perpetuity under existing law, regardless of whether the chemical was ever produced or entered into commerce. These constraints limit the ability of outside parties to independently scrutinize, validate, and improve upon EPA/SAT models. Access to the data used in SAR development is valuable for defining the proper boundaries of application of the SAR model, for developing hypotheses concerning the structural basis for the toxic mode-of-action, and for refining or developing alternative SAR models.

QSAR/SAR models are also developed and used by industry for addressing TSCA require-

ments. However, these models tend to be tailored and restricted to the specialty chemicals produced and used within a particular industry (e.g. solvents, adhesives, etc.). These SAR models, and data on which they are based, are usually considered proprietary. There tends to be more limited SAR expertise within industry with respect to the wide range of toxicity endpoints of potential concern under TSCA, and more limited access to data than is available to the EPA/SAT. Particularly when in-house expertise is lacking, there is incentive for industry to take advantage of the PMN process for toxicity screening prior to large dollar investments in research and/or development. A PMN submission costs little and is performed within a short time frame. Even when expertise is available to industry, there is incentive to anticipate PMN toxicity estimates that would trigger a testing requirement, rather than to develop independent estimates. An EPA testing requirement for a PMN chemical often provides sufficient incentive for industry to redirect a line of research or abandon plans to manufacture a potentially toxic chemical, providing an effective means for serving the interests of environmental protection. For the review of existing chemicals in commerce, the economic incentive to avoid possible regulatory action is much greater since considerable investment in the chemical has already taken place. In this case, industry is more inclined to challenge testing requirements by independent SAR estimates.

CURRENT SAR APPROACHES AND EMERGING TECHNOLOGIES

The most widely used paradigm for QSAR study is the linear free-energy relationship (LFER) or Hammett equation approach, based on statistical linear regression fit of steric, electronic, and hydrophobicity terms to biological potency. This is a chemical class-based approach, designed to be applied to a range of structurally similar, or "congeneric" chemicals that are assumed to have a common mechanism of action. LFER equations are the basis of the ECOSAR compilation of QSARs for ecotoxicology used by the SAT, and the majority of published QSARs for biological endpoints. Within the QSAR paradigm, further incremental advances will come from better mechanism-based chemical classifications, generation of additional test data, and development of more mechanistically relevant molecular parameters and descriptors. For example, while it is now possible to predict with reasonable accuracy the acute toxicity of "unreactive" (i.e. narcosis mechanism) chemicals to many aquatic species, models are less reliable for chemicals acting by alternate mechanisms of action. In particular, there is movement in the QSAR field towards incorporation of more rigorous quantum mechanical properties related to potential reactivity and energy characteristics of molecules derived from their three-dimensional structure.

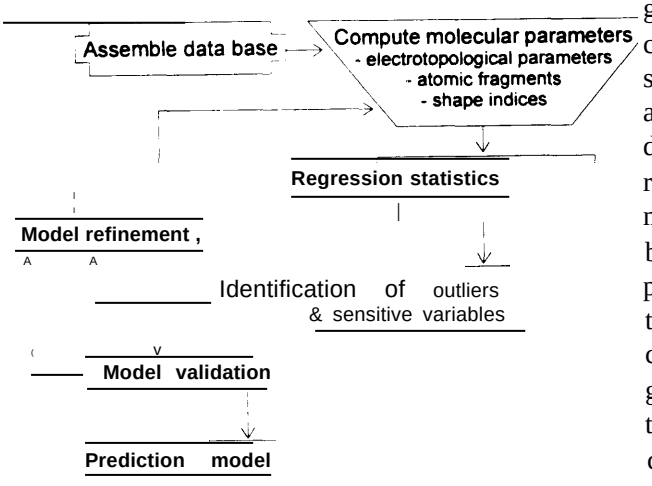
Each individual LFER QSAR equation is associated with a relatively narrow range of chemicals and a specific biological endpoint and, thus, has limited applicability to other toxicity prediction problems. An approach being advocated by Corwin Hansch, one of the pioneers of the QSAR field, is to process larger units of existing information than individual QSARs in order to generate insight into unifying features of biological processes (12). Over the past several years, Hansch and coworkers have compiled over 3000 existing QSARs from the literature into a computerized data base, CQSAR (9), for easy access, comparison, and study. The CQSAR data base is also being used as a validation tool, to judge individual QSARs in a larger biological context by lateral examination of related, or overlapping QSARs, i.e. QSARs for similar chemicals/different endpoints, QSARs for similar endpoints/different chemicals, or QSARs for different chemicals/different endpoints having a similar functional form. For example, a very general feature of the CQSAR data base is that >85% of the QSARs contain a major contribution from a hydrophobicity term (logPo/w), and the coefficient of this term is almost always in the range of 1-2. This argues that QSARs without a logP term, or with a logP coefficient significantly deviating from 1-2 should be considered either novel or suspect.

Table 12-1: TOPKAT Available Models

- Rat acute oral LD50 (19 submodels, 400 chemicals)
- Rat chronic oral LOAEL
- Mouse inhalation LC50
- Developmental toxicity potential (3 submodels, w/ and w/o maternal toxicity)
- Carcinogenicity
- Mutagenicity
- Fathead minnow acute LD50
- Daphnia EC50
- Biodegradability
- Skin/eye irritancy (Draize test)

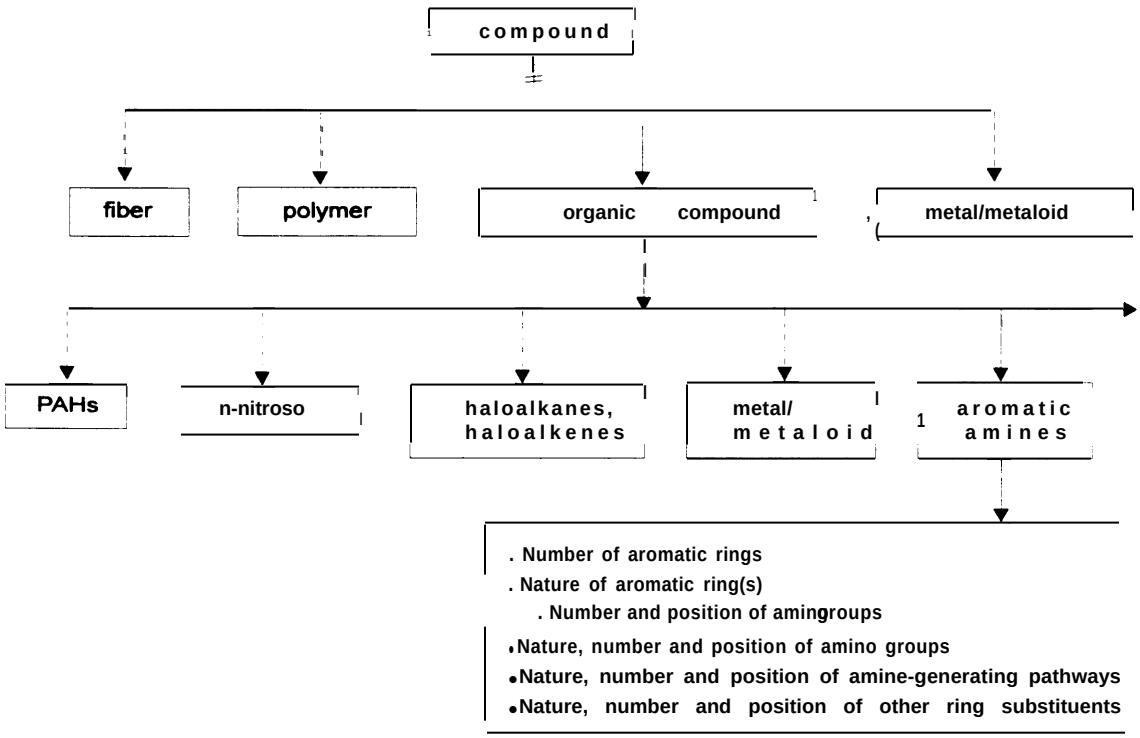
An example of a commercially available SAR program, in use by some industry and government groups, is the TOPKAT computer-based toxicity prediction program (18). TOPKAT is based on LFER concepts, but has typically been applied to SAR modeling of large data sets of "non-congeneric" chemicals, i.e. chemicals representing many chemical classes and mechanisms of action. TOPKAT applies traditional statistical approaches, such as multiple-linear regression and discriminant analysis, to identify SAR associations between structure-derived chemical properties and activity. Indicator variables, i.e. variables that take on a value of O or 1 depending on the presence or absence of a molecular feature, provide an approximate means for incorporating multiple chemical classes into a common SAR model, e.g. a parameter could be "turned on" or "turned off' if a molecular feature such as a nitro group were present. TOPKAT models have been developed and are available for the endpoints listed in table 12-1. The model development procedure is summarized in figure 12-3 and culminates in an SAR model for toxicity prediction. The main limitations of the TOPKAT approach, shared by other statistically-based computerized SAR programs such as CASE and ADAPT, are: limitations in chemical descriptors; dissociation of SAR model development from biological mechanism considerations; and the abandonment of the chemical class restrictions of traditional QSAR (16). Models have been developed for large, chemically diverse data sets associated with complex toxicity endpoints known to represent many possible modes-of-action and, since little effort has been made to incorporate mechanism considerations, models tend to be difficult to interpret and scientifically rationalize. In addition, there has been a tendency towards over-reliance on statistical indicators of model predictive capabilities and underestimation of the inherent uncertainty of these models due to their biological component. For these reasons, TOPKAT, and other statistically-based toxicity prediction programs are not currently used by the SAT for TSCA screening.

Figure 12-3: TOPKAT Method



TOPKAT does, however, have some useful features and legitimate uses. One of TOPKAT's greatest strengths is its high quality data bases compiled from private sources and an exhaustive search of the literature, where each experiment and activity call is carefully evaluated prior to data base incorporation. TOPKAT provides ready access to this existing data and an automated means for identifying chemical analogues based on structural features. TOPKAT also employs conservative statistical analysis and validation procedures. A TOPKAT analysis of a non-congeneric data set is potentially useful for generating mechanism hypotheses when very little prior knowledge is available for classifying chemicals according to mechanism. TOPKAT can also serve as a potentially valuable supple

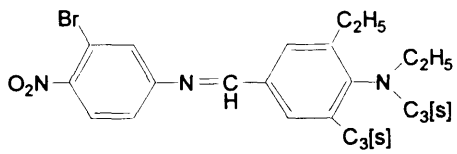
Figure 12-4: OncoLogic Cancer Expert System



ment to expert rule-based approaches when the latter are relatively undeveloped (e.g. in the developmental toxicity field), or as an approximate preliminary screening capability when little expertise is available. Recognizing current limitations of their approach, TOPKAT developers are moving towards the goal of more "mechanistic" models in the sense of restricting model development to smaller, more well-defined activity endpoints and chemical classes (1 1). (See examples in table 12-1.) These initiatives offer hope for improving the utility of such models for toxicity prediction in TSCA screening.

Traditional QSAR studies attempt to discover new, previously unknown mathematical relationships for predicting activity from chemical structure. In contrast, an expert system aims to reproduce the human expert decision process for evaluating chemical toxicity by codifying current knowledge. An example of the latter is the OncoLogic cancer prediction expert system (15), being developed as a collaborative effort between the cancer experts within the EPA/SAT and outside expert systems programming consultants. OncoLogic is an artificial intelligence, rule-based expert system that can be applied to a wide range of non-congeneric chemicals, but that relies on a chemical class, mechanism-based approach to cancer prediction (22). It incorporates literally thousands of discreet rules for characterizing each of a variety of chemical classes based on the cancer expertise of the SAT. Due to the enormous size of this undertaking, the program is currently operational only for metals, polymers, fibers, and a few classes of organic chemicals, with capabilities for other chemical classes still under development. OncoLogic has a hierarchical structure as represented in figure 12-4. In the top-most levels, structural considerations are used to exclude molecules from concern on the basis of factors such as molecular weight, volubility, and bioavailability. A chemical is then classified according to properties and structure features until sufficient characterization allows application of

Figure 12-5: Sample OncoLogic Carcinogenesis Evaluation Justification Report



Summary: The level of caranogenicity concern from this aromatic amine is LOW - MODERATE.

Justification: In general, the level of carcinogenicity concern of an aromatic amine is determined by considering the number of rings, the presence or absence of heteroatoms in the rings; the number and position of amino groups; the nature, number and position of other nitrogen-containing "amine generating groups", and the type, number and position of additional sustituents.

...

The evaluation of this compound proceeds as if the di-akyl substituted amino group, NR1R2 [where R1= ethyl; R2=sec-propyl] were a free amino group. The influence of the N-akyl groups on the bioactivation of the compound is considered at the end of the evaluation.

...

The reduced electron conducting properties of the intercyclic linkage are expected to lower the overall level of concern. Therefore the level of carcinogenicity is reduced to MODERATE.

the SAR rules for a specific chemical class, such as shown for aromatic amines. A sample output is also shown in figure 12-5, illustrating a portion of the lengthy, mechanism-based rationale provided to support the prediction of LOW-MODERATE concern for this sample chemical. The rules consider issues such as metabolism and activation to reactive electrophiles, and accurately reproduce and communicate the mechanism-based rationale of the EPA/SAT cancer experts. Since these rules are distinct program units, they can be modified as knowledge advances. By communicating the detailed rationale for EPA/SAT cancer predictions, OncoLogic allows industry and others to identify and challenge prevailing SAR assumptions by directed research. OncoLogic also makes the current cancer prediction expertise of the SAT more accessible and widely available within and outside EPA.

An expert system is only as good as the rules and knowledge upon which it is based. Although a few other expert systems are currently available for toxicity prediction, most are based on more limited knowledge and expertise than OncoLogic, and have more limited appeal for toxicity screening. One possible exception is in the area of metabolism prediction. Many chemicals require metabolic activation as a precondition to toxicity. In many of these cases, modeling the conditions for metabolic activation, or modeling the metabolizes instead of the parent compounds, is the key to developing a successful SAR for toxicity prediction. OncoLogic, for example, incorporates numerous metabolism rules for organic chemicals. Metabolism expert systems, such as MetabolExpert (14), provide industry and the larger SAR community valuable access to expertise concerning likely metabolic pathways and products for many chemicals of concern.

As has been stressed, mechanism-based chemical classification is one of the primary requirements for successful SAR model development, providing the scientific basis for the chemical analogue selection step in figure 12-2. The criteria for analogue selection is a key area of uncertainty in many SAR models since it is usually based on organic chemistry principles derived independent of the biology, and may not reflect similarity in terms of biological mechanisms of toxicity. The problem of choosing appropriate analogues is illustrated by the example of peroxisome proliferators. These chemicals are structurally diverse, yet have highly similar pleiotropic, toxicological responses that strongly suggest a common receptor-mediated mode-of-action. Hence, it is the biological

response, not the apparent chemistry, that directs one to group these chemicals into a common class for the purpose of SAR model development.

A few research groups are considering biological means for classification of chemicals for use in SAR studies. In recent years, Bradbury and coworkers at EPA's Environmental Ecology Laboratory have moved away from traditional chemical class-based QSARs for predicting aquatic toxicity and towards the generation of biological mechanism-based QSARs (7). They first established a mode-of-action knowledge base covering a broad range of chemicals, exposure regimes and endpoints. Empirical assessment of toxicity mechanism was then determined by consideration of joint toxic action studies, physiologically-based toxic response syndromes, and single chemical dose-response curves, yielding a variety of toxic mode-of-action classifications (e.g. baseline narcosis, oxidative phosphorylation uncouplers and respiratory inhibitors). Only after such biological classifications were determined were efforts centered on QSAR analysis and understanding the chemical mechanisms and structural criteria for underlying activity. Hence, in applications, biologically-based chemical classifications would define criteria for choosing appropriate chemical analogues and identify the relevant QSAR for use in a toxicity screening application.

A second example is provided by the Rules Induction Method for Predicting Chemical Carcinogenesis developed by Bahler and Bristol in a collaborative effort between NIH/NIEHS and academia (3). This is an automated, decision-tree approach where rules for use in prediction are mathematically induced from available data, rather than obtained by human experts. In contrast to TOPKAT, the rules are derived from both chemical and biological "attributes". These include: Salmonella mutagenicity (SAL); electrophilic structural alerts; route of administration; MTD; subchronic organ pathology (up to 59 organ types, up to 40 morphological lesions); and miscellaneous *in vitro* short term test results. As a predictive screening tool, this approach has the limitation that it requires subchronic pathology information from the rodent bioassay, and some *in vitro* assay information. However, significant cost savings would be realized in generating these data as opposed to carrying out a fill 2-year rodent carcinogenicity bioassay. This biological information was also utilized by "human experts" in a recent NTP-44 prospective carcinogenicity prediction exercise where human expert predictive performance was judged superior to the performance of "pure" SAR methods, i.e. methods based on chemical structure alone (l). Perhaps because the Rules Induction Method and human experts used much the same information in their assessments, the Rules Induction Method performed nearly as well as the human experts in the prediction exercise. An alternative use of this information, germane to the present discussion, is as a means for defining biological mechanism-based chemical classifications for subsequent SAR analysis, i.e. using each rule branch to define a possibly distinct mode-of-action chemical class. Two sample rules are shown in table 12-2. All chemicals satisfying Rule#l comprise a subclass of active carcinogens likely to be mechanistically distinct from other active carcinogens in terms of biological attributes. Hence, the structural features common to chemicals in this rule class, and distinct from chemicals belonging to the remaining actives or inactives could constitute an SAR model for prediction of carcinogenicity.

Table 12-2: Sample Induction Rules for Chemical Carcinogenisis Predictor

Rule #1:

IF chemical mutates Salmonella
AND adjusted rat MTD=<750 mg/kg/day,
THEN class is positive.
(Rule true for 90% of 147 chemicals in training set)

Rule #5:

IF chemical does not mutate Salmonella
AND there is no subchronic pathology in male rat pituitary, spleen, or urinary/bladder,
AND there is no subchronic pathology in female rat kidney
THEN class is negative

A discussion of promising technologies for toxicity screening also should include mention of major advances in computational chemistry and 3D modeling and visualization that are yielding greater understanding of the detailed molecular-level interactions and changes ultimately responsible for the toxicity. Such methods are advancing in tandem with increasing computational capabilities, and increasing knowledge of chemical reaction mechanisms, the structure and function of biological receptors, metabolic enzyme activity (cytochrome P-450s, and glutathione), and DNA interactions implicated in various forms of toxicity. Computational chemistry studies have had a significant impact, for example, on understanding of the structural and electronic requirements for DNA-adduct formation and carcinogenicity of polycyclic aromatic hydrocarbons (incorporated as rules in OncoLogic).

Another example illustrating how 3D modeling tools, more commonly employed in drug design, can be applied to toxicity problems once a receptor-based mechanism for a toxicity endpoint has been proposed or established is provided by a recently published 3D-QSAR model for endocrine disrupters developed by Wailer and coworkers at the National Health and Environmental Effects Research Lab of EPA (21). The toxicity of a class of endocrine disrupters was postulated to be due to interaction with the steroid hormone receptor. Since the structure of this receptor was unknown, ligand requirements of the receptor were inferred from a comparison of the three-dimensional structures of known steroid receptor ligands (such as estradiol). The final computerized 3D QSAR model provides a means for predicting the potential receptor binding affinity of any chemical relative to endogenous steroids given the 3D structure of the untested chemical. After sufficient validation, such a model could serve as a rapid screen for potential endocrine disrupters and be used for setting testing priorities, i.e. by identifying chemicals most likely to compete with endogenous steroid ligands.

Major advances are being realized in information and computational fields that could eventually lead to improved SAR models for toxicity prediction. Advances in neural networks, artificial intelligence, molecular visualization and modeling all have the potential to generate previously undiscovered models from existing data. However, these models will be subject to the same biology-imposed constraints as previously discussed, and share many of the same limitations as current methods. For example, the major disadvantage of current neural network-based SARs for toxicity prediction is that the model cannot be easily interpreted in terms of the original molecular parameters and, hence, the scientific basis for the NN model is practically undecipherable, making it difficult to scientifically rationalize a model prediction or define the bounds of application of the model.

CONCLUSIONS

Improvement in current SAR models used in TSCA chemical screening will be achieved most effectively by close interaction and feed-back between SAR application and toxicity prediction, laboratory testing, validation, and research into chemical mechanisms of toxicity. SAR is an extremely multidisciplinary field, applicable to a wide range of problems and endpoints. Since SAR modelers often lack expertise in toxicology, and toxicologists tend to be unfamiliar with the tools and assumptions of SAR modelers, there is a need for increased interaction, collaboration, and education between these two groups. The SAR modeler can guide the toxicologist in choosing experimental measures of toxicity, appropriate chemicals for SAR model design, and, in cases where a preliminary SAR model exists, approximate dose ranges to test for an effect. The toxicologist can provide the SAR modeler with insight into possible modes-of-action, practical and experimental design constraints (i.e. a reality check), and sources of uncertainty and error in the data.

The EPA/SAT is the regulatory arm that bears primary responsibility for the development and application of SAR to TSCA chemical screening, and the SAR expertise, models, and data used by

the SAT represent an extremely valuable resource for serving the interests of health and environmental protection. However, the SAT operates in relative isolation from the larger SAR community, and each could benefit from increased communication and collaboration. Although some outreach efforts have been made by the SAT, through development and dissemination of computerized SAR programs such as ECOSAR and OncoLogic, a major obstacle to increased collaboration is the confidential nature of the data used in SAR model development. Similarly, industry and other government regulatory agencies, such as FDA, often have large stores of toxicity data that are considered proprietary. A recent ECVAM workshop on "Integrated Use of Alternative Approaches for Predicting Toxic Hazard" produced the following recommendations (4):

- "Companies should be encouraged to make non-confidential data available to external groups, perhaps via an independent organisation such as ECVAM. For confidential data, they should be encouraged to review the need to maintain that confidentiality on a regular (continual) basis."
- "Regulatory agencies should be encouraged formally to establish (Q)SARs utilising submission data. . . . Companies should also be encouraged to develop (Q)SARs using their confidential data. Such (Q)SAR models should then be placed in the public domain, along with supporting non-confidential data."

While legitimate and defensible concerns of industry regarding the need for confidentiality of chemical structures and processes should not be minimized, there also should be greater acknowledgment of the value of available toxicity data, and recognition that more universal access to high quality toxicity data for SAR model development serves the best interests of the entire SAR community.

A number of issues have been identified that impact on the accuracy and utility of current SAR models for TSCA screening. To reiterate, the major trends likely to lead to the greatest improvement in SAR models for toxicity prediction are:

- greater understanding of mechanisms of toxicity for endpoints of potential concern;
- greater use of biologically-based chemical classifications in SAR development;
- better ways to represent molecules and their detailed biological interactions in SAR models;
- continued use of prospective validation and testing to validate and refine SAR models;
- greater understanding of role of metabolism in various forms of chemical toxicity;
- increased knowledge of role of biological receptors in toxicity and elucidation of the structure, function, and ligand requirements of relevant receptors;
- testing to fill crucial data gaps for chemical classes and toxicity endpoints of potential concern;
- greater effort to declassify some CBI and proprietary toxicity data that have little commercial value for use in SAR development;
- improved interaction between SAR users in industrial, academic and government research, and regulatory agencies to improve SAR models .

Screening the TSCA existing chemical inventory for all manner of potentially harmful effects in a timely manner is a huge challenge that cannot be met by testing alone. While testing deals with the generation of new data, SAR is above all the study of existing data and how to make best use of these data to predict the biological activity and properties of chemicals for which data are unavailable. SAR provides the only real alternative to expensive and time consuming laboratory testing. Hence, reliance on SAR methods will no doubt increase in response to increased budgetary and societal pressures to reduce costs and limit the use of animals in toxicity screening. While these methods hold great promise, the danger is that in response to such pressures SAR models will be invoked prema-

turely for some toxicity endpoints, will be extended beyond where they are likely to be valid or reliable, and will be used without sufficient oversight and testing verification. SAR offers a means for achieving better health and environmental protection by enabling a strategic and intelligent application of limited testing resources, by identifying the highest priority risk chemicals, and by attacking a much larger portion of the problem than is currently being addressed by testing alone. Better SAR models also have the clear potential to further reduce the need for testing or eliminate testing in some circumstances. SAR models will improve in tandem with increased understanding and availability of data upon which to base and refine such models. However, fulfilling the promise of SAR requires proper application of these methods, clear recognition of the limitations of SAR, and appreciation for the essential roles of research and strategic testing in SAR model development and refinement.

REFERENCES

1. Ashby, J., and Tennant, R. W., "Prediction of Rodent Carcinogenicity for 44 Chemicals: Results," *Mutagenesis* 9:7-15, 1994.
2. Auer, C. M., et al., "SAR – The U.S. Regulatory Perspective," *SAR and QSAR in Environmental Research* 2:29-38,1994.
3. Bahler, D., and Bristol, D. W., "The Induction of Rules for Predicting Chemical Carcinogenesis in Rodents," *Intelligent Systems for Molecular Biology,* L. Hunter, et al. (eds.) (Menlo Park, CA: AAAI/MIT Press, 1993).
4. Barret, M. D., et al., "The Integrated Use of Alternative Approaches for Predicting Toxic Hazard. Report and Recommendations of ECVAM (European Centre for the Validation of Alternative Methods) Workshop," *Alternatives to Laboratory Animals,* in press.
5. Benigni, R., Andreoli, C., and Giuliani, A., "QSAR Models for Both Mutagenic Potency and Activity: Application to Nitroarenes and Aromatic Amines," *Environmental and Molecular Mutagenesis* 24:208-219, 1994.
6. Benigni, R., and Giuliani, A., "Quantitative Structure-Activity Relationship (QSAR) Studies in Genetic Toxicology: Mathematical Models and the 'Biological Activity' Term of the Relationship," *Mutation Research* 306:181-186, 1994.
7. Bradbury, S., "Quantitative Structure Activity Relationships and Ecological Risk Assessment: An Overview of Predictive Aquatic Toxicology Research," *Proceedings of Workshop on Screening Technologies (1995),* in press.
8. Clements, R. G., et al., "The Use of Quantitative Structure-Activity Relationships (QSARs) as Screening Tools in Environmental Assessment," *Environmental Toxicology and Risk Assessment - 2nd Volume,* ASTM STP 1216, J.W. Gorsuch et al. (eds.) (Philadelphia, PA: American Society for Testing and Materials, 1993.
9. CQSAR: Comprehensive Structure Activity Relationships Using Multivariate Information, Biobyte Corp., Claremont, CA.
10. ECOSAR: A Computer Program for Estimating the Ecotoxicity of Industrial Chemicals Based on Structure Activity Relationships, National Center for Environmental Publications and Information, U.S. Environmental Protection Agency, Cincinnati, OH.
11. Enslein, K., et al., "Computation of Developmental Toxicity Potential by QSTR Models in the TOPKAT Program, " *Animal Test Alternatives,* in press.
12. Hansch, C., "Quantitative Structure-Activity Relationships and the Unnamed Science," *Accounts of Chemical Research 26:147-153, 1993.*
13. Leo, A., and Weininger, D., "CLOGP: Medchem Software Release 3.5," *Medicinal Chemistry Project* (Claremont, CA: Pomona College, 1992).
14. MetabolExpert: Expert System for Predicting Metabolizes, CompuDrug, USA, Austin, TX.
15. OncoLogic: An Expert System that Evaluates the Carcinogenic Potential of Chemicals, LogiChem, Inc., Boyertown, PA.
16. Richard, A. M., "Application of SAR Methods

to Non-Congeneric Data Bases Associated with Carcinogenicity and Mutagenicity: Issues and Approaches," *Mutation Research* 305:73-97, *1994.*

17. Richard, A. M., "Role of Computational Chemistry in Hazard ID: Mechanism-based SARs," *Proceedings of ATSDR sponsored Workshop on Screening Technologies (1995),* in press.
18. TOPKAT: Toxicity Prediction Software, Health Designs Incorporated, Rochester, NY, 14604
19. U.S. Environmental Protection Agency/ European Community, *Joint Project on the Evaluation of (Quantitative) Structure-Activity Relationships: Final Report,* July, 1993.
20. Wagner, P.J., Nabholz, V. and Kent, R.J., "The New Chemicals Process at EPA: Structure Activity Relationships for Hazard Identification and Risk Assessment," *Proceedings of ATSDR sponsored Workshop on Screening Technologies (1995),* in press.
21. Wailer, C.L., Minor, D. L., and McKinney, J. D., "Examination of the Estrogen Receptor Binding Affinities of Phenyl-substituted Hydrocarbons of Environmental Concern Using Three-dimensional Quantitative Structure-activity Relationships," *Environmental Health Perspectives,* in press.
22. Woo, Y.-T., et al., "Development of Structure Activity Relationship Rules for Predicting Carcinogenic Potential of Chemicals," *Proceedings of ATSDR sponsored Workshop on Screening Technologies (1995),* in press.
23. Zeeman, M., Nabholz, J. V., and Clements, R. G., "The Development of SAR/QSAR for Use Under EPA's Toxic Substances Control Act (TSCA): An introduction" *Environmental Toxicology and Risk Assessment - 2nd Volume,* ASTM STP 1216, J.W. Gorsuch et al. (eds.) (Philadelphia, PA: American Society for Testing and Materials, 1993).

13

Predictive Methods for Chemical Fate

John D. Walker

ABSTRACT: *A strategy is proposed to predict physical properties, partitioning, bioconcentration and degradability of nonionized, nonpolymeric organic chemicals. The strategy uses structural features of organic chemicals, structure activity relationships, sequenced estimation techniques (including recently developed computerized methods), and laboratory screening tests. The cost to use this strategy is compared with the cost of conducting a standard menu of tests to measure physical properties, partitioning, bioconcentration and degradability.*

This chapter proposes a strategy for using structural features of organic chemicals, structure activity relationships, sequenced estimation techniques (including some recently developed computerized methods) and laboratory screening tests to predict the chemical fate of nonionized, nonpolymeric organic chemicals (figure 13-1). It emphasizes the need to consider computerized methods to make these predictions, especially as these methods relate to the "cost and time considerations" emphasized by the April 24-25, 1995, Congressional Office of Technology Assessment (OTA) "Workshop on Testing and Screening Technologies for Review of Chemicals in Commerce. "

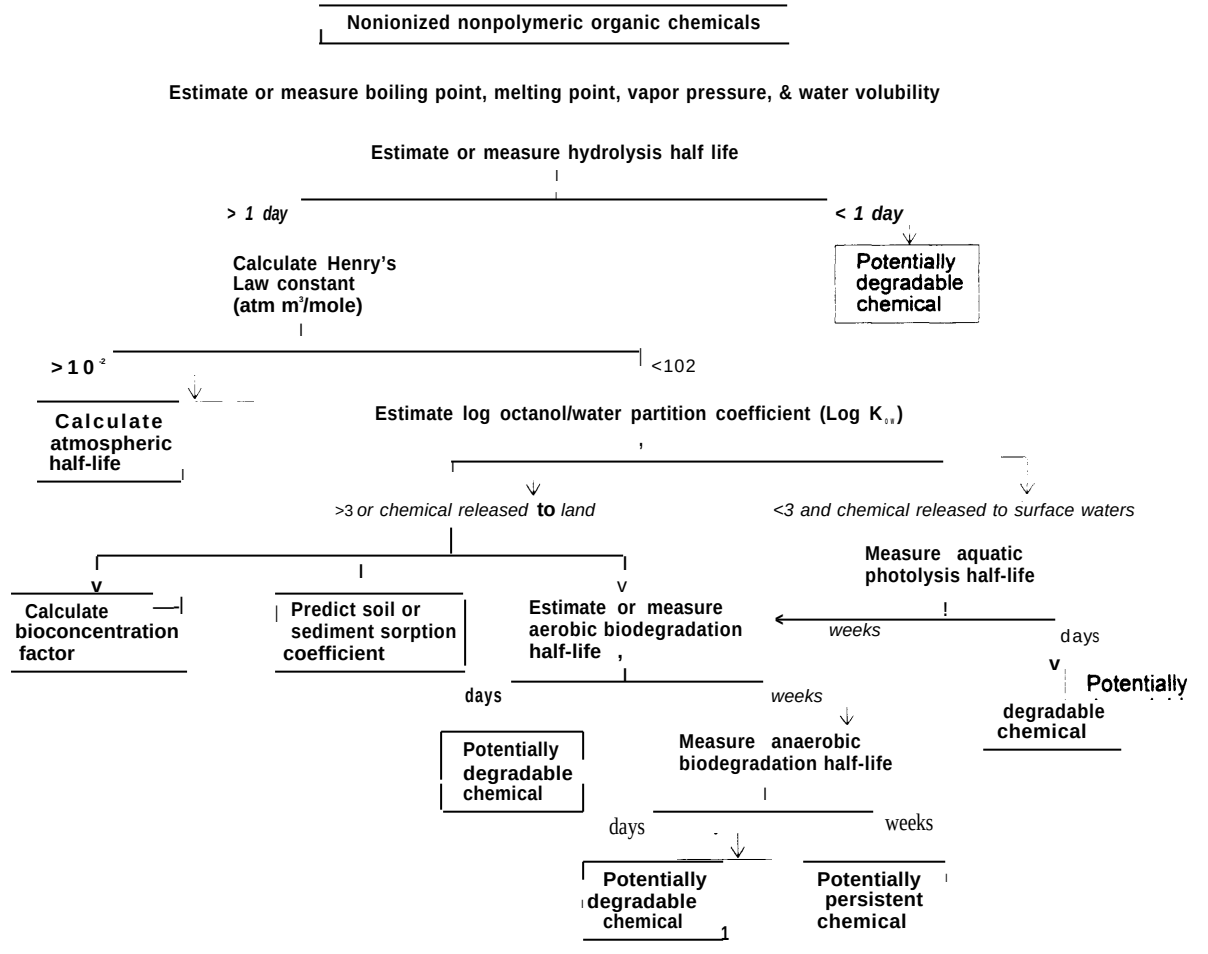
Many of the computerized estimation methods discussed in this chapter have been validated with large data sets of chemicals that were not used to derive the methods, e.g., the computerized methods developed by Meylan and Howard for estimating aerobic biodegradation, boiling point, Henry's law constant, octanol-water partition coefficient, soil or sediment sorption coefficient, vapor pressure and water volubility are quite accurate and certainly accurate enough for the proposed strategy. While laboratory measurements may be preferred, it should be noted that these measurements cost time and money and for many chemicals, the resources to make these measurements may not be warranted, especially with the availability of computer software and mathematical models that can be used to accurately predict the data that would be obtained from measurements.

The strategy for making chemical fate predictions for physical properties, partitioning, bioconcentration and degradability uses a number of decision criteria to progress through these predictions. While all these decision criteria are supported by data, there is still need for professional judgment when progressing through the predictions. The initial predictions use estimation techniques, structure activity relationships and computerized methods to predict physical properties, partitioning, bioconcentration and degradability. The final predictions also use these methods, but may also require some laboratory screening tests to predict potential degradability.

The proposed strategy as well as the schemes used to implement ecological effects testing under the Toxic Substances Control Act (TSCA) are different than the chemical testing approaches taken by the Organisation for Economic Cooperation and Development (OECD) and others. The OECD's Screening Information Data Set (SIDS) program for example, requires that a standard menu of chemical fate, ecological effects and health effects tests be conducted for all chemicals. Clearly there are advantages and

The author acknowledges the review **and** comments from Bob Boethling of EPA and Bill Meylan of Syracuse Research Corporation. The contents of this chapter do not necessarily reflect the views and policies of the Environmental Protection Agency.

Figure 13-1: Proposed Strategy to Identify Persistent Organic Chemicals



disadvantages to both approaches, e.g. an advantage of the SIDS approach is that all the data are available for all chemicals should it be necessary to make comparative assessments related to pollution prevention, use of substitutes, risk reduction, etc. To illustrate the economic aspects of these different approaches, the cost of developing data for individual predictions in the proposed strategy is compared to the cost of conducting a base set or standard menu of tests to measure physical properties, partitioning, bioconcentration and degradability (table 13-1).

In response to the request from OTA, the chapter focuses on TSCA, the cost and time considerations for screening chemicals, and the organizations that are required to recommend, implement and conduct testing under TSCA. These organizations include the TSCA Interagency Testing Committee (ITC), EPA's new and existing TSCA chemical testing programs and the manufacturers and processors of TSCA chemicals. The proposed strategy was previously presented at an American Society for Testing and Materials Symposium on Environmental Toxicology and Risk Assessment (27).

CHEMICAL TESTING UNDER TSCA

Three sections of TSCA relate to chemical testing, *viz.*, sections 2, 4 and 5.

Section 2 of TSCA states that the U.S. Congress finds that humans and the environment are exposed to large numbers of chemicals, and that

Table 13-1: Cost of Laboratory Tests for Basic Physical Properties

Test	Cost ($)a
Boiling point	420-560
Melting point	430-565
Vapor pressure	1,360-2,520
Water volubility	2,870-14,080
Physical property data	5,080-17,725
Octanol-water partition coefficient	1,310-2,370
Soil sorption Coefficient	6,680-9,430
Volatilization	6,180-8,800
Partitioning data	14,170-20,600
Hydrolysis	5,470-7,790
Photolysis	11,880-16,910
Aerobic biodegradation	8,250-11,800
Anaerobic biodegradation	5,980-8,480
Degradability data	31,580-44,980
Bioconcentration data	15,580-41,690
TOTAL	66,410-124,995

a Cost estimatesfrom Mathtech (1995)

some of those chemicals may present an unreasonable risk of injury to health or the environment. Section 2 states that it is the policy of the United States that "adequate data should be developed with respect to the effect of chemical substances and mixtures on health and the environment and that the development of such data should be the responsibility of those who manufacture and those who process such chemical substances and mixtures. "

Section 4 requires the testing of existing chemical substances and mixtures ("chemicals"). Existing chemicals are those chemicals that were produced in or imported into the United States before TSCA became effective in 1977 or those chemicals for which notice to commence production were issued after 1977 by EPA under TSCA section 5. Section 4 of TSCA was enacted by Congress in response to concerns that the effects of existing chemical substances and mixtures on human health and the environment were inadequately documented and understood. To alleviate these concerns, Congress created the ITC to screen, prioritize and recommend existing chemicals for testing to the EPA Administrator and empowered the EPA Administrator with author ity to require that manufacturers or processors test their chemicals to develop adequate data. These data are used by EPA, other U.S. Government, Foreign Government and International Organizations as well as state and local governments to develop hazard and exposure assessments that are necessary to promote pollution prevention or chemical regulation (27).

Section 4(a) ensures that existing chemicals which may present an "unreasonable risk" to human health or the environment, or may involve substantial production or exposure, receive priority testing consideration and that manufacturers or processors of these chemicals test them to assure that adequate data are developed to assess their potential risk to humans or the environment. Section 4(a) requires that the EPA Administrator make three findings before requiring the manufacturers or processors of a chemical to conduct testing. Historical details of these findings are discussed in EPA's first two TSCA section 4(a) test rules (20, 21). Section 4(b) requires that EPA publish standards for development of test data and review the standards at least every 12 months. Section 4(c) allows manufactures and processors of chemicals to apply for exemptions for testing under section 4(a). Section 4(d) requires EPA to publish in the *Federal Register* the receipt of any data developed under section 4(a) within 15 days of its receipt.

Section 4(e) describes the statutory responsibilities of the ITC. The ITC is described in a previous publication (27) as is an ITC case study that provides all of the milestones from the time of ITC's initial work until EPA completes its risk assessment (30). Section 4(f) requires EPA to prevent or reduce risks to substances that present a significant risk of serious or widespread harm from cancer, gene mutations or birth defects. Section 4(g) allows manufactures to petition the EPA to prescribe standards for the development of data. Section 5 requires manufactures that want to produce new chemicals to submit premanufacturing notices to the EPA before initiating commercial production. New chemicals are those that were not produced in or imported

into the United States before 1977. EPA has 90 days to approve a premanufacturing notice. EPA can approve these notices with no contingencies, with contingencies for chemical testing or pollution prevention or not approve them, thereby banning the chemical from production. EPA approves a premanufacturing notice by issuing a commencement notice. Those chemicals for which commencement notices are issued and for which commercial production is initiated become existing chemicals and are subject to the requirements of TSCA section 4. Additional details about the TSCA section 5 process have been recently published (16).

TSCA ENVIRONMENTAL TESTING SCHEMES

Two testing schemes have been used to implement environmental testing under TSCA sections 4 and 5. Both are used to implement ecological effects testing. Details of the ecological effects testing scheme used for existing chemicals under TSCA Section 4 have been described in detail previously (23, 24). The TSCA section 4 ecological effects testing scheme uses physical property, partitioning and degradability data to develop predicted environmental concentrations (PECs) of chemicals and then uses these PECs as one factor to determine whether ecological effects testing should be considered.

Details of the ecological effects testing scheme used for new chemicals under TSCA section 5 have been described in a series of papers published in the proceedings of a 1991 American Society for Testing and Materials Symposium (2, 16, 18, 33). The TSCA section 5 scheme uses ecological effects decision criteria or "concern levels" to determine whether higher tier tests should be conducted. The TSCA section 5 scheme relies on structure activity relationships, because of the scarcity of data associated with new chemicals.

The proposed strategy described in this chapter is designed to complement both the TSCA section 4 and section 5 ecological effects testing schemes. This strategy focuses on chemical fate structure activity relationships, predictions and testing to provide information that can be integrated with that generated from the ecological effects testing schemes to estimate environmental risk, promote pollution prevention, etc.

THE PROPOSED STRATEGY

The proposed strategy uses structural features of organic chemicals, structure activity relationships, sequenced estimation techniques (including some recently developed computerized methods) and laboratory screening tests to predict physical properties, partitioning, bioconcentration and degradability of nonionized, nonpolymeric organic chemicals.

Structural Features and Structure Activity Relationships

Structural features and structure activity relationships have been used for many years to develop mathematical models that can be used predict the fate of chemicals (5). It is well known, for example, that chemicals with certain structural features (described below) will be more susceptible to hydrolysis than chemicals not containing those features. Similarly, it is well known that structural features related to molecular topology can be used to estimate physical properties, partition coefficients, bioconcentration potential and degradability. Structural features and structure activity relationships have been used to develop the estimation techniques and computerized methods described below.

Cost Comparisons for Estimation Techniques and Screening Tests

Laboratory tests that provide a standard menu of basic physical chemical property, partitioning, bioconcentration and degradability data are listed in table 13-1. If all of these tests were conducted, the cost could range from $66,410 to $124,995 U.S. dollars (table 13-1). If empirical data are not required, then estimation techniques (including some recently developed computerized methods) can be used to estimate some of these data at substantial time and cost savings compared to laboratory measurements.

As explained below, some of the cost ranges for tests to develop the physical properties and partitioning data listed in table 13-1 are large because they were estimated for more than one method. Costs for conducting hydrolysis, photolysis, aerobic biodegradation, anaerobic biodegradation and fish bioconcentration tests are all estimates for one method and the size of the cost range is usually a reflection of the cost of the analytical method used to develop the data.

The proposed strategy described in figure 1 does not require all the tests listed in table 13-1, but suggests that it may be possible to sequentially use estimation techniques (including some recently developed computerized methods) or screening tests to estimate physical properties, partitioning, bioconcentration and degradability of nonionized, nonpolymeric organic chemicals. Based on the type of organic chemical that is being evaluated, the sequencing allows the minimum predictions to be made or the minimum number of screening tests to be conducted. For example, if there were no data for a chemical, the structure suggested that it would be susceptible to hydrolysis and a decision was made to develop empirical data *in lieu* of making predictions, then it should cost from $2,210 to $3,645 to conduct boiling point, melting point and vapor pressure tests and $5,470 to $7,790 to conduct a hydrolysis test (table 13-1). The water volubility test would not be conducted, until after the results from hydrolysis were available. If the hydrolysis half life was > 1 day, then water volubility should be measured at a cost of $2,870 to $14,080 (table 13-1). If the hydrolysis half life was < 1 day, the total cost of conducting chemical fate testing for this chemical would range from $10,550 to $25,515 as opposed to $66,410 to $124,995 for conducting all the chemical fate tests listed in table 13-1. The methods in table 13-1 for which costs have been estimated are all TSCA test methods that have been harmonized with OECD and others. These methods (listed below) are described in Volume 40 of the Code of Federal Regulations (CFR), parts 795, 796 and 797. The methods that have been used to develop data for existing chemicals recommended by the ITC to measure physical-chemical properties and environmental persistence have been described (25) as have those to measure bioconcentration (28, 29). The TSCA section 4 chemical fate and ecological effects test data developed from using those methods have been published (26, 31, 32).

Physical Properties

Estimated costs for conducting boiling point, melting point, vapor pressure and water volubility tests range from $5,080 to $17,725 (table 13-1). The cost range estimated for vapor pressure testing ($1,360 to $2,520) is actually from two estimates for measuring vapor pressure by isoteniscope ($1,360 to $1,900) or by gas saturation ($1,790 to $2,520). The large cost range estimated for water volubility testing ($2,870 to $14,080) is a composite of three estimates for measuring water volubility by an analytical method after centrifugation ($2,870 to $4,070), by high pressure liquid chromatography after column generation ($7,740 to $1 1,080) or by gas chromatography after column generation ($9,840 to $14,080). Lyman et al. (5) describe methods for estimating boiling point, vapor pressure and water volubility. Recently, methods for personal computers have been developed to estimate melting point, boiling point and vapor pressure (11) and water volubility (12). If reliable measured data for these physical properties are not available for a chemical (especially vapor pressure and water volubility data at 20°C or 25°C), they should all be predicted using the computerized methods listed above, unless as noted above the structure suggests that the chemical would be susceptible to hydrolysis.

Partitioning, Bioconcentration and Degradability

Estimated costs for measuring octanol-water partition coefficient, soil or sediment sorption coefficient and volatilization range from $14,170 to $20,600 (table 13-1). The cost range estimated for measuring octanol-water partition coefficient is a composite of two estimates by liquid chromatography ($1,310 to $1,730) or by generator column ($1,820 to $2,370). Lyman et al. (5) de-

scribe methods for estimating octanol-water partition coefficient, soil or sediment sorption coefficient and volatilization. Recently, methods for personal computers have been developed to estimate octanol-water partition coefficients (13) and soil or sediment sorption coefficients (10).

- **Hydrolysis.** Hydrolysis is a chemical fate process in which an organic chemical, reacts with water resulting in the cleavage of a carbon-X bond and the formation a new carbon-oxygen bond. X could be represented by structural features (groups) such as alkyl halides, epoxides, esters, hydrazines, isocyanates, nitriles, silanes, etc. Hydrolysis is probably the first indicator of degradability that should be estimated, because almost all chemicals entering the environment contact water (fig. 13-1). If a chemical contains a structural feature that is susceptible to hydrolysis, then hydrolysis should be estimated or measured. Hydrolysis half life can be estimated using the method of Mill et al. (15) or the personal computer methods developed by Meylan and Howard (8). If empirical data are required, then hydrolysis can be measured using the methods in 40 CFR 796.3500 at a cost of $5,470 to $7,790 (table 13-1). If the half life is < 1 day, then the chemical is likely to degrade in aqueous systems (figure 13-1). If the hydrolysis half life is > 1 day, then the Henry's Law Constant should be estimated (figure 13-1). A one day half life was selected as the decision criteria based on an analysis of the data of Maybe and Mill (6). Factors that may influence hydrolysis rates include the pH of the receiving aqueous system (especially if hydrolysis is acid or based catalyzed), temperature of the receiving aquatic system (hydrolysis increases with temperature), concentration of a chemical entering an aqueous system (at high concentrations some chemicals will polymerize before they can hydrolyze) and potential of a chemical to sorb to sediment, suspended sediment or soil (especially if the chemical is released to landfills). Sorption can be estimated as described below.

. **Henry's Law Constant.** Henry's Law Constant is the vapor pressure (in atm) divided by water volubility (in moles/m^3) of a chemical at one temperature and physical state; it provides information on air-water partitioning. The Henry's Law Constant can be calculated using the personal computer method of Meylan and Howard (9). If measured or estimated vapor pressure and water volubility data are not available to calculate Henry's Law Constants, then they can be estimated or measured using the methods described above. A chemical with a Henry's Law Constant > 10^{-2} atm m^3/mole should partition to air and its half life in the atmosphere should be calculated (Fig. 1). If the Henry's Law Constant is < 10^{-2} atm m^3/mole, it should partition to water and its log octanol-water partition coefficient (log K_{ow}) should be estimated (figure 13-1).

. A Henry's Law Constant of 10-3 atm m^3/mole is usually used to describe highly volatile chemicals (29) and is a requirement if volatilization is measured using the method described below. A Henry's Law Constant of 10^{-2} atm m^3/mole was selected as a more conservative decision criteria than 10^{-3} atm m^3/mole, because there are likely to be fewer chemicals with a Henry's Law Constant of 10^{-2} atm m^3/mole that could also partition to sediment, suspended sediment or soil if they could not rapidly escape from aqueous systems. Chemicals such as chloroethane, chloropropane and dichloroethylene with Henry's Law Constants >10^{-2} atm m^3/mole and water solubilities >= 1 mole/m^3 and log K_{ow} values < 2 are likely to rapidly escape from aqueous systems before they are sorbed. In contrast a chemical such as octane with a Henry's Law Constant > 10^{-2} atm m^3/mole, water solubility value < 10^{-2} mole/m^3 and log K_{ow} values > 2 may sorb to sediments or soil before significant quantities can escape from aqueous systems. If there uncertainties regarding the influence of sediment, suspended sediment or soil on air-water partitioning, then the log K_{ow} should be estimated as described below and be used to estimate the influence of sorption on air-water partitioning. In general, chemicals with Henry's Law Constants > 10^{-2} atm m^3/mole and log K_{ow} values <2 should almost totally partition to air.

. **Atmospheric half life.** Atmospheric half life is an estimate of the degradation time of chemicals that partition to air and react with hydroxyl radicals or ozone. Atmospheric half life can be calculated using the personal computer method of Meylan and Howard (lo).

. **Octanol-water partition coefficient (K_{ow}).** K_{ow} is used to estimate a chemical's potential to partition between water and octanol (as a surrogate for lipids) and to bioconcentrate in fatty tissues of aquatic organisms. Log K_{ow} can be estimated using the personal computer method of Meylan and Howard (13) or measured using the methods in 40 CFR 796.1570 or 1720 at a cost of $1,310 to **$2,370** (table 13-1). If the log K_{ow} >3 or if the chemical is released to land, a bioconcentration factor should be calculated, a soil or sediment sorption coefficient should be predicted and aerobic biodegradation half life should be estimated or measured (figure 13-1). If the log K_{ow} < 3 and the chemical is released to surface waters, then aquatic photolysis should be measured (figure 13-1).

. **Bioconcentration factor.** The bioconcentration factor is used to estimate the potential for chemicals to concentrate in aquatic organisms after uptake from water. The bioconcentration factor can be calculated using the method of Veith et al. (22) or measured using the methods in 40 CFR 797.1520 at a cost of $15,580-$41,690 (table 13-1). If the estimated bioconcentration factor is > 10,000, some consideration should be given to measuring the bioconcentration factor to determine if metabolism significantly decreases actual bioconcentration.

. **Soil or sediment sorption coefficient.** The soil or sediment sorption coefficient is used to estimate the potential of a chemical to partition between water and the organic fraction of soil or sediment. The soil or sediment sorption coefficient can be predicted using the personal computer method of Meylan et al. (10) or measured using the method in 40 CFR 796.2750 at a cost of $6,680 to $9,430 (table 13-1).

. **Aquatic photolysis half life** Estimates of aquatic photolysis half lives predict the potential of chemicals released to surface waters to degrade directly in sunlight or indirectly in sunlight or sunlight in the presence of humic acids. Chemicals that degrade directly must contain a structural feature that absorbs sunlight. Aquatic photolysis can be measured using the methods in 40 CFR 795.70 at a cost of $11,880-$16,910 (table 13-1). If the aquatic photolysis half life is a few days, then the chemical is likely to degrade (figure 13-1). However, if the aquatic photolysis half life is a few weeks, then aerobic biodegradation half life should be estimated or measured (figure 13-1).

. **Aerobic biodegradation half life.** Aerobic biodegradation is used to estimate the degradation of chemicals by microorganisms in oxygenated water, sediment, sludge, soil, etc. Aerobic biodegradation half life can be predicted using the method of Boethling et al. (l), Gombar and Enslein (3), Klopman et al. (4), Niemi et al. (17) or Tabak and Govind (19) or estimated from the measured values obtained from using the method in 40 CFR 796.3100 at a cost of $8,250-$11,800 (table 13-1). If the predicted or estimated aerobic biodegradation half life is a few days, then the chemical is likely to degrade (figure 13-1). However, if the aerobic biodegradation half life is a few weeks, then anaerobic biodegradation half life should be measured (figure 13-1).

. **Anaerobic biodegradation half life.** Anaerobic biodegradation is used to estimate the degradation of chemicals by microorganisms in anoxic water, sediment, sludge, soil, etc. Anaerobic biodegradation half life can be estimated from the measured values obtained from using the method in 40 CFR 796.3140 at a cost of $5,980-$8,480 (table 13-1). If anaerobic biodegradation half life is a few days, then the chemical is likely to degrade (figure 13-1). However, if the anaerobic biodegradation half life is a few weeks, then the chemical is likely to persist (figure 13-1).

. **Estimation techniques that need to be developed.** Aquatic photolysis, anaerobic biodegradation, oxidation and reduction are important environmental processes that deserve

some consideration for development of validated computerized estimation techniques. For example, it is known that certain chemicals (multi-halogenated aliphatic hydrocarbons) undergo reductive dehalogenation by anaerobic bacteria while other chemicals (nitroaromatics) undergo abiotic reduction in anaerobic environments. Generating rules that could be used as in computer programs to reliably predict aquatic photolysis, anaerobic biodegradation, oxidation and reduction would considerably reduce the time and money needed to generate measured data.

Hypothetical Case Studies

Chemical A. Chemical A is a short-chain phthalate ester that is likely to be released to surface waters. Reliable physical property data are available and do not have to be measured or predicted. The measured hydrolysis half life was > year, the calculated Henry's Law Constant = 1 x 10^{-7} atm m^3/mole, the estimated log K_{ow} = 2, the measured aquatic photolysis half life was > year and the estimated aerobic biodegradation half life was 2 days. It cost $17,350 to $24,700 to develop these data, a savings of about $49,060 to $100,295, when compared to the cost of a standard menu of chemical fate tests (table 13-1). These data suggest that if the chemical were released to surface waters, that it would partition to water and be rapidly biodegraded.

Chemical B. Chemical B is a branched chain alcohol that is likely to be released to surface waters. No reliable physical property data were available and a decision was made to develop measured data at a cost of $5,080 to $17,725 (table 13-1). The estimated hydrolysis half life was > year, the calculated Henry's Law Constant 1 x 10^{-9} atm m^3/mole, the estimated log K_{ow} = -1, the measured aquatic photolysis half life was > year and the measured aerobic biodegradation half life was 0.5 days. It cost $20,130 to $28,710 to develop these data, a savings of about $46,280 to $96,285, when compared to the cost of a standard menu of chemical fate tests (table 13-1). These data suggest that if the chemical were released to surface waters, it would partition to water and be very rapidly biodegraded.

Chemical C. Chemical C is a halogenated silane. Measured boiling point, melting point and vapor pressure were available and did not have to be measured or predicted. The measured hydrolysis half life was 3 minutes. It cost $5,470 to $7,790 to develop these data, a savings of about $60,440 to $117,205, when compared to the cost of a standard menu of chemical fate tests (table 13-1). These data suggest that if the chemical were to contact water it would be rapidly degraded.

Chemical D. Chemical D is a phenol with branched hydrocarbon chain that is likely to be released to land. No reliable physical property data were available and a decision was made to predict the physical properties listed in table 13-1 using the personal computer methods of Meylan and Howard (1 1, 12). The measured hydrolysis half life was > year, the calculated Henry's Law Constant 1 x 10^{-5} atm m^3/mole, the estimated log K_{ow} = 5, the measured fish bioconcentration factor = 347, the predicted sediment sorption coefficient was > 5,000, the estimated aerobic biodegradation half life was 3 weeks and the measured anaerobic biodegradation half life was 7 months. It cost $39,380 to $57,960 to develop these data, a savings of about $27,030 to $67,035, when compared to the cost of a standard menu of chemical fate tests (table 13-1). These data suggest that if the chemical were released to land or water, it would partition to water, soil and sediment, be available for bioconcentration by aquatic organisms, be tightly bound to soil and sediment and persist in aerobic and anaerobic environments.

REFERENCES

1. Boethling, R. S., et al., "Group contribution method for predicting probability and rate of aerobic biodegradation," *Environmental Science and Technology* 28;459-465, *1994.*
2. Clements, R. G., et al.,'' The Use and Application of QSARs in the Office of Toxic Substances for Ecological hazard Assessment of New Chemicals," *Environmental Toxicology and Risk Assessment: First Volume,* ASTM STP 1179, W.G. Landis and J. Hughes (eds.)

(Philadelphia, PA: American Society for Testing and Materials, 1993).

3. Gombar, V. K., and K. Enslein, K., "A Structure-Biodegradability Relationship Model by Discriminant Analysis," *Applied Multivariate Analysis in SAR and Environmental Studies*, J. Devillers and W. Karcher (eds.) (Boston, MA: Kluwer, 1991).
4. Klopman, G., Balthasar, D.M., and Rosenkranz, H. S., "Application of the Computer-automated Structure Evaluation (CASE) Program to the Study of Structure-biodegradation Relationships of Miscellaneous Chemicals," *Environmental Toxicology and Chemistry* 12:231-240, *1993*.
5. Lyman, W.J., Reehl, W. F., and Rosenblatt, D. H., *Handbook of Chemical Property Estimation Methods (New* York, NY: McGraw-Hill, 1982).
6. Maybe, W., and Mill, T., "Critical Review of Hydrolysis of Organic Compounds in Water Under Environmental Conditions. *Journal of Physical Chemistry* 7:383-415, 1978.
7. Mathtech, Inc., *TSCA Test Guidelines Cost Estimates*, prepared under contract for the U.S. Environmental Protection Agency, Office of Pollution, Prevention and Toxics' Economics, Exposure and Technology Division, February 1995.
8. Meylan, W. M., and Howard, P. H., *Users guide for HYDRO, PC Software to Estimate Aqueous Hydrolysis Rates* (Syracuse NY: Syracuse Research Corporation, 1991).
9. Meylan, W. M., and Howard, P. H., "Bond contribution method for estimating Henry's law constants," *Environmental Toxicity and Chemistry* 10:1283-1293, *1991*.
10. Meylan, W.M. and P.H. Howard, P.H., "Computer Estimation of the Atmospheric Gas-phase Reaction Rate of Organic Compounds with Hydroxyl Radicals and Ozone," *Chemosphere* 26:2293-2299, *1993*.
11 Meylan, W. M., and Howard, P. H., *Users guide for MPBPVP, Melting Point, Boiling Point, Vapor Pressure Estimation Program for Microsoft Windows 3.1* (Syracuse NY: Syracuse Research Corporation, 1994).
12. Meylan, W.M., and Howard, P. H., *Users guide for WS-KOW, Water Solubility from Log K_{ow} Program for Microsoft Windows 3.1* (Syracuse, NY: Syracuse Research Corporation, 1995).
13. Meylan, W. M., and Howard, P. H., "Atom/fragment Contribution Method for Estimating Octanol-water Partition Coefficients," *Journal of Pharmcological Science* 84:83-92, *1995*.
14. Meylan, W. M., and Howard, P. H., and Boethling, R. S., "Molecular Topology/fragment Contribution Method for Predicting Soil Sorption Coefficients," *Environmental Science Technology* 26:1560-1567, *1992*.
15. Mill, T., et al., *Environmental fate and exposure studies development of a PC-SAR for hydrolysis: esters, alkyl halides and epoxides*, EPA Contract No. 68-02-4254, SRI International, 1987.
16. Nabholtz, J. V., Miller, P., and Zeeman, M., "Environmental Risk Assessment of New Chemicals under the Toxic Substances Control Act (TSCA) Section 5," *Environmental Toxicology and Risk Assessment: First Volume*, ASTM STP 1179, W.G. Landis and J. Hughes (eds.) (Philadelphia, PA: American Society for Testing and Materials, 1993).
17. Niemi, G. J., et al., "Structural Features Associated with Degradable and Persistent Chemicals," *Environmental Toxicology and Chemistry* 6:515-527, *1987*.
18. Smrchek, J.R., et al., "Assessing ecological hazard under TSCA: Methods and Evaluation of Data," *Environmental Toxicology and Risk Assessment: First Volume*, ASTM STP 1179, W.G. Landis and J. Hughes (eds.) (Philadelphia, PA: American Society for Testing and Materials, 1993).
19. Tabak, H.H., and Govind, R., "Prediction of Biodegradation Kinetics Using a Nonlinear Group Contribution Method," *Environmental Toxicology and Chemistry* 12:251-260, 1993.
20. U.S. Environmental Protection Agency, "Chloromethane and Chlorinated Benzenes," proposed rule, *Federal Register 45:48524-48566, 1980*.

21. U.S. Environmental Protection Agency, "Dichloromethane, Nitrobenzene and 1,1,1 - Trichloroethane," proposed test rule, *Federal Register* 46:30300-30320, *1981*.
22. Veith, G.D., DeFoe, D. L., and Bergsdedt, B. V., "Measuring and Estimating Bioconcentration Factor of Chemicals in Fish," *Journal of Fisheries Resources Board Can. 36:1040-1048, 1979.*
23. Walker, J. D., "Bioconcentration, Chemical Fate and Environmental Effects Testing Under Section 4 of the Toxic Substances Control Act," *Toxicity Assessment* 5:61-75.
24. Walker, J. D., "Bioconcentration, Chemical Fate and Aquatic Toxicity Testing under the Toxic Substances Control Act: Proposed Testing and Decision Criteria," *Toxicity Assessment* 5:103-134, *1990.*
25. Walker, J. D., "Review of Chemical Fate Testing Conducted Under Section 4 of the Toxic Substances Control Act: Chemicals, Tests and Methods," *Aquatic Toxicology and Risk Assessment: Thirteenth Volume,* ASTM STP 1096, W.G. Landis and W.H. Vanderschaile (eds.) (Philadelphia, PA: American Society for Testing and Materials, 1990).
26. Walker, J. D., "Acrylamide Aquatic Effects: Potential Impact of Extended Exposure" *Environmental Toxicology and Water Quality* 6:363-369, *1991*.
27. Walker, J. D., "Can Chemical Structures and Physical Properties be used to Define Required Chemical Fate and Ecological Effects Tests?" *3rd American Society for Testing and Materials Symposium on Environmental Toxicology and Risk Assessment* (Atlanta, Georgia, 1993).
28. Walker, J. D., "Review of Ecological Effects and Bioconcentration Testing Recommended by the TSCA Interagency Testing Committee and Implemented by EPA under the Toxic Substances Control Act: Chemicals, Tests and Methods," *Environmental Toxicology and Risk Assessment,* ASTM STP 1179, W.G. Landis, J.S. Hughes and M.A. Lewis (eds.) (Philadelphia, PA: ASTM, 1993).
29. Walker, J. D., "The TSCA Interagency Testing Committee's Role in Facilitating Development of Test Methods: Toxicity and Bioconcentration Testing of Chemicals Added to Sediment," *Environmental Toxicology and Risk Assessment: Second Volume,* ASTM STP 1216, J.W. Gorsuch et al. (eds.) (Philadelphia, PA: American Society for Testing and Materials, 1993).
30. Walker, J.D., and Smock, W. H., "Chemicals Recommended for Testing by the TSCA Interagency Testing Committee: A Case Study with Octamethylcyclotetrasiloxane," *Environmental Toxicology and Chemistry,* in press.
31. Walker, J. D., "Recommendations of the TSCA Interagency Testing Committee: Aquatic Toxicity, Bioconcentration and Chemical fate Data Developed Under Section 4 of the Toxic Substances Control Act," *Fundamentals of Aquatic Toxicology II,* G.M. Rand (cd.) (Washington, DC: Taylor & Francis, 1995).
32. Walker, J. D., "Testing Decisions of the TSCA Interagency Testing Committee for Chemicals on Canada's Domestic Substances List and Priority Substances List: Di- *tert* -butylphenol, ethyl benzene, brominated flame retardants, phthalate esters, chloroparaffins, chlorinated benzenes and anilines," *Environmental Toxicology and Risk Assessment: Fourth Volume,* ASTM STP 1241, T.W. LaPoint, F.T. Price, and E.E. Little (eds.) (Philadelphia, PA: American Society for Testing and Materials, 1995).
33, Zeeman, M., and Gilford, J., "Ecological Hazard Evaluation and Risk Assessment Under EPA's Toxic Substances Control Act (TSCA): An Introduction," *Environmental Toxicology and Risk Assessment: First Volume,* ASTM STP 1179, W.G. Landis and J. Hughes (eds.) (Philadelphia, PA: American Society for Testing and Materials, 1993).