Report of the NTP Ad Hoc Panel on Chemical Carcinogenisis Testing and Evaluation

Board of Scientific Counselors
National Toxicology Program

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service

REPORT OF THE

AD HOC PANEL ON CHEMICAL CARCINOGENESIS TESTING AND EVALUATION

OF THE

NATIONAL TOXICOLOGY PROGRAM

BOARD OF SCIENTIFIC COUNSELORS

August 17, 1984

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The Ad Hoc Panel on Chemical Carcinogenesis Testing and Evaluation was charged by National Toxicology Program's (NTP) Board of Scientific Counselors to examine the current state of the science. The Panel has met with the NTP staff, considered the bioassay program and related short-term, subchronic and other tests, reviewed the literature and received oral and written presentations from interested colleagues. The Panel requested wide circulation of the draft report and considered the 50 or so formal responses as well as numerous informal contacts in preparing the final text.

The conclusions of the Panel are stated as a series of recommendations to the Board of Scientific Counselors. These deal with practical aspects of the bioassay as well as possible future studies extending methodology and resolving persistent questions. The Panel considers all of these to be relevant to the NTP mission.

In reviewing the bioassay and related studies, the Panel found that the on-going effort has benefited from the existing peer review process and was continuing to evolve scientifically, that the issue of quality assurance was gaining appropriate recognition and that many of the Panel's recommendations were in fact anticipated by the NTP managers.

NTP Board of Scientific Counselors'
Ad Hoc Panel on Chemical Carcinogenesis Testing and Evaluation

Dr. John Doull (Chairman)
Professor, Department of
Pharmacology and Toxicology
University of Kansas Medical Center
Kansas City, Kansas 66103
913/588-7140

Dr. Richard Adamson
Director, Division of Cancer Cause
and Prevention
National Cancer Institute
Building 31, Room 11A04
Bethesda, Maryland 20205
301/496-6618

Dr. Perry J. Gehring
Vice President of Agricultural
Products R & D and Director of
Health and Environmental Science
Dow Chemical Company
Midland, Michigan 48640
517/636-1089

Dr. Richard A. Griesemer
Director, Biology Division
Oak Ridge National Laboratory
P. O. Box Y
Oak Ridge, Tennessee 37830
615/574-0213 or FTS 624-0213

Dr. Kim Hooper
Chief, Hazard Evaluation System and
Information Service
State of California
Department of Health Services
2151 Berkeley Way, Room 504
Berkeley, California 94704
415/540-2012

Dr. Sanford Miller
Director, Bureau of Foods
Food and Drug Administration
Federal Building 8
200 C Street, SW
Washington, D.C. 20201
202/245-8850

Dr. Ruggero Montesano
Head, Unit of Mechanisms of
Carcinogenesis
International Agency for
Research on Cancer
World Health Organization
150 Cours Albert-Thomas
69372 Lyon Cedex 2
FRANCE
337 8 758 181

Dr. Ian Munro
Director, Canadian Centre for
Toxicology
University of Guelph
645 Gordon Street
Guelph, Ontario N1G 2W1
CANADA
519/824-4120, ext. 8917 or 18

Dr. Frederica Perera
Division of Environmental Sciences
Columbia University
School of Public Health
Senior Science Advisor
The Natural Resources Defense
Council
60 Haven Avenue Room B-106
New York, New York 10032
212/949-0049 SECY: Carolyn Parris
Columbia Univ. 212/694-3465

Dr. Robert A. Scala
Senior Scientific Advisor, Medicine
and Environmental Health Department
Research and Environmental Health Division
Exxon Corporation
P.O. Box 235, Mettlers Road
East Millstone, New Jersey 08873
201/873-6061

Dr. Andrew Sivak
Vice President, Biomedical Science
Arthur D. Little, Inc.
Acorn Park
Cambridge, Massachusetts 02140
617/864-5770, ext. 2525

Dr. I. Bernard Weinstein
Professor of Medicine and Public Health
Faculty of Medicine
Columbia University
701 W. 168th Street
New York, New York 10032
212/694-6921

Dr. Gerald N. Wogan
Head, Department of Nutrition
and Food Science
Massachusetts Institute of Technology
77 Massachusetts Avenue, Bldg. 16-333
Cambridge, Massachusetts 02139
617/253-3188

NTP Board of Scientific Counselors

Dr. Norman Breslow
Professor and Chairman
Department of Biostatistics SC-32
University of Washington
Seattle, Washington 98195
206/543-1044

Dr. Henry C. Pitot
Director
McArdle Laboratory for Cancer Research
Medical School
University of Wisconsin
450 North Randall Avenue
Madison, Wisconsin 53706
608/262-2177

Dr. James A. Swenberg
Chief
Chemical Industry Institute
of Toxicology
P.O. Box 12137
Research Triangle Park, N.C. 27709
919/541-2070

NTP Board of Scientific Counselors'
Ad Hoc Panel on Chemical Carcinogenesis Testing and Evaluation

Subpanels

Data Required from Prechronic Studies

Dr. Andrew Sivak, Chairperson

Dr. Richard Adamson

Dr. Perry J. Gehring

Design of Chronic Studies

Dr. Robert A. Scala, Chairperson

Dr. Norman Breslow

Dr. Richard A. Griesemer

Dr. Ruggero Montesano

Dr. Henry C. Pitot

Dr. I. Bernard Weinstein

Techniques to Supplement or Foreshorten Cancer Tests

Dr. Frederica Perera, Chairperson

Dr. Kim Hooper

Dr. James A. Swenberg

Dr. Gerald N. Wogan

Regulatory Aspects

Dr. Sanford Miller, Co-chairperson

Dr. Ian Munro, Co-chairperson

Dr. John Doull

Introduction and Overview

The Ad Hoc Panel on Chemical Carcinogenesis Testing and Evaluation was established in response to a recommendation from the March 1983 meeting of the National Toxicology Program (NTP) Board of Scientific Counselors. The charge to the Panel was to review the basic biology and chemistry of chemical carcinogenesis and to recommend to the NTP Board of Scientific Counselors, methods that NTP should use for the detection and evaluation of chemical carcinogens. Recommendations for Panel membership and specific topics for consideration by the Panel were also provided and additional suggestions were received from the NTP staff and the public. At the first meeting of the Panel, Dr. M. Mendelsohn, Chairman, NTP Board of Scientific Counselors, requested the Panel to evaluate current knowledge in the area of carcinogenesis testing and to make recommendations that meet the scientific, regulatory and industrial needs in this area. To accomplish this task, the Panel was divided into subpanels to deal with the general areas of Short Term Tests, Subchronic Studies and Related Issues, Chronic Studies and Regulatory Aspects. Meetings of the full Panel were held on May 17, 1983, August 23, 1983, January 13, 1984, and May 3, 1984, and Subpanel meetings were held on July 14, 1983, September 14, 1983, September 15, 1983, and September 21, 1983. Notices for these meetings were published in the Federal Register and all meetings were open to and well attended by the public.

A draft report of the Panel was released on February 15, 1984, with an invitation to interested parties for review, evaluation and comment. Responses were received from numerous trade organizations, scientific bodies, governmental groups including the NTP staff, and concerned individuals both in this country and abroad. The Panel has reviewed all of the comments and incorporated some of the material into this report. Several comments concerned topics which

were relevant but not specifically included in the topics covered in this report. Since many comments were extensive and well documented and often included innovative and thoughtful suggestions, we have arranged to make all of the comments available through the National Technical Information Service repository service, PB-84-225945. The Panel is grateful to our colleagues in industry, government and academia for their comments and participation in this effort. We recognize that there are additional topics and issues that need to be considered by the scientific community and that several of the topics considered in this report have not been fully resolved. In these areas we consider our efforts to be a reasonable first step and recommend that the NTP Board of Scientific Counselors conduct workshops and other activities designed to maintain and promote the productive and cooperative interaction which has characterized the activities of this Panel.

Mission of the NTP Program

The Department of Health, Education and Welfare announced the establishment on November 18, 1978, of the National Toxicology Program. The broad goal of this Program is to strengthen the Department's activities in the testing of chemicals of public health concern, as well as in the development and validation of new and better integrated test methods.

To accomplish this goal, the Program was established as a Department-wide effort to provide needed information to regulatory and research agencies and to strengthen the science base. The Program is presently composed of the relevant activities of the Food and Drug Administration/National Center for Toxicological Research; the Centers for Disease Control/National Institute for Occupational Safety and Health; and the National Institutes of Health/National Cancer Institute, National Institute of Environmental Health Sciences.

In this report, the Panel has focused on activities associated with the testing of chemicals for toxicity rather than on the use of NTP data for

regulatory purposes since the latter is not part of the NTP mission nor of our charge. The inclusion of Panel recommendations regarding validation and additional methods as well as those relating to the current protocols is intended to enhance the quality and utility of NTP test data for both scientific and regulatory purposes. We recognize the need to distinguish between the responsibility of NTP to determine whether a chemical is carcinogenic to animals under the conditions of the test and the regulatory responsibility for the determination of risk to humans from chemicals found to be carcinogenic in animals. NTP test data constitutes only part of the total toxicology data base for an agent that must be considered by the regulatory agency. Thus classifications which utilize the total toxicology data base for an agent such as the International Agency for Research on Cancer (IARC) assessment of animal data (sufficient, limited, inadequate, no evidence) are not appropriate for individual NTP bioassays or short-term test data. The Panel recommends, however, the use of descriptive terms in the technical report for each chemical which clearly and succinctly indicate both the scientific quality of the data and its possible relevance to human risk assessment. An effort should be made to develop descriptive terms which facilitate the subsequent use of the NTP data for IARC classification and other regulatory purposes.

Confounding Factors Related to the Design, Conduct and Interpretation of NTP Studies

Toxicology and safety assessment are emerging disciplines and the methodologies associated with these activities are under constant development and change. This is particularly true of chronic testing where we have witnessed remarkable changes in testing protocols from the early 1960's to the present. It has been established, for example, that certain confounding variables, if not properly controlled, will profoundly influence the outcome of long-term

animal tests. It was because of this that guidelines for the conduct of these tests were initially developed and are continuing to evolve. Regulatory agencies have recognized that acceptable standards for the conduct of animal bioassays need to be modified and updated as knowledge of the science of toxicology expands. Studies conducted 10 to 15 years ago cannot be expected to provide the degree of certainty regarding their predictive value as would be expected of studies conducted by current standards. It must also be recognized, however, that the development of guidelines that specify, in a general way, the protocols for conducting chronic tests address only part of the problem. Failure to follow accepted principles of scientific investigation including meticulous attention to experimental detail, generally accepted quality control procedures and adequate record keeping and reporting have cast doubt on the validity and usefulness for regulatory purposes of large amounts of data generated in the past. Thus factors affecting the usefulness of animal data include two components: one related to the appropriateness of the methodologies employed and a second related to the scientific integrity of investigators and the adherence to accepted principles of scientific investigation.

It should be recognized that no one experiment can be expected to provide answers to all potential questions. Indeed, more often than not, well designed experiments lead to the need for further studies to elucidate questions raised by the initial investigations. Attention should also be focused on the control of variables that may potentially confound results. Intercurrent disease, inappropriate choice of the dosing levels, regimens and vehicles, inadequate attention to clinical observations and a host of other factors may confound the experiment in a way that makes it impossible to clearly interpret the results and this renders the data useless or of limited value for regulatory purposes.

Major questions have been raised in recent years concerning the use of high doses in cancer bioassays. This has resulted in numerous attempts to define the highest dose level to be used in such studies. The Panel recognized the need for sufficiently high doses to characterize toxicity but also recognized the possibility that using such doses may complicate evaluation of the results. In this report, the Panel has made a number of general recommendations in this area but we recognize that in practice, decisions concerning the selection of doses for the testing of individual chemicals must be made on a case by case basis.

In recent years, considerable attention has been focused on the question of design aspects of chronic studies but little attention has been given to the development of criteria for acceptance/rejection of results on the basis of confounding variables. The Panel recommends that NTP look into this question and develop a set of principles and guidelines that will assist in determining whether the results of a particular bioassay are clearly due to treatment or whether the results are confounded in a way that makes such a determination impossible or equivocal. Among the factors which should be included in these guidelines are the appropriateness of the animal species and strain, the level and duration of dosing, the choice of route, duration of testing, statistical methods, group size, nutritional state and disease prevalence. The relevance of the animal studies to the human setting should also be considered. Factors which confound experimental observations should be clearly pointed out in the bioassay report and a determination should be made regarding the extent to which each of the confounding factors affect the interpretation of the data in the species tested as well as their effect on the usefulness of the data for regulatory purposes.

Because of the cost associated with chronic animal testing and the potentially immense economic and public health importance of the results, it is imperative that good laboratory practices (GLP) be adhered to. This consists not only of meeting the legislated requirements (Federal Register, 1978), but also of ensuring that documented evidence exists regarding all changes in protocols, disease outbreaks and all other factors that affect the outcome of the test. This information should be available to those responsible for assessing experimental results. Good experimental practice goes beyond legislated requirements regarding GLP and every effort should be made to achieve the highest possible standards of scientific excellence in the conduct of chronic bioassays. It is, therefore, incumbent on those responsible for conducting and monitoring such tests to pay meticulous attention to details, particularly unexpected observations or observations that may, at first glance, not be expected to influence the outcome of the study but which may, in the final evaluation, have a very significant impact on the interpretation and usefulness of the data.

Considerations Related to Previously Conducted Studies

The criteria for judging the adequacy (value and importance) of a rodent carcinogenesis bioassay as a basis for a regulatory decision cannot be determined solely on the basis of whether a study meets current or previously acceptable protocol requirements. As discussed above, the scientific quality and predictive value of both "old" and "recent" bioassays is dependent on the adequacy of the experimental design, the competency and completeness with which the study was conducted and reported and on the evaluation and interpretation of the results. The Panel recognizes that confounding factors may limit the adequacy of "recent" bioassays and that "old" studies may serve as an adequate basis for regulatory decisions even though they have protocol

deficiencies in the number of animals/group, number of dose levels, absent clinical observations, etc. The requirements of the current NTP bioassay protocol and the enhancements recommended in this report are intended to improve the predictive value as well as the scientific quality of NTP test data. The Panel recognizes that protocol modifications and enhancements which clearly improve the quality of the study as a basis for regulatory decisions may raise questions about the adequacy of previously conducted tests. Our intent is not to imply that previous studies would or should be judged inadequate on the basis of modern criteria, but that in selected instances modern criteria of bioassay design, conduct, evaluation and interpretation might alter previous decisions. The Panel recommends, therefore, that in those instances where older NTP data served and continues to serve as the primary basis for regulatory decisions, the NTP staff re-evaluate selected studies in the light of new knowledge and determine the impact, if any, on the interpretation of the results.

The introduction of GLP legislation has had a substantial influence on the conduct and reporting of animal studies carried out for regulatory purposes. Many previously conducted bioassays and some ongoing NTP studies have not yet benefited from detailed appraisals aimed at ensuring their integrity. Since the credibility of regulatory decisions on specific agents depends in large part on the integrity and quality of the toxicology data base for that agent, it is considered advisable that special attention be focused on determining whether the interpretation of previously conducted studies would be altered by the application of GLP principles. The Panel recommends, therefore, that the NTP staff consider auditing some of the older studies that were pivitol to decision making and where no new or additional studies have come forward in the intervening years. The integrity of NTP test data and the credibility of the NTP program

are of fundamental importance not only to the regulatory process but also to the science of toxicology. With this in mind, the Panel also recommends that the NTP staff consider other approaches for ensuring the integrity of test data and maintaining the credibility of all such programs. Such approaches might include, for example, the testing of an agent known to be a human carcinogen but not previously tested in animals and the validation of a previously tested animal carcinogen by retesting.

Potential Role of Short Term Tests in Screening for Carcinogenic Potential and in Risk Estimation

At the present time, short term in vitro and in vivo tests are considered to be useful supplements to the chronic bioassay in identifying and assessing potential human cancer risks. Although in some cases positive short term test results by themselves may serve as a basis for considering a substance to be a potential human carcinogen, in general they are used to flag the need for a chronic bioassay and to set priorities for long term testing. They are also viewed as helpful in studying possible mechanisms and sites of action of carcinogens identified as such in the bioassay.

Considerable progress has been made in establishing the sensitivity and positive predictive value of the commonly used assays for genetic toxicity/potential carcinogenicity. However, an important gap remains concerning their specificity. Morever, there is a recognized deficiency of short term tests to detect agents that do not exert genetic effects (such as some promoting agents). Research should be emphasized in these areas.

At the same time, a thorough analysis of the extensive existing data base on short term tests now being updated by the EPA Gene-Tox, the NTP and the International Collaborative Programs could provide a means of identifying the best available chemical-specific or chemical class-specific batteries of short

term tests. Ultimately, such batteries might replace the long term bioassay as the major carcinogenicity testing method.

Another application of short term tests is in monitoring humans with exposure to carcinogens. The panel recommends the validation of potential markers of biologically effective dose and preclinical response to carcinogens in humans. Parallel, comparative studies in laboratory animals and humans are a logical approach. If studies are successful, these methods could improve human risk assessment in two ways. First, calibration of markers in experimental models and humans can provide a better basis for using non-human studies to assess human risk. Second, their incorporation into conventional epidemiology can result in an increased power to detect carcinogenic risks earlier and at lower exposures as well as to more precisely estimate the magnitude of human risk.

Acknowledgement

The Panel is particularly grateful for the dedication and hard work of the Subpanel Chairpersons - Dr. Frederica Perera, Dr. Ian Munro, Dr. Sanford Miller, Dr. Robert Scala, and Dr. Andrew Sivak - our Executive Secretary, Janet A. Riley, and the NTP staff and to all of the participants in this endeavor.

Report Of Subpanel On Short Term Tests

A. Introduction and Charge to the Subpanel

A Subpanel consisting of Dr. Frederica Perera, Chairperson, Dr. Kim Hooper, Dr. James A. Swenberg, and Dr. Gerald N. Wogan was asked to make recommendations to NTP for supplementing the chronic carcinogenicity bioassay with short term tests. To this end, the Subpanel reviewed various aspects of short term tests (advantages, disadvantages, usefulness in studies of mechanism, correlation data) as well as the present NTP short term testing program and existing relevant data base (in Section C). The Subpanel then evaluated specific short term tests now considered to be possible candidates for incorporation into the NTP bioassay program (background; state of the art; pros/cons; specific recommendations) as well as tests for transformation and promotion (Section D). Finally, the Subpanel offered specific recommendations to NTP (in Section E).

Focus of Review

In addition to reviewing short term tests according to the above format, this review focuses on those short term tests for genetic and other endpoints related to carcinogenesis which have actual or potential applicability to both laboratory (*in vitro* and *in vivo*) and human (*in vivo*) studies.* Put another way, these are short term tests that can be used in parallel to study nonhuman mammalian and human cells/tissues both *in vitro* and *in vivo*.

---

*We use the term *in vitro* to describe short term tests that involve exposure of cultured cells and tissue in the test tube; while the term *in vivo* comprises both those tests in which the intact animal or human is exposed and cells are removed, cultured and assayed (sometimes termed *in vivo/in vitro*) and those tests which involve exposure and study of the intact animal or human (cells or biological fluids) only.

This focus is not intended to detract from the value of the many *in vitro* and *in vivo* tests used in the detection and study of chemical carcinogenesis (e.g., cell transformation assays and altered liver foci) that do not at present have applicability to human studies.

In adopting this focus, the Subpanel is responsive to the recognition by a number of committees and individual experts of the likely role of genetic damage in the etiology of cancer and of the need for sensitive validated methods to estimate individual human exposure and preclinical response to carcinogenic chemicals. These methods would then be used in the interpretation of *in vitro* and animal data, in epidemiological studies of cancer causation, and in the direct evaluation of human exposure and risk. (See Banbury Report, 1982; Bloom (ed.) 1981; Vainio *et al.*, 1983, NAS, 1983; NAS Briefing Panel, 1983; ICPEMC 1983; Legator, 1983; and Perera and Weinstein, 1982.)

This approach is a way of achieving two related goals:

(1) to improve our ability to use non-human studies as a basis for assessing human risk by directly comparing and calibrating findings in humans with those in experimental animals *in vivo* and *in vitro* (In this regard it should be possible eventually to reduce the need for the chronic *in vivo* bioassay through greater reliance on short term tests.); and

(2) to increase the power of conventional epidemiology to detect carcinogenic risks earlier and at lower exposures than heretofore possible and to more precisely assess the magnitude of human risk by incorporating "markers" for internal activated dose, biologically effective dose, and subclinical, preclinical or clinical genetic response.* These markers could be provided by short term tests in human cells and tissue once validated for that purpose.

Figure 1 shows the relationship between external exposure, several measures of dose, and genetic response with examples of tests in each category. Of course, it remains to be definitively established if the tests listed under "Genetic Response" are in fact indicating an early step in chemical carcinogenesis. Hence for the present they can be regarded as a subset of indicators of biologically effective dose and as *potential* markers of early carcinogenic response.

*Internal dose here refers to (qualitative or semi-quantitative) evidence of the activated agent in body fluids and tissues; biologically effective dose to quantitative measures of interaction by an agent that has reacted with critical cellular targets (DNA, RNA, protein) but has not necessarily induced a heritable genetic effect; and genetic response refers to the induction of heritable changes in the genome (see Perera and Weinstein, 1982).

Figure 1

SOME POTENTIAL MARKERS FOR ANIMAL/HUMAN IN VIVO STUDIES

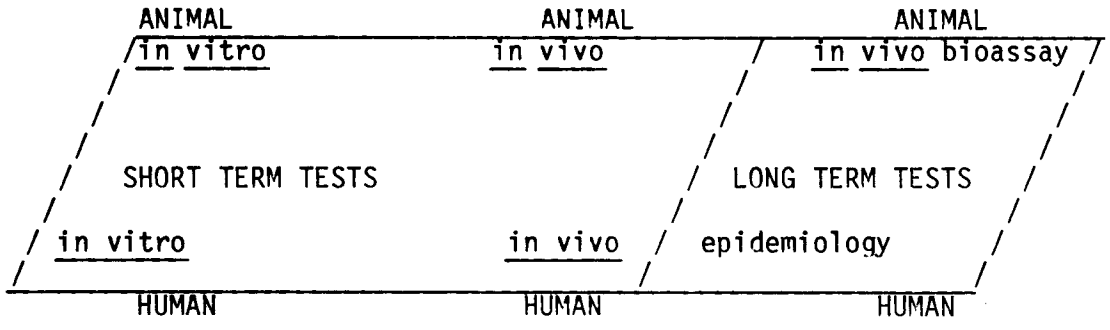
External Exposure	Internal Dose	Biologically Effective Dose	Early Genetic Response
	--Mutagenicity of urine	--DNA, RNA or protein adducts	--Chromosomal aberrations
	--Adduct excision	--SCE	--Somatic cell mutation
	--DNA repair	--Micronucleus	--Sperm abnormalities*

*Sperm abnormalities may reflect a genetic or a phenotypic change.

Figure 2 shows the relationship between the four types of testing data (in vitro vs. in vivo in experimental animals and in humans). It is obvious that development of comparable laboratory and human data is necessary to carry out the "parallelogram" approach.

Figure 2

PARALLELOGRAM APPROACH



Concerning the practical possibility of parallel testing in research studies, while most applicable noninvasive short term methods are still at an early stage, a number of them have been widely applied to laboratory animals, cells and in a small number of limited human studies. These can be considered either practically useful or well on the way to being applicable (Bridges, 1982; Banbury Report, 1982; Bloom (ed.) 1981, Vainio et al., 1983; Legator, 1983; Perera and Weinstein, 1982). Furthermore, biological fluids from human and experimental animals -- urine, blood, sperm, etc. -- are available for analysis in a variety of short term tests for mutagenicity, DNA damage and repair, binding to DNA and protein, and chromosomal effects. Populations suitable for parallel study could include patients receiving chemotherapeutic drugs, certain worker populations, and individuals or groups exposed to carcinogens through accidents or a significant environmental source such as hazardous waste sites. Costs of such studies are "within bounds of funded research projects" (Bloom, ed., 1983).

Scope and Process of Review

Given the particular focus of the Subpanel's review and the large quantity of material to be covered in that regard, a detailed examination of the basis for NTP's selection of its current battery of in vitro tests was not made. Reliance instead was on the major published correlation studies and summary data made available from the EPA Gene-Tox Program to answer the question: "Is this set of tests a reasonable qualitative screen for potential carcinogenicity for purposes of ordering priorities for chronic bioassay testing?." (Thanks are expressed to Michael Waters and Barry Howard at EPA for their valuable assistance in providing the necessary information.) Deficiencies in the data base were noted as well as the

ongoing studies aimed at filling these gaps.

A number of issues and methods are not addressed in detail in this report because they do not fall within the focus adopted by the Subgroup. This in no way detracts from their importance in carcinogenicity research and testing. They include:

(1) the many valuable *in vitro* tests as well as tissue and organ specific short term *in vivo* animal tests (e.g., altered liver foci, and lung, skin and breast tumor induction in rodents) which are not readily applicable to human monitoring;

(2) host factors in chemical carcinogenesis;

(3) interaction between initiating agents and chemicals that enhance their initiating effect (i.e., cocarcinogens) and interaction between chemicals and viruses; and

(4) the role of oncogenes and the interactions between oncogenes and chemical carcinogens for which tests are currently unavailable. The Subgroup noted that, given the rapid development in understanding of the importance of oncogenes in cancer, it is possible that short term tests based on oncogene activation may become available in the not-too-distant future; such tests should then be incorporated into the NTP program.

Also outside the scope of this review is the value of short term tests in designing large scale bioassays (e.g., dose selection, study size etc.).

Another area of interest which is only briefly discussed is the potential usefulness (and current limitations of) short term tests as a means of interpreting bioassay results in terms of potential human risk. This topic was recently covered in depth by an IARC working group. (See IARC, 1983 for review).

Regarding the process of review, tests were initially selected for evaluation based on their applicability to parallel testing as discussed above, their state of development, and their complementarity with each other. Although not now applicable to human studies, *in vitro* and *in vivo* tests for cell transformation and promotion were also reviewed because of their potential importance in understanding the complex, multi-stage process of cancer. (See

Section D.) The Subpanel members then conducted reviews of the recent relevant literature, especially major summaries, and selected interviews with researchers in the field. The Subpanel solicited contributions from the public at an open meeting on September 21, 1983, and has considered all written and verbal comments.

B. Background

Gene mutation, errors of chromosome structure and errors of chromosome number are known to occur in humans with clear casual relationship to human genetic disease and strong association with human cancer (Gilbert, 1983; Radman, et al., 1983; Yunis, 1983). Available short term tests can be classified by endpoint measured: mutagenicity, chromosomal effects, DNA damage and repair, and neoplastic transformation. They can also be classified according to whether they are potential indicators of internal dose, biologically effective dose, or genetic response as defined above. Over 100 short term tests exist -- each with peculiar advantages and disadvantages. Only a small percentage of these tests have been applied widely enough to "own" a substantial enough data base for consideration by the Subpanel. Waters et al., (1983a) have pointed out that not one short term test meets all of the seven optimal criteria for such methods; i.e., has established adequate predictive capacity, theoretical basis, reproducibility, rapidity, feasibility, dose sensitivity, and, especially, ability to predict human carcinogenic potency. Moreover, chemicals vary in the kind of genetic changes they induce (e.g., a chemical may be hazardous because it induces aneuploidy while not being active as a clastogen or gene mutagen).

A review of in vitro and in vivo short term tests reveals the following limitations. Many are generic problems common to both in vitro and in vivo short term methods and are even shared to some degree with the chronic bioassay.

- Tests are subject to confounding by species and strain as well as sex, organ, or tissue specificities.

- Laboratory tests in vitro and with experimental animals in vivo generally involve exposure to a single agent whereas humans are generally exposed to multiple agents which may compete for absorption, metabolism, and excretion (Garattini, 1983).

- Although short term tests can be helpful in studying the mode of action of a carcinogen - as a recent IARC report has pointed out - they cannot at present definitively identify the mechanism of action relevant to carcinogenic or classify carcinogens according to mechanism of action, (IARC, 1983). (See discussion below.)

- Most short term methods (at least in theory) test for the mutagenic and/or initiating action of agents. They are unlikely to detect promoters -- a significant gap given the substantial human exposure to such agents. This is a major problem in the linking of dose and gene response data with cancer incidence (Bridges, 1982). In addition, short term methods do not exist for chemical interaction with oncogenes.

- Data are limited with which to compare and correlate in vitro and in vivo short term methods against each other and against the standards of animal and human carcinogenicity. (See below). In particular, there are few data on animal noncarcinogens with which to determine the specificity of tests. Also, only tests for mutagenicity have been widely used.

- There is no conclusive etiological linkage between specific genetic effects measured by available short term tests and specific neoplastic events or phenotypes (Tennant, et al., 1983).

- A major unresolved problem is the value of short term laboratory data to estimate the carcinogenic potency of chemicals in humans. There are conflicting data on this question (McCann et al., 1982). Therefore, the tests are generally regarded as qualitative rather than quantitative predictors.

- There is a lack of adequate data on the endpoints measured by short term tests: background level, inter- and intra-individual variation, and their persistence.

- False negative results can occur unless an effort is made to test the ultimate reactive metabolite or to ensure that an appropriate metabolic activating system is used with in vitro test methods. In theory, both false negative and false positive results might also occur because the short term test does not reflect the complexity of the intact animal.

- Certain testing systems (especially in vitro) are unresponsive to some chemicals or classes of chemicals. Ideally, one would want to tailor class-specific combinations of tests in order to avoid false negatives; however, marked differences are also possible within a chemical class making this approach difficult (Waters et al., 1983a).

At the same time, short term tests also have significant advantages vis-a-vis other testing methods:

- Short term tests can be highly sensitive, precise and can be used over a wide range of doses to develop comprehensive dose-response data within that system.
- In combination or a "battery," they can measure a wide range of genetic endpoints.
- They are relatively quick and low cost, especially compared with the chronic bioassay.
- They can be useful in setting long term testing priorities; in studying mechanisms of toxicity; and, as discussed above, have a significant potential for use in human risk assessment. Although short term tests have been recommended as a replacement for chronic bioassays (Weisburger and Williams, 1983) they are generally regarded as supplements rather than as an alternative.
- A battery of short term tests can now be used as a qualitative predictor of human response because of the high positive correlation between the carcinogenic and mutagenic activities of chemical carcinogens (IARC, 1983).

Potential Usefulness of Short Term Tests in Better Understanding Mechanisms of Carcinogenesis

There is considerable interest in the development of methods to help understand the mechanism(s) of chemical carcinogenesis. Obviously, the scope of such an undertaking could vary from applying relatively well developed organ-specific protocols for initiation/promotion to pure basic research. A recent IARC Working Group evaluated present abilities to classify chemical carcinogens according to mechanism of action and concluded that "at present, no classification of carcinogens according to mechanism could be exhaustive or definitive. On the other hand, classification of mechanisms has considerable value for particular scientific purposes" (IARC, 1983, p.61). IARC further cautioned:

> "Carcinogens can be grouped according to the biological effects they exhibit in short term tests, but this does not imply that the mechanisms producing those effects are identical with the mechanisms involved in their carcinogenic effect" (p.57).

Therefore, it appears that classification of carcinogens into different categories with varying degrees of risk for humans is premature.

In the course of testing a chemical, the NTP develops a data base of short term test results. As mentioned earlier, chemicals vary in terms of the kind of genetic effects they exert. One frequently sees a mixture of positive and negative results in a battery of short term tests on the same chemical, each of which measures a different effect. Although not necessarily biologically inconsistent, such mixed results further complicate risk assessment.

In addition to short term test results, NTP develops data on acute and chronic toxicity, as well as carcinogenicity. In the future, data on pharmacokinetics, metabolism and distribution will also be available. These data should enhance studies of the mechanism(s) of carcinogenesis.

In particular, short term testing data may eventually shed light on the mode(s) of action of a chemical and hence be helpful in interpreting animal bioassay data. For example, the data may be highly complementary to each other, i.e., no effect in an animal carcinogenicity bioassay and battery of short term tests for mutagenicity or positive in both. Conversely, the data may not be complementary, i.e., positive results in short term tests but without effect in animal bioassays or the reverse. The standard battery of tests provides some insight into possible mechanisms responsible for three of the four examples cited; however, the case of a chemical which is negative in a battery of short term tests but positive in a cancer bioassay could reflect the limitations of the short term test battery (e.g., the absence of tests to detect promoters).

Specifically, because of the limitations of the present battery (for example, relative inability to detect non-initiating/non-mutagenic carcinogens, the fact that it is not known which endpoint(s) are critical in carcinogenesis, etc.), a lack of positive results, by itself, could not justify a conclusion that

a positive animal test was a "false positive." Therefore, for purposes of interpreting animal data, negative short term test data do not by themselves provide a basis for discounting positive results from animal bioassays. As IARC has already done for purposes of classifying a chemical regarding carcinogenicity where animal data were limited, short term test positives can be regarded as support for a conclusion of potential human carcinogenicity based on experimental data. However, a positive bioassay cannot be invalidated or discounted by negative short term data alone.

The development and use of additional test systems should provide greater understanding of the mechanisms involved in such cases. Techniques that identify promotors are presently available for a number of tissues including mouse skin (Berenblum, 1954; Boutwell, 1974; Slaga et al., 1978, 1980), rat liver (Peraino et al., 1971, 1980; Pitot et al., 1980, 1982; Farber, 1982), urinary bladder (Hicks, 1978), lung (Armuth and Berenblum, 1972), mammary gland (Welsch, 1977), intestine (Narisawa et al., 1974), thyroid (Hiasa et al., 1982a) and kidney (Hiasa et al., 1982b). Of these, the skin, liver and bladder test systems have been evaluated more completely. As mentioned in the section on tests for promotion, the NTP should support further development of these and related techniques.

While the incorporation of additional test systems aimed at understanding the mechanism(s) of carcinogenesis will not definitively or exhaustively prove said mechanisms, they will provide a greater scientific understanding of chemical carcinogenesis.

Validation and Correlation Data

Individual tests have been validated to a greater or lesser degree. This term refers to the test's performance with regard to the following characteristics:

(1) Sensitivity (proportion of positive results among carcinogens tested);

(2) Specificity (proportion of negative results among noncarcinogens tested);

(3) Predictive Value (proportion of correct test results among all chemicals tested);

(4) Reproducibility (intra- and inter-laboratory); and

(5) Dose Sensitivity.

The term correlation refers to 1-3 above in that it reflects the degree to which the qualitative response to a particular agent in a short term test is the same as that in the chronic bioassay and/or epidemiological studies. The determination of the extent to which short term test results correlate with animal carcinogenicity and epidemiological data is impeded by major gaps in all three data bases. In particular, the "standard" against which short term tests are evaluated is usually animal carcinogenicity data (Dunkel, 1983); yet disparate results could be a consequence of an inadequately conducted "false" negative animal test as was the case with the Japanese food additive, AF-2 or of an apparent "false positive" bioassay due to variability in spontaneous tumor incidence.* Another problem is that because of the lack of an extensive data base on chemicals which are not positive in animal bioassays, one cannot accurately calculate the specificity of a test (Shelby and Stasiewicz, 1983). Furthermore, only a limited range of chemical classes has been tested in the various short term tests for effects other than mutagenicity (Tennant et al., 1983). Similarly, the epidemiological data are sparse and will continue to be so because of the limitations of conventional methods to identify carcinogens (Karstadt, 1981). The ongoing update of the EPA Gene-Tox Program and the effort to update the NTP/NCI data base for about 300 chemicals (recently begun by NIEHS and expected to be complete in 1984) should cast a great deal more light on the subject.

*See discussion of the "false positive" issue, on page 215.

With these limitations in mind, Tennant et al. (1983) have recently compiled results of short term tests for 74 chemicals previously tested in the rodent bioassay, and have observed a reasonable concordance between rodent carcinogenicity and genetic toxicity. The short term tests were grouped in four categories: mutagenicity, chromosomal effects, primary DNA damage, and mammalian cell transformation. Adequate results (in at least 3 of the above categories) were available for 33 of the 74 chemicals, of which 25 chemicals were positive for animal carcinogenicity. Of those 25 animal carcinogens, 23 (or 92%) of the sample were positive for genetic toxicity; i.e., in at least one test/category. Furthermore, genetic toxicity was particularly concordant for 13 chemicals which induced both uncommon and common tumors in two rodent species. Five substances were negative in the chronic animal bioassay; of these, 2 were consistently negative in short term tests and 3 gave at least one positive and one negative result.

A recent report has summarized the results from genetic bioassay systems (in the same four test categories listed above) for agents or groups of agents classified by IARC as known or suspected human carcinogens (Waters et al., 1983b). The report reviewed the recent published literature and the Gene-Tox data base regarding those substances. All 24 of these substances having genetic toxicity data were positive in at least 2 different categories of tests. This report demonstrates the high qualitative correlation between human carcinogenicity and genetic toxicity; it also illustrates the wide variability between compounds in terms of the amount of genetic testing data available even for well-known compounds and in the percentage of positive tests for individual compounds. For example, ethylene oxide was uniformly positive in all test categories for which data were reported; whereas there were no short term data whatsoever for oxymethalone and only two reports for phenytoin.

In a novel attempt to relate genetic short term bioassay data to human genetic monitoring data, Waters et al. (1983a) compared results of laboratory tests (Salmonella, HGPRT, SCE) with in vivo data in humans on chromosomal damage for cyclophosphamide, among alkylating anticancer drugs now classified by IARC as human carcinogens. They concluded that the genetic bioassays detect effects at dose levels similar to those that produce chromosomal damage in humans.

Table 1, based upon Vainio et al. (1983), compares human and animal genetic toxicity and carcinogenicity results for several chemicals or categories of chemicals.

Table 1

CORRELATION BETWEEN EXPERIMENTAL AND EPIDEMIOLOGICAL FINDINGS

	AGENTS				
Method/Endpoint	Cigarette Smoke	Hair Dyes	Anaesthetic Agents	Cytostatic Drugs	Ethylene Oxide
In Vitro/Genotoxicity	+	+	+	+	+
Bioassay/Carcinogenicity	+	+	..	+	+
Human Studies/					
•agent in body fluids	+	+	+	+	+
•mutagenicity of urine	+	+	+	+	..
•covalent adducts	+	..	..	+	+
•chromosomal changes	+	+	..	+	+
Epidemiology/Cancer	+	±	±	+	(+)

Key: + Positive data; (+) Suggestive; ± Contradictory; .. No data

(Vainio et al., 1983, as modified)

In addition to the ongoing Gene-Tox/NTP validation programs, the NTP may be able to draw on the results of international collaborative studies (ICS) aimed at evaluating in vitro and in vivo short term tests, although there seem to be some problems interpreting results of the first ICS. The International Collaborative Program obtained data on 42 chemicals in 35 assays ranging from

bacterial to *in vivo* mammalian tests (de Serres and Ashby, 1981). NTP is currently participating in the International Program on Chemical Safety (IPCS) aimed at first evaluating *in vitro* tests in eukaryotes using 10 selected chemicals and, in a second phase, at assessing the *in vivo* activity of two pairs of carcinogens/putative noncarcinogens (Ashby *et al.*, 1984).

C. The Present NTP Genetic Toxicity Program

At the present time (NTP, 1984) the NTP genetic toxicity program is evaluating four categories of *in vitro* tests:

Mutation
- Salmonella typhimurium
- L5178Y lymphoma (TK+/-forward mutation assay)

Chromosome Aberration and Sister Chromatid Exchange (SCE)
- Chinese hamster ovary *in vitro*

Unscheduled DNA synthesis
- Rat hepatocyte (rat liver primary cell culture assay)

Cell Transformation
- BALB/C 3T3
- SHE
- SHE/SA7
- Rat cell/RLV

When NTP's compilation of results in these test systems is complete, the Panel recommends that the Board of Scientific Counselors convene an expert group to evaluate the results and identify future directions accordingly.

The stated rationale for selection of the tests is that they represent four broad classes of genetic toxicity assays, and they are among the most widely available tests in commercial testing laboratories. NTP's aim is to obtain results for about 40 chemicals in all tests in order to validate them against the animal data. The program is emphasizing the testing of animal noncarcinogens in the battery of *in vitro* assays in order to assess specificity. At present, however, selected tests, rather than the entire battery, are "routinely" used on NTP chemicals because of limited resources. For example, only the Ames Salmonella test is used on virtually all chemicals.

NTP is also developing additional tests, largely through extramural contracts or interagency agreements (NTP, 1984). The NTP test development

program consists of technical validation studies, development of standardized protocols and, in some cases, coded testing by two laboratories with selected chemicals. According to NTP (1984) and Tennant (1983) these additional tests, now undergoing development, include:

1. *in vivo* mammalian somatic cell test systems:
   a. rodent bone marrow cytogenetic systems, and
   b. cytogenetic damage in human lymphocytes (chromosomal aberrations and SCEs*);
2. human cell assay systems using co-cultivation of activating cells;
3. the V79 cell metabolic cooperation/inhibition assay;
4. the micronucleus test;
5. the assay for *in vivo/in vitro* rat and mouse hepatocyte DNA damage;
6. tests for aneuploidy in yeast and Drosophila;
7. assays for gene transposition in Drosophila and in rodent cells;
8. genetic toxicity endpoints in human cells *in vitro*; and
9. tests for the effects of chemicals on germ cells in mouse and Drosophila.

Regarding the data base supporting NTP's selection of the four categories of *in vitro* tests as a prechronic "screen" for potential mutagenicity/carcinogenicity, the EPA Gene-Tox Program probably provides the most comprehensive information and general support although limited to pre-1979 data. Table 2 summarizes the available data regarding the NTP test battery.** It should be kept in mind that the table may reflect bias in the selection of chemicals tested as well as interlaboratory variation in performing the assays.

The Subpanel regards NTP -- combining as it does activities of three agencies, National Institute of Environmental Health Sciences (NIEHS), the National Center for Toxicology Research (NCTR), and the National Institute for Occupational Safety and Health (NIOSH) -- as a unique resource for validating short

*NTP is developing data on human lymphocyte cytogenetics of a medically monitored population to look at background and inter-individual variation.
**Data kindly provided by Dr. Mike Waters and Staff at EPA.

Table 2

GENE-TOX DATA FILE, NTP BIOASSAYS

"CCG" Subset of GENE-TOX (391 compounds, as of 8/23/83)
Bioassays, from Zetafield

	SAL	L51	UDP	SC2	CY&	CT7	CTB
No. '+'	153	27	23	52	24	44	22
No. '-'	47	2	15	8	8	11	24
Other	47	3	7	12	0	0	0
TOTAL	247	32	45	72	32	55	46

Total GENE-TOX Data file (3218 compounds as of 8/23/83)
Bioassays, from Zetafield

	SAL	L51	UDP	SC2	CY&	CT7	CTB
No. '+'	566	40	49	68	42	82	24
No. '-'	139	3	35	21	17	38	27
Other	376	5	14	21	1	1	0
TOTAL	1081	48	98	110	60	121	51

ASSAY CODE

CCG	Chemical designated as carcinogen in rodent bioassays using the Gene-tox criteria for carcinogenicity
CTB	Cell transformation studies in BALB/C - 3T3 cells
CT7	Cell transformation studies in SA7/SHE cells (Syrian hamster embryo cells)
CY&	Mammalian cytogenetics-in vitro cell culture studies, less human
L51	Mouse lymphoma (L5178Y) cells in culture-gene mutation in TK locus
SC2	Sister chromatid exchange - in vitro studies - all animals less human
UDP	Unscheduled DNA synthesis in mammals - studies with rat primary hepatocytes
SAL	Salmonella typhiumurium (histidine reversion test) - use of one or more of the five standard strains (TA98, TA100, TA1535, TA1537 and TA1538)

term tests and in particular for obtaining complementary and parallel data on tests *in vitro* and *in vivo* in animals and humans. NIEHS and NCTR, in their own laboratories and by extramural agreement, are able to develop and validate laboratory testing methods; while NIOSH is in a position to apply assays for mutagenic and genetic effects to human samples from selected worker groups.

D. Evaluation of Specific Tests:*

In the following sections, selected short term tests are discussed primarily in light of their potential usefulness in parallel laboratory animal and human studies and secondarily as screens for potential carcinogenicity. Specific recommendations are included. The Subpanel's overall summary in Section V is based upon these reviews.

Detection Of Mutagens In Body Fluids Or Excreta

1. Background

A presumption of genetic hazard can be made if mutagens are found to be circulating in blood or are present in other body fluids. However, one cannot necessarily assume the absence of hazard if mutagens are not detected. The availability of well-characterized mutation assay systems utilizing microbial cells or mammalian cells in culture makes it feasible to detect and quantify the presence and amounts of mutagenic substances present in body fluids. Further indication of their character can be gained by determination of the requirement for metabolic activation for mutagenicity.

Although many mutagenesis bioasay systems are now available utilizing eucaryotic as well as procaryotic cells, practically all of the existing literature concerns detection of bacterial mutagens in urine of exposed persons. The nature of the available findings is illustrated in Table 3.

*See also NAS (1983); IARC (1980); Bloom (ed.) (1981); Banbury Report (1982); and ICPEMC (1983); for discussion of these tests.

TABLE 3

DETECTION OF MUTAGENS IN URINE BY MICROBIAL ASSAYS*

EXPOSURE	POPULATION	RESPONSE	REFERENCE
DRUGS			
Cytoxan	Patients	+	Siebert and Simon, 1973
Metronidazole	Patients	+	Legator et al., 1975
Niridazole	Patients	+	Legator et al., 1975
Cytostatic Drugs	Nurses	+	Falck et al., 1979
CIGARETTE SMOKING			
	General	+	Yamasaki and Ames, 1977
	General	+	Putzrath et al., 1981
	Workers	+	Dolara et al., 1981
	Workers		Van Doorn et al., 1979
OCCUPATIONAL EXPOSURE			
Styrene production	Workers	-	Cerna and Dobias, 1980
Rubber Mfr.	Workers	+	Falck et al., 1980
Coke Mfr.	Workers	-	Moller and Dybing, 1980
Epichlorhydrin	Workers	+	Legator et al., 1979
Chemical Mfr.	Workers	+	Dolara et al., 1981

*Modified from Bloom, 1981

2. Advantages and Disadvantages

An important advantage of these tests is that they can be directly applied in human and animal studies in their present form. Therefore, they can be useful in detecting exposure of humans to potential carcinogens and mutagens. The currently existing data base consists largely of studies in human populations; few comparable parallel studies have been done in animals treated experimentally with carcinogens. The tests used thus far have been non-quantitative (or at most semi-quantitative) and thus provide data of limited value beyond simply indicating that exposure has occurred.

Another important advantage of this approach is that interpretable data can be obtained in single individuals. This makes the test especially useful in assessing exposure of small groups.

A further advantage of the urine analysis approach is that the samples can be processed before testing. Organic compounds can be extracted and/or concentrated by a variety of techniques, and enzyme treatment will liberate conjugated chemical species. These technical adaptations have not yet found wide application but their feasibility is illustrated by the report of Putzrath et al., 1981, who found that the mutagenic activity of smokers' urine was concentrated by passage through an XAD-2 resin column and could be separated from about 90% of nonmutagenic material by CH_2Cl_2 extraction. Separation of these extractables on HPLC revealed multiple non-polar fractions which were activated to mutagenicity by methylcholanthrene-induced rat liver microsomes. They also remained stable at -20°C for at least 3 months.

Disadvantages include potential confounding sources of mutagenic activity in body fluid samples, such as drugs, food constituents, beverages and lifestyle factors such as smoking and cosmetics in humans, and incidental food ingredients in animals. In some cases, mutagens are extracted in conjugated forms which may represent detoxified material, and may require hydrolysis before assay. Certain chemical classes of mutagens (e.g., nitrosamines) may have short half-lives in blood and excreta and therefore be undetectable. Therefore, urinary mutagens are considered a measure of internal dose (See pp. 18, 19).

Recommendations

Measurement of urinary mutagens is not currently part of the NTP testing protocol. Because it is currently being applied in some human studies, it is recommended that studies on levels of mutagenic material in blood and urine be carried out in bioassay test species as a part of the pre-chronic testing of at least a few selected chemicals in order to provide a background of information

on the quantitative relationships between carcinogen intake and mutagenicity of blood and urine. These chemicals should be chosen on the basis of their mutagenicity as established by the short term test battery, and if possible, should include chemicals encountered in the workplace so that parallel human studies could be possible.

Detection of Covalent Adducts

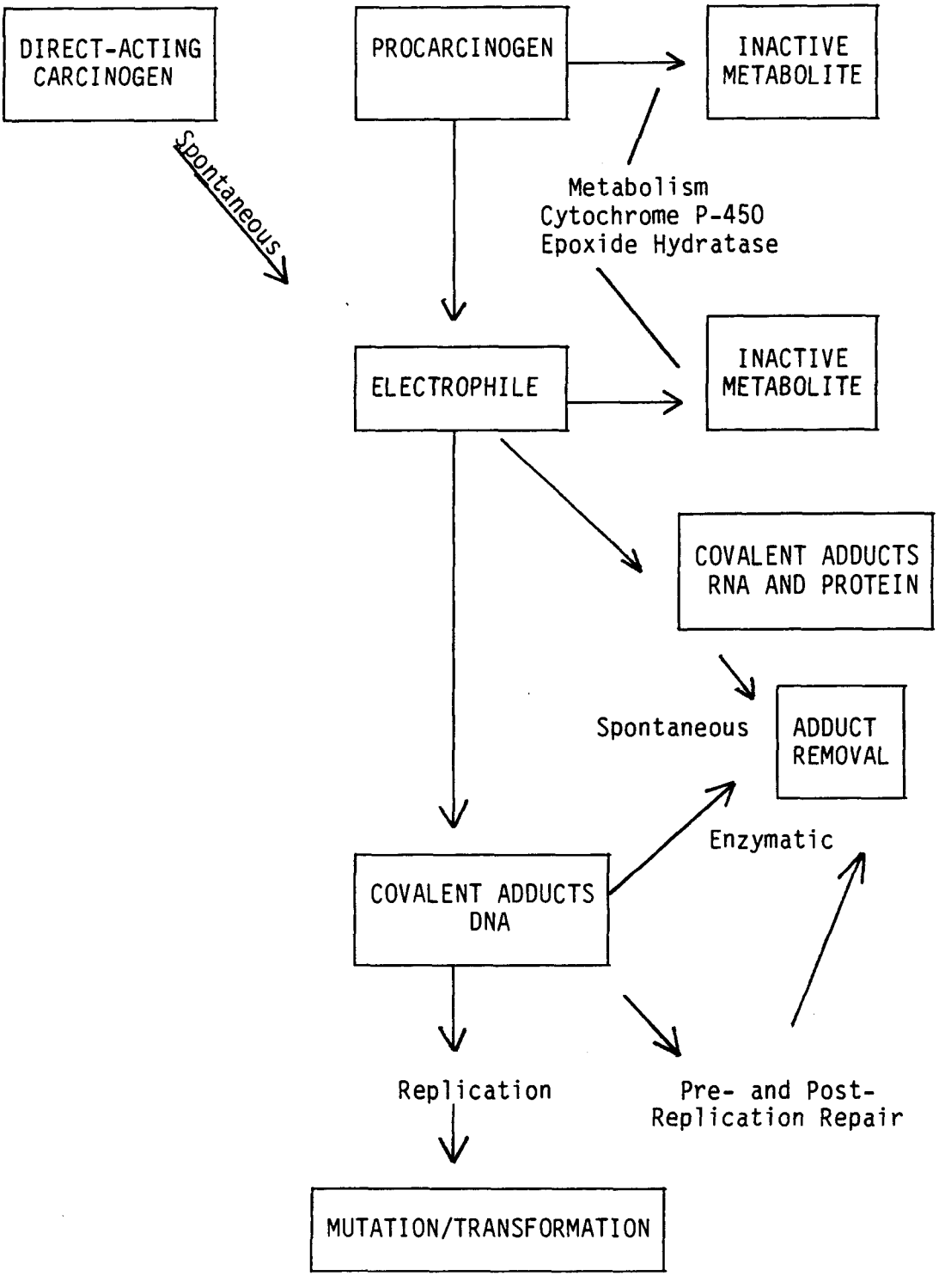
1. Introduction

The rationale underlying the strategy of chemical dosimetry by determining levels of derivatives covalently bound to cellular macromolecules is based on current understanding of the mode of action of genotoxic carcinogens and mutagens. Most chemicals that are active as carcinogens and mutagens have electrophilic properties, or are metabolically converted into electrophiles (see Figure 3). These reactive forms of the compounds undergo attack by nucleophilic centers in nucleic acids and proteins, resulting in the formation of covalent adducts. Particular emphasis has been placed on DNA adducts, since these are postulated to represent initiating events leading to mutation and/or malignant transformation. It has been empirically established that carcinogenic potency of a large number of chemicals bears a rough proportionality to their ability to bind to DNA - the so-called "covalent binding index" - when reacted *in vivo* with DNA (Lutz, 1979). (There are, however, some notable exceptions to this general correlation (Lijinsky and Phillips *et al.*)). Covalent adducts formed in RNA and proteins have no putative mechanistic role in carcinogenesis, but may relate quantitatively to total exposure and activation, and therefore represent dosimeters for both exposure and activating capability. It should be stressed that adducts are an average dosimeter at the time of sampling unless serial measurements are taken. Moreover, the fluids sampled are generally remote from the target tissue, and hence give an indirect measure of the adducts of interest.

In the case of DNA and RNA, it is known that covalent adducts have differing levels of stability. Some are removed spontaneously through depurination, for

Figure 3

FORMATION OF COVALENT ADDUCTS OF CARCINOGENS AND MUTAGENS



example, whereas others are removed enzymatically in the process of DNA repair. A few are known to remain in DNA for long periods of time (Croy and Wogan, 1981). In the few experimental models in which appropriate measurements have been made, adducts removed spontaneously or enzymatically from DNA have been excreted in urine in amounts that were reflective of total binding levels (Bennett and Essigman, 1981). In contrast, those protein adducts that have been examined were stable over the lifespan of the protein, and therefore accumulated over time to give an integrated measure of exposure (Ehrenberg and Osterman-Golkar, 1980).

These properties collectively form the basis for several complementary approaches to development of chemical dosimeters, each with its own characteristics, providing different kinds of information. Measurement of DNA adducts *in situ* in the DNA of cells should give the most direct evidence of the average genotoxic exposure of that tissue. Measurement of DNA adducts (or products of them) in urine may give an indication of recent exposure. Protein adducts, by contrast, may provide an index of total exposure integrated over the life-span of the target proteins. This rationale is reflected in the experimental approaches summarized in the following four sections.

2. Protein Adducts

Background

Ehrenberg and Osterman-Golkar (1980) reviewed the rationale and technical requirements for the use of protein alkylation for detecting mutagenic agents. Important among these requirements is that exposure must result in the formation of stable covalent derivatives of amino acids for which assay methods of adequate sensitivity and specificity can be devised. Further, the target protein should be found in easily accessible fluids (e.g., blood), and should be present in concentrations adequate to provide sufficient material for analysis. Among

the amino acids likely to be alkylated following exposure are cysteine, histidine, the N-terminal amino acid of the protein, and lysine. Although any protein could be used for monitoring of alkylated derivatives of these amino acids, hemoglobin was suggested by Osterman-Golkar et al. (1976) as a suitable dose-monitoring protein, and virtually all of the available literature on this subject concerns studies of hemoglobin alkylation.

Osterman-Golkar et al. (1976), established the stability of alkylated residues in hemoglobin modified by ethylene oxide or dimethylnitrosamine, and the equivalence of the half-life of alkylation levels produced by a single dose to the life-span of hemoglobin in the mouse. Alkylation of hemoglobin in mice treated with vinyl chloride was reported by the same investigators (Osterman-Golkar et al., 1979). Segerbeck et al., (1978) further characterized the experimental model in mice treated acutely and chronically with the direct-acting alkylating agent methylmethane sulfonate in which they demonstrated the validity of the steady-state level of alkyl residues in hemoglobin as a measure of chronic repeated exposure.

Subsequently, Calleman et al., (1978) carried out a study of hemoglobin alkylation in people occupationally exposed to ethylene oxide. Blood samples were obtained from persons exposed to doses of ethylene oxide established through continuous air-monitoring. Hemoglobin was analyzed for the presence of N-3-(2 hydroxyethyl) histidine by mass spectrometry and by ion-exchange amino acid analysis. The authors concluded that the hemoglobin alkylation values accurately reflected exposure, and were in good agreement with data derived earlier for ethylene oxide in the mouse.

Farmer et al., (1980) have developed a high-resolution GC/MS method for estimating the production of S-methylcysteine in hemoglobin following exposure to methylating agents. This method was used to study in vivo alkylation of

hemoglobin in rats dosed with methylmethane sulfonate (MMS), dimethylnitrosamine (DMN), and the antitumor agent 5-(3,3-dimethyl-1-triazeno) imidazole-4-carboxamide (Bailey et al., 1981). A linear dose-response curve for MMS was observed over a 100-fold dose range, but the dose-response curve for DMN was non-linear. No alkylation was observed with the anti-tumor agent, but may have been overshadowed by a low level of naturally-occurring S-methylcysteine which was found to be present in hemoglobin of the rat and 13 other animal species. These findings emphasize the importance of careful dose-response studies in animals for each compound for which human exposure data is to be sought by this approach. Farmer et al., (1982) have also devised a GC/MS method for the detection of hydroxypropyl histidine in hemoglobin as a measure of exposure to propylene oxide.

Pereira and Chang (1981) surveyed the ability of carcinogens and mutagens representing a broad spectrum of chemical classes to bind to hemoglobin in rats. Animals were dosed with ^{14}C-labeled test compounds at levels of 0.1 to 10 mols/kg body weight and blood was collected 24 hours later. Covalent binding was determined by analysis of purified hemoglobin for radioactivity. Their results are summarized in Table 4, in which compounds are arranged within classes in order of decreasing binding capability. All carcinogens/mutagens had radioactivity associated with hemoglobin, but the extent of this association varied over a wide range ("binding index" from 10.2 to 3320). It should be emphasized that the magnitude of this index is not reflective of potency as carcinogens for the rat. In those instances (11) in which the compounds were administered at more than one dose, the hemoglobin binding index was also dose related. The authors conclude, therefore, that this approach has potential applicability for dosimetry of some environmental carcinogens at ambient levels of exposure. Some association was also shown for two noncarcinogens evaluated.

Advantages and Disadvantages

Detection of covalent protein adducts in proteins such as hemoglobin and

Table 4

COVALENT BINDING IN VIVO OF ^{14}C-LABELED CARCINOGENS/MUTAGENS TO RAT HEMOGLOBIN

CARCINOGENS	BINDING INDEX*
Direct-Acting Carcinogens	
Methylmethane Sulfonate	3320
Methylnitrosourea	302
Ethylnitrosourea	217
Methylnitro-nitrosoguanidine	180
Carcinogens Requiring Metabolic Activation	
Dimethylnitrosamine	697
Benzo(a)pyrene	181
3-Methylcholanthrene	165
Diethylnitrosamine	153
Benzene	144
Acetylaminofluorene	108
Dimethylbenzanthracene	99.8
Aflatoxin B_1	39.4
Aniline	24.8
Benzidine	22.5
Chloroform	14.2
Carbon Tetrachloride	10.2
Non-Carcinogens	
Naphthalene	28.3
Phenol	0.3

*Binding Index: pmol bound/g Hemoglobin/ mol/kg body weight

From: Pereira and Cheng (1981)

albumin have several advantages as dosimeters. In those models studied thus far, the adducts once formed remain stable over the lifespan of the erythrocyte in the species being studied (45-50 days in the rat, 120 days in man). Thus, hemoglobin adducts represent an accumulating dosimeter which integrates exposure over a substantial period of time. Furthermore, the blood proteins (hemoglobin and albumin) are present in relatively large amounts in an accessible body fluid which can be readily obtained in sufficient quantity for analysis.

Disadvantages include the fact that little is known about the pharmacokinetics of activated carcinogens and mutagens, and adduct levels in hemoglobin may not be accurately reflective of actual exposure of target tissues in which the activated compounds are being formed. Additionally, for certain moieties (e.g., methyl and ethyl residues), a background of alkylation of unknown origin seems to exist which limits the sensitivity of detection for some alkylating agents.

3. DNA/RNA Adducts

Background

The scheme summarized in Figure 3 illustrates the scientific rationale for the measurement of DNA adducts in chemical dosimetry. It is apparent that two experimental avenues are available to obtain information on levels of DNA adducts formed in a given set of circumstances. On the one hand, measurements could be made of the levels of DNA adducts derived from a chemical of interest in cells of an accessible tissue (e.g., white blood cells, biopsy, or autopsy samples). Providing that the chemical nature and stabiltiy of the DNA adducts for the compound of interest had been fully characterized, qualitative as well as quantitative identification of adduct levels could provide for that individual not only an indication of exposure history but also of his capability to activate the carcinogen to DNA-binding forms. This issue is greatly complicated by differential rates of DNA repair.

A second approach to monitoring DNA adducts is also being explored, taking advantage of the fact that adducts removed from cellular DNA (and also from RNA) are excreted in urine. Their detection and measurement of excretion rates would, in principle, provide information on (recent) exposure history of the subject, and possibly also indications of that individual's capability for DNA repair. Thus, studies of urinary excretion of adducts would provide data complementary to measurement of adduct levels in cellular DNA in the same individual. Experimental progress in this area is summarized below.

a. Analysis of Cellular DNA

Among several methodologies being developed for analysis of DNA for the presence of carcinogen-derived adducts, two recent approaches can be differentiated by the manner in which the adducts are detected - by immunological techniques in one case and by radiochemical labelling in the other.

For immunological detection, antisera have been raised in rabbits against some of the RNA and DNA adducts of aromatic amine, polycyclic aromatic hydrocarbons, aflatoxins, and methylating and ethylating carcinogens. High affinity antisera have been elicited with either nucleoside adducts covalently bound to a protein carrier, or modified DNA electrostatically coupled to a protein carrier. The properties of these antisera have been reviewed by Poirier (1981). Monoclonal antibodies have also been produced that bind carcinogen-DNA adducts with high affinity. These are reviewed by Muller and Rajewsky (1981).

Characteristics of currently available antibodies against carcinogen modified DNA are summarized in Table 5. Two preliminary studies are in progress which utilize these antibodies in attempts to determine the occurrence of benzo(a)pyrene adducts in white blood cells and tissues of individuals who might have received substantial exposure to this compound, such as roofers, shale oil workers, and lung tumor patients. Experience to date indicates that the methodology has potential value for epidemiologic studies, but that significant

Table 5

ANTIBODIES AVAILABLE AGAINST DNA MODIFICATIONS INDUCED BY CARCINOGENS[1]

CARCINOGEN	DNA ADDUCT	AFFINITY CONSTANT (L/mol)	SENSITIVITY	ASSAY
N-NITROSO- AND	Me-O^6-Gua	2.7×10^{10}	60 fmol	RIA[2]
ALKYLATING	Et-O^6-Gua	2.0×10^{10}	40 fmol	RIA
AGENTS			10 fmol	ELISA[3]
	Bu-O^6-Gua	2.7×10^{10}	40 fmol	RIA
	Et-O^4-Thy	1.3×10^{9}	240 fmol	RIA
	Bu-O^4-Thy	8.8×10^{8}	450 fmol	RIA
	Bu-O^2-Thy	1.1×10^{10}	70 fmol	RIA
AROMATIC AMINES	AAF-C^8-Gua	6.0×10^{9}	100 fmol	RIA
			5 fmol	ELISA
PAH	B(a)P-DNA	1.1×10^{8}	5 pmol	RIA
			5 pmol	ELISA
MYCOTOXINS	AFB_1-DNA		50 fmol	ELISA
	AFB_1-N^7-Gua DNA[2]		100 fmol	ELISA
	AFB_1-FAPY-DNA[2]		100 fmol	ELISA
	AFB_1-N^7-Gua[2]		100 fmol	ELISA
	AFB_1-FAPY		100 fmol	ELISA
UV-LIGHT	THYMINE DIMER-DNA		80 fmol	RIA

[1]Modified from IARC/IPCS Working Group Report (1982)
[2]RIA (Radioimmunoassay): 50% tracer-antibody binding by adduct
[3]ELISA (Enzyme linked immunosorbent assay): Sensitivity is that giving 50% inhibition in competitive assays

problems remain to be resolved. These problems and some suggestions for potential areas for field studies are discussed in the IARC/IPCS Working Group Report (1982).

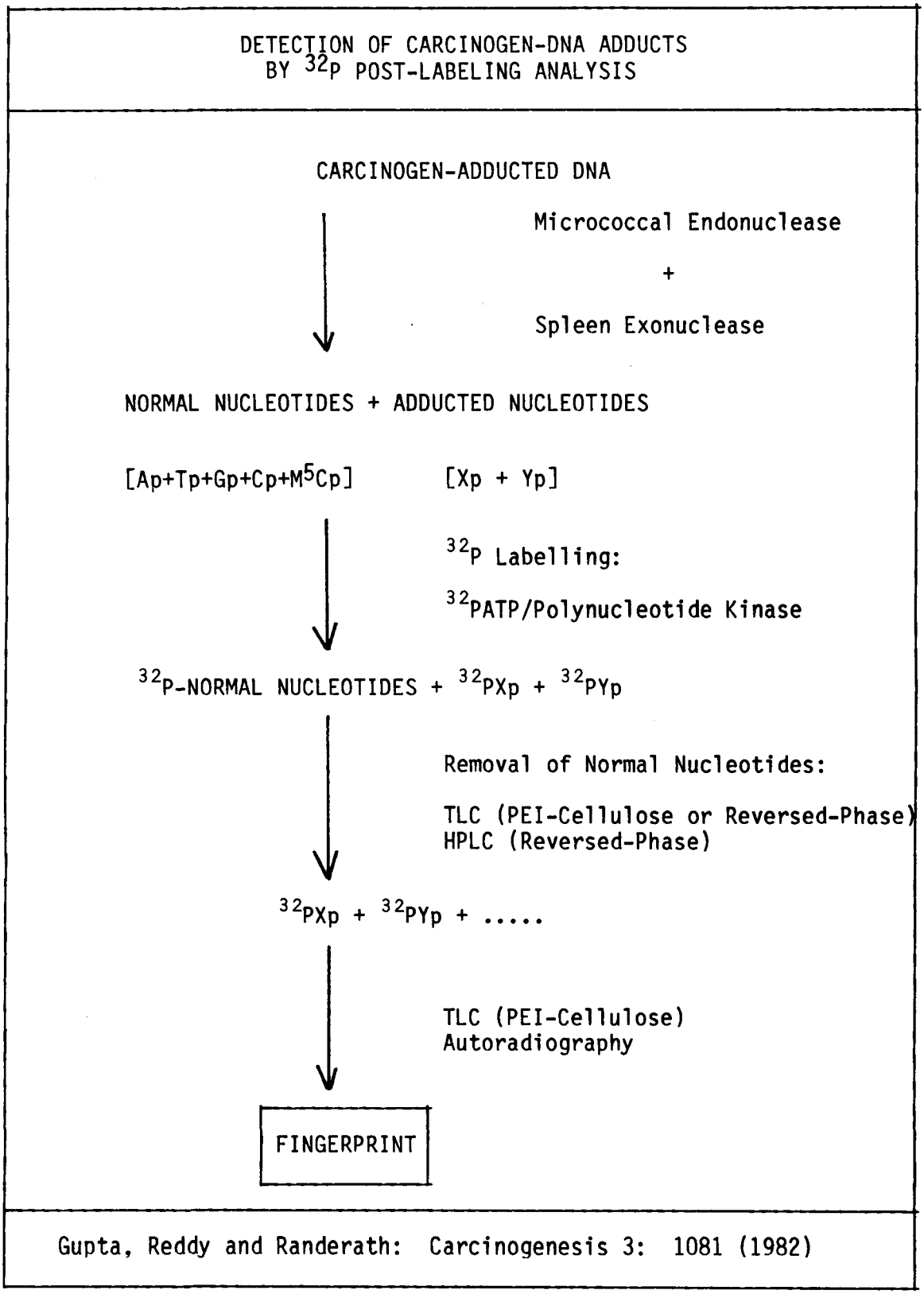
Post labeling methods to detect and characterize carcinogen-DNA adducts have been described by Gupta et al., (1983) and Haseltine et al., (1983). The experimental strategy and procedures involved in the method are summarized in Figure 4. Carcinogen-adducted DNA is subjected to enzymatic analysis under conditions which, when carried to completion, produce a mixture of normal and adducted nucleotides, with the phosphate localized on the 3' position of deoxyribose. These nucleotides are then subjected to phosphorylation through the action of polynucleotide kinase, using ^{32}P-ATP as the source of ^{32}P. This substrate can be obtained in extemely high specific activity, so that nucleotide are radiolabeled in the 5'-deoxyribose position also at a high specific activity. Unmodified nucleotides are removed by TLC or HPLC, and the mixture of adducted nucleotides is resolved on 2-dimensional TLC and subjected to autoradiography. The presence of adducts and semi-quantitative estimation of their levels is achieved by densitometry of the autoradiograms.

Gupta et al., (1982) have applied this method to studies of DNA modified by bulky aromatic carcinogen adducts formed in vivo with derivatives of 2-aminofluorene, acetyl-aminofluorene, and benzo(a)pyrene. In its most sensitive form, the method showed great sensitivity, with the ability to detect adducts at a frequency of 1 in 10^7 - 10^8 DNA bases. It is much less sensitive with small adducts ($1x10^{-4}$, 10^{-5}).

b. Urinary Excretion of Adducts

Exploitation of the detection of DNA, RNA and protein adducts in urine for dosimetry has been undertaken in two experimental models. Bennett et al., (1981) found that rats dosed with aflatoxin B_1 excreted in their urine a large

Figure 4



Gupta, Reddy and Randerath: Carcinogenesis 3: 1081 (1982)

fraction (about 35%) of the major N^7-guanine adduct of the carcinogen over the 48 hours after a single injection. The method allowed reproducible quantitative measurement of adduct in urine from rats treated with doses as low as 0.125 mg/kg body weight. Further, application of the method in rats treated with different doses of carcinogen showed that the amount of adduct excreted bore a constant relationship to peak adduct levels in the livers of animals treated with the same doses. Thus, one condition of adequacy as a dosimeter was met for one chemical, viz, quantitative reflection of adduct levels in the target tissue.

However, the limit of sensitivity of the method was still inadequate to detect ambient exposure levels of aflatoxins known to occur in human populations, and further methodologic development was required. Donahue et al., (1982) reported improvements that attain that objective. These improvements consist of modifications in the chromatography, but most importantly substitution for the absorbance measurement of the earlier method radiolabeling of the adduct with ^3H-dimethylsulfate and ultimately determination of the radioactive product, ^3H-7-methylguanine.

Experiments with similar objectives were carried out with the carcinogen dimethylnitrosamine by Hemminki and Vainio (1982) and Hemminki (1982). Rats were injected with ^{14}C-dimethylnitrosamine and urine was collected over the succeeding five days. Radioactivity was extracted and separated by Sephadex G-10 chromatography, and the main DMN-derived adducts were tentatively identified. These included N-acetyl-S-methylcysteine, 1-methylhistamine, S-methylcysteine, and methionine, allantoin and 7-methylguanine. Although dose-response experiments were not carried out, these results illustrate the potential applicability of the approach to human monitoring if suitable detection methods for the adducts can be devised. They also illustrate the complexity of the adduct mixtures formed by alkylating agents such as DMN.

Advantages/Disadvantages

It is well established that all carcinogens that have been shown to form adducts form a complex spectrum of DNA adducts, involving covalent binding to various nucleophilic sites on all four DNA bases as well as on the phosphate residues of DNA (Singer and Grunberger, 1983). Thus, from a qualitative viewpoint, detection of all DNA adducts derived from even a single carcinogen can present a very complex analytical challenge. Quantification of adduct levels is complicated even further by the fact that adducts are removed from DNA by chemical or enzymatic processes at different rates, even within the same cell; these rates can also vary substantially from one cell type to another.

Most of the currently available information on DNA adducts in experimental systems has been obtained through the use of physicochemical or radiochemical methods of detection. Usefulness of these methods for detection in human monitoring is limited by their relative insensitivity and inapplicability, respectively. However, immunological techniques are being developed which have promise of utility in detecting DNA adducts in people exposed to environmental carcinogens under ambient conditions. The determination of carcinogen-DNA adducts by immunologic procedures has certain advantages over other techniques. The sensitivity is frequently better than that obtainable with radiolabeled carcinogens (which are useful only for experimental purposes in any event). Antibodies are specific for particular three-dimensional structure and can be used to probe the conformation of unknown adducts on DNA. However, this specificity introduces a major complication in development of immmunoassays to be used in dosimetry, in that a different antibody must be produced for each adduct to be detected.

Immunologic assays are rapid, highly reproducible and can be used in situations where the cost or availability of radiolabeled carcinogens would be prohibitive. The high sensitivity and capability of detecting non-radioactive

adducts would therefore suggest a use in monitoring of human tissues. In addition, immunologic techniques can be applied together with morphologic procedures (electron microscopy and immunofluorescence) to localize adducts in particular cells, subcellular compartments, or DNA molecules.

Detection of nucleic acid adducts by the post-labeling technique has great potential as a method for application to human population studies. However, much further development will be required. For example, in its present form, the method is only applicable to DNA modified by bulky, aromatic carcinogens, and new hydrolysis and separation techniques will be required to enable its application to DNA alkylated by small moieties such as methyl or ethyl groups. Furthermore, identification of individual adducts will be impossible until reference standards for each adduct are made available and their properties are determined. Since, as discussed earlier, each carcinogen forms a complex mixture of DNA adducts, this development will require much additional research effort. In addition, specific technical problems remain to be overcome, such as a differential in labeling efficiencies for different nucleotides, especially those structurally modified by adducts.

Recommendations

It is clear that the complexities of this field necessitate extensive further research in method development and careful validation in animal models before interpretable data can be obtained from studies in human populations. However, the potential usefulness of the information to be gained justifies the additional research effort, as discussed in several recent reviews (IARC/IPCS Working Group Report, 1982; Perera and Weinstein, 1982; Weinstein, 1983).

We therefore recommend that the NTP introduce into its program studies on covalent adduct formation in animals treated with chemicals under the carefully controlled conditions of the bioassay protocol. Similarly, clinical monitoring of adduct formation should also be approached on a sound scientific basis. At

this stage, requisite methology is available for only a limited number of model compounds, but studies on them would provide extremely valuable data for validation of these methodologic approaches and interpretation of data obtained by parallel studies being conducted in human populations. In this context, the NTP program represents an extremely valuable, perhaps unique, source of such quantitative data in animals.

DNA Damage and Repair

Background

The interaction of chemicals with DNA is generally accepted to be a requisite step for the initiation of chemical carcinogenesis. A number of techniques have been developed to quantitate chemically-induced DNA damage and its subsequent repair. These recently have been reviewed in detail (IARC 1980; Larsen et al, 1982; Mitchell et al., 1983). It should be noted that DNA repair or misrepairs can in theory induce gene mutation and that a negative response may indicate that a substance does not elicit excision repairs but may still be mutagenic. While some universality in biologic systems is to be expected, there is currently little or no evidence supporting a role of misrepair in mammalian cells. Briefly, methods for determining DNA damage include measurement of specific DNA adducts (covered elsewhere in this document), single strand breaks, alkali sensitive sites, and DNA cross-links. Several of these same methods can be used to follow DNA repair. Such repair usually includes loss or excision of the altered regions of DNA, followed by resynthesis of a small segment of DNA. This resynthesis of DNA provides a mechanism for identifying DNA repair, since radiolabeled precursors can be incorporated to monitor the process, commonly referred to as unscheduled DNA synthesis, (UDS), (Mitchell et al., 1983). UDS can be measured in a wide variety of systems including cultured cells from animals and humans, and tissues exposed to chemicals in vivo, but assayed in vitro

(Mirsalis et al., 1982). The most extensively studied technique is hepatocyte UDS, where hepatocytes of rats are freshly isolated and placed in culture, exposed to the test substance and ^{3}H-thymidine, and subjected to autoradiographic analysis (Mitchell et al., 1983). This system has the distinct advantage that the cells under test are metabolically competent. Furthermore, hepatocytes from any species can be utilized. The method has also been modified so that the test substance is administered to the host animal prior to hepatocyte isolation and culture, allowing all of the host's metabolic capacity to interact, including enterohepatic circulation and metabolism by intestinal microflora (Mirsalis et al., 1982). Preliminary data are also evolving from systems employing other tissues and cell types, including human hepatocytes (Butterworth et al., 1982).

Advantages and Disadvantages of DNA Damage and Repair Tests

1. DNA Damage

DNA damage assays may be more relevant to carcinogenesis than UDS, since it is generally accepted that mutation results from the presence of adducts during DNA replication, not from the repair of such lesions. The alkaline elution technique has been adapted as a screening method using various cell lines with and without metabolic activation and more recently freshly isolated hepatocytes (Sina et al., 1983). While the latter system requires further validation in other laboratories it appears to be a rapid and accurate method deserving further consideration. Alkaline elution or sucrose gradients also can be used on tissues from a variety of organs from animals exposed to test substances, although only a small number of laboratories have taken this approach (Larsen et al., 1982).

2. Unscheduled DNA Synthesis

Several laboratories have utilized the rat hepatocyte UDS test (Mitchell et al., 1983). Its major advantages include the use of metabolically competent

cells and rather straightforward methodology. Interspecies comparisons are possible with this method; however, care must be taken to maintain similar viability and metabolic competence. The *in vivo/in vitro* hepatocyte UDS test also identifies chemicals requiring metabolism by the intestinal microflora and provides some information on the mitogenic response to the test substance (Mirsalis *et al.*, 1982). Disadvantages of the hepatocyte UDS system include a lack of correlation with carcinogenic potency. (This is true of most short term assays.) They also include the fact that groups of chemicals like the nitroaromatics are not identified unless the *in vivo/in vitro* system is employed; and a limited sensitivity due to high background in the *in vitro* hepatocyte UDS autoradiographs. Finally, UDS methods that employ scintillation counting must use agents that suppress scheduled DNA synthesis. These agents could affect other metabolic processes. Furthermore, scintillation counting eliminates information on the mitogenic response.

Recommendations

The rat hepatocyte UDS test using metabolically competent cells should remain on NTP's standard battery of short-term tests. Chemicals of specific classes known to require metabolism by the intestinal microflora should be tested in the *in vivo/in vitro* test. In addition, when a chemical induces hepatocellular tumors in the bioassay, but is negative in the *in vitro* hepatocyte UDS test, it should be tested in the *in vivo/in vitro* test. This will assist in understanding the mechanisms of liver carcinogenesis in cases involving microbial metabolism. Interspecies comparisons between human rodent hepatocytes using chemicals known to be carcinogenic to humans and animals will further validate this system. Unless unique advantages of other tissues become apparent, minimal effort should be exerted in this direction. One such area, however, may be the use of urothelial cells (Oglesby *et al.*, 1983).

The lack of quantitation between carcinogenic potency and UDS response makes UDS a less appropriate choice for studies on internal dosimetry than studies on specific adducts. (However, as already mentioned, the correlation between specific adducts and carcinogenic potency may not be perfect.) Nonetheless, if compound related UDS is present, it represents a qualitative indication of potential genotoxicity.

Micronucleus Test

Background

The micronucleus test detects chromosomal fragments that were not incorporated into the cell's nucleus at the time of cell division. It represents a simpler method for detecting clastogens than conventional techniques of chromosomal analysis. As discussed in recent reviews, over 150 chemicals have been evaluated using the micronucleus test, with approximately 50% of the carcinogens causing an increase in micronuclei (Heddle et al., 1983; Jenssen and Ramel, 1980; Tsuchimoto and Matter, 1979; and Heddle and Bruce, 1977). More stringent test protocols have been proposed which expand the number of animals and polychromatic erythrocytes to be examined and extend the period between chemical administration and evaluation for micronuclei (Heddle et al., 1983). These improvements are expected to increase the sensitivity of the micronucleus test. New staining procedures also promise to reduce animal to animal variation and make the technique amenable to automation (Hayashi et al., 1983; MacGregor et al., 1983). While the conventional "micronucleus test" utilizes polychromatic erythrocytes derived from bone marrow of exposed animals, new methods are being developed circulating erythrocytes and for other tissues, including liver and transplacentally exposed fetuses.

Advantages and Disadvantages of the Micronucleus Test

The major advantage of the micronucleus test is that it is a relatively simple in vivo method that can provide comparisons with in vitro or in vivo SCE

or chromosomal analysis. While only limited numbers of studies have used animals other than mice, the available data across species (including man) have been remarkably consistent (Heddle et al., 1983). The method has few false positives and these should be decreased even more by applying newer staining procedures. The latter also should be amenable to automation, such as flow cytometry. Few laboratory-to-laboratory discrepancies have occurred. The micronucleus test could be carried out on the same animals undergoing prechronic toxicity studies.

Major disadvantages of the micronucleus test include its lack of sensitivity, (i.e., requiring doses near the LD_{50}) and its inability to detect tissue specific chemicals such as those requiring specialized metabolism or having limited distribution. Furthermore, the micronucleus test requires dividing cell populations. Artifacts such as stain precipitate and leukocyte or mast cell granules can increase the difficulty in identification of micronuclei. Lastly, the micronucleus test is not well suited for evaluating cumulative biological dosimetry, since these cells appear to be preferentially removed by the reticulo-endothelial system. This is a major limitation in applying the test to human erthrocytes but it may be applicable to other human cells (e.g. lymphocytes) and is now applicable to rodent studies.

Recommendations

Presently, the micronucleus test is not employed in the NTP genetic toxicology screen. New test protocols and staining procedures should increase its usefulness. In light of this, it is recommended that the NTP evaluate the micronucleus test as a secondary or tertiary screen and if useful consider using automated methods. When compounds produce inconsistent results in a battery of short term tests, blood samples could be collected during prechronic testing for analysis of micronuclei. It also would be useful to evaluate with this technique any chemical that has induced leukemia.

Sister Chromatid Exchange (SCE)

Background, Method, and Significance

SCEs involve symmetrical exchange between sister chromatids at one locus, in contrast to the asymmetrical exchange in chromatid aberrations, and thus do not result in any alteration in chromosome morphology. These exchange events can be visualized in cells that have gone through two cell cycles in the presence of 5-bromodeoxyuridine (BrdUrd) and stained with a fluorescent dye and/or Giemsa to give differential staining of the sister chromatids. Fluorescence in the Brd-Urd containing DNA fades, and the SCE events appear as switches in staining along the length of the chromosome that are reciprocal between homologous chromatids (Taylor, 1982; Perry, 1982, IARC, 1980).

There have been numerous recent reviews of SCEs which decribe the test methods and standard protocols used in vivo and in vitro assays in rodents and humans (Bloom, 1981; IARC, 1980; Carrano, 1982; Evans, 1982; Lambert et al., 1982; U.S. E.P.A. Gene-Tox Program, Latt et al., 1981; Perry, 1982; Taylor, 1982; Wolff, 1982). Significant aspects of the test procedures are summarized from these sources and described below.

Little is known about the molecular mechanism of SCE formation, except that the exchange process must involve DNA breakage and rejoining (Wolff, 1982). Despite this uncertainty, SCEs have been widely used to identify exposure to genotoxic agents. Many clastogens and mutagens produce SCEs, particularly those that form DNA adducts or distort the DNA backbone (Carrano, 1982; Wolff, 1982). For most substances tested, the induction of SCEs is linearly dependent upon dose over the dose range tested. Moreover, induced SCEs and single-gene mutations demonstrate linear relationships in in vitro studies with a variety of physical and chemical agents, although the slope of the SCE-mutation relation differs for each agent (Carrano, 1982; Carrano and Thompson, 1982). Because of this, and because many carcinogens and mutagens produce SCEs in in vitro

systems, the analysis of SCEs is widely used as an assay for genetic damage (Latt et al., 1981).

SCE versus Chromosome Aberrations

Differences and overlaps exist between chromosome aberrations and SCEs but most evidence now indicates that the two cytogenetic endpoints are produced by different lesions and/or different processes. Thus, assays for SCEs and standard cytogenetic analysis tend to complement one another. For example, induction of SCEs is an S-phase phenomenon, whereas aberrations can be induced by radiation or in any phase of the cell cycle (Wolff, 1982). Lymphocytes from patients with ataxia telangectasia and Fanconi's anemia are characterized by increased chromosome aberrations but are without increases in SCEs. Alternatively, lymphocytes from patients with Bloom's syndrome, another chromosome instability syndrome, have 10-15 fold higher SCE frequencies than controls, but the distribution of SCEs in cells does not correspond to that of chromosome aberrations (Perry, 1982; Evans, 1982).

Agents that cause double-strand breaks (e.g., x-rays and bleomycin) are excellent producers of chromosomal aberrations but only inefficiently induce SCEs. Conversely, many alkylating and mutagenic agents cause increases in SCEs without causing chromosomal aberrations, or require much higher doses to do so. In general, the SCE assay appears to be more sensitive, responding to a wider variety of agents and/or requiring 10-100 fold lower doses (Perry, 1982).

Evidence Linking Increased SCE Frequency with Increased Risk of Cancer

There is information from a variety of sources that indicates a link between elevated SCE frequencies and exposures to carcinogens as well as some data on elevated SCE frequencies among individuals at increased risk of cancer.

Humans

The frequency of SCEs in lymphocytes from patients with Bloom's syndrome, an autosomal recessive disease that predisposes individuals to increased risks of

malignancy, was 10-15 times that for lymphocytes from normal individuals (Evans, 1982b). The frequency of SCEs in 13 untreated patients with malignant lymphoma, both Hodgkin's and non-Hodgkin's, was significantly higher (12.7/cell) than the mean frequency of SCE's in lymphocytes from 40 healthy control individuals (6.1/cell) (Evans, 1982b). There are exceptions, however, such as patients with Fanconi's syndrome who do not exhibit elevated SCE's. Furthermore, treatment with carcinogens produces increases in SCE incidences. SCE frequencies are elevated in cancer patients following treatment with the chemotherapeutic agents adriamycin, cyclophosphamide, busulfan, melphalan, and various mixtures which themselves are cytotoxic and carcinogenic (Evans, 1982; Perry, 1982). Peripheral lymphocytes from cigarette smokers who smoke more than 20 cigarettes/day generally have a 20-30% increase in SCE frequency in comparison to lymphocytes from suitable control non-smokers (Lambert et al., 1982).

Occupational exposures to a number of chemicals which are known to be carcinogenic in humans or test animals and/or mutagenic in various short term tests also produce higher levels of SCE's in peripheral lymphocytes from exposed workers. For example, production workers exposed to vinyl chloride monomer and laboratory workers exposed to organic solvents developed higher SCE levels in peripheral lymphocytes (Evans, 1982), as do workers with significant asbestos exposures (Rom, 1983). Exposures to ethylene oxide produced similar significant increases among hospital and sterilization plant workers (Yager et al., 1983; Garry et al., 1979).

Occupational exposures to other agents produce results that are less clearly positive but are strongly suggestive of an effect on SCE incidence. For example, exposure of production workers to styrene or of hospital personnel to chemotherapeutic drugs have a measurable effect on SCE frequencies (Lambert, 1983; Sorsa, 1983). A group of 11 nurses handling cytostatic drugs in a cancer

clinic had a small but significantly increased frequency of SCE's in peripheral lymphocytes when compared to a control group of 10 female hospital clerks (Lambert, 1982).

At present, it is not clear that most occupational exposures produce significant increases in SCE incidences. For a number of studies, no measurable or significant increases have been reported (Lambert, 1982). This may be because the SCE assay is not responsive to these dose levels (i.e., for most chemicals the in vivo assay may require exposures to large doses before a significant increase in SCE incidence is produced). Alternatively, the study design may have conferred limited sensitivity to the assay, because of the small numbers of subjects tested in most studies. This clearly limits the importance of some of the negative results.

Animals

Rodents exposed to urethane or related carcinogenic carbamate derivatives showed increases in SCE frequencies which paralleled the increases in tumor incidence in neighboring cell populations (Connor, 1983; Cheng et al., 1981, and 1982). For a number of mutagenic or carcinogenic compounds tested in Chinese hamster ovary (CHO) cells in culture, increases in mutation frequencies at the HGPRT locus were paralleled by increases in SCE frequencies (Carrano, 1982).

Many mutagens and carcinogens produce increases in SCEs in human or rodent cells in culture. Of the 29 chemicals listed in the Gene-Tox program with adequate evidence for carcinogenicity, 25 produced positive responses in the SCE assay (Latt et al., 1980). However, the paucity of non-carcinogens on the list of chemicals tested for both SCE and carcinogenicity severely limits the assessment of the predictive value of SCE assay or carcinogenicity.

Conclusion

An inferential conclusion can be drawn from the above data. Although, there

is no substantial evidence that neoplastic transformation is accompanied by a dramatic increase in SCE incidence, the evidence in humans from studies of the effects of cigarette smoking, anti-cancer agents, and occupational exposures to carcinogens and mutagens, and in rodent and *in vitro* test systems, all support the notion that elevated SCE frequencies may reflect exposure to carcinogens and thus may signal an increased risk of cancer.

General Methodologic Considerations

In analyzing the data from the SCE assay, there are three major considerations in interpreting the significance of the results: possible SCE artifacts; possible insensitivity to low level chronic exposures; and the dependence of SCE on time and dose.

First, the incubation with BrdUrd alone can induce SCEs and may be responsible for the baseline level of SCEs in untreated cells (Wolff, 1977). Although, this introduces variability into the assay and has cast doubt on the significance of some of the results reported in the past, the effects of BrdUrd concentration and time of incubation on SCE formation have been well-studied, and standardized protocols now exist which greatly enhance the reproducibility of tests and reduce inter-lab variability. (For review, see Latt *et al.*, 1981; Kligerman *et al.*, 1981; Carrano and Thompson, 1982; Lambert *et al.*, 1982; Erexson *et al.*, 1983; Kligerman *et al*, 1982).

Second, for a chromosome lesion to be converted to an SCE, the lesion must be present during DNA synthesis when BrdUrd is also present. Circulating lymphocytes are in an arrested state at G_0 and must be incubated with a mitogen such as phytohemaglutinin (PHA) to initiate DNA synthesis. Routinely, 24 hours pass before such lymphocytes enter into S-phase, and lesions which are repaired during this one-day interval would not get expressed as SCEs. Similarly, the SCE-inducing lesion must persist, and not be removed by repair processes or by

lymphocyte turnover in the circulating blood, from the time of exposure *in vivo* until samples are taken for *in vitro* incubation with PHA or BrdUrd. For these reasons, the SCE assay may not give a response for chronic, low-level exposures or in blood samples taken long after an acute exposure. (Evans 1982a; Lambert *et al.*, 1982; Yager, 1982).

Third, it is important to note that the time course of appearance and disappearance of SCEs seems to depend upon the severity of the exposure. For example, elevated incidences of SCEs may decline within hours following exposure to 1-2 cigarettes, or take several weeks following high exposures to cancer chemotherapeutic drugs. In humans, studies with smokers indicate that a low baseline frequency of SCEs is obtained following 18 hours of non-smoking, which remains unchanged immediately following the first cigarette. Six hours later after six cigarettes, a small increase in SCE frequency was noted which returned to baseline levels following an 18-hour period of non-smoking (Lambert *et al.*, 1982).

In general, it seems the greater the dose, the longer the elevated SCE levels are found to persist. With higher exposures, such as occur in cancer patients treated with chemotherapeutic agents, the SCE frequencies remain elevated for long periods of time, returning to pretreatment values a few months (4-16 weeks) after cessation of therapy (Lambert, 1982). A surprising finding is a three-fold increase in SCE frequencies relative to controls in patients who were treated with potassium arsenite more than 24 years before and continue to show clinical symptoms of arsenic treatment, including skin cancer (Evans, 1982).

Methodologic Considerations Re: Parallel Testing in Rodents and Humans

In studying the formation of SCEs in circulating lymphocytes, one is making the assumption that the blood is a responsive tissue that signals damage

occurring throughout the body. Thus, by examining blood cells one hopes to indirectly monitor effects that are occurring at distant and inaccessible tissues, and to obtain an indication of genetic damage to the individual. There are, however, several uncertainties (generic to all short term methods discussed) regarding this assumption:

First, as is the case with most, if not all, assays discussed, there are few studies to date in animals or humans that continuously monitor effects of chronic exposures on SCE frequency in lymphocytes, and in most studies samples are taken at one point in time. A notable exception is the Johnson and Johnson study in which ethylene oxide workers have been followed for 2 years (Johnson and Johnson, 1982). In vivo studies with human lymphocytes have received more attention than with rodent lymphocytes in part because they are easier to grow in culture. However, procedures have recently been developed which apparently make culturing of rodent lymphocytes quite manageable (Kligerman, 1982; Bishop; 1982).

Second, chronic administration of a chemical, such as would occur in the proposed bioassay, would in theory produce steady-state levels of SCEs which reflect the equilibria of lesions formed and lesions repaired in response to repeated daily doses. It is not known what the comparative sensitivity of bone marrow or lymphocytes would be in this situation, although studies in humans indicate the mean SCE frequency in bone marrow cells is lower (4-6/cell) than that for blood lymphocytes (8/cell) (Evans, 1982). In acute studies, the responses of B- and T-lymphocytes do not appear to be markedly different (Lambert et al., 1982).

Third, the extent of individual variation in SCE frequencies is not established but could be an important confounding variable. For example, it isn't known if the variation in SCE frequencies among individual test animals

would be expected to be so great as to obscure differences that may exist between tumor-free and tumor-bearing animals. It is possible that only very large persistent differences would be detectable. In humans, the coefficient of variation (CV) of SCE frequencies for an individual over time is about 10%, and the CV for individuals in a group is about 10-15%. Both depend upon the number of cells analyzed per individual (Carrano et al., 1980; Lambert et al., 1982). Inter- and intra-individual variations would probably be less in inbred strains of test animals held under tightly controlled conditions than in the genetically heterogeneous human population exposed to a variety of background agents. Studies in humans are needed to record inter- and intra-individual variation over time during chronic or repeated exposures to chemotherapeutic agents. To calibrate and compare human and animal responses, it would be very useful to make measurements in test animals.

Fourth, it appears that the time-course of SCE incidence will vary over the rodent's lifetime in an animal bioassay, a factor to consider in designing studies. In humans there do not appear to be major differences in SCE frequencies among adults that are attributable to age or sex. In newborns, however, the mean incidence of SCE in lymphocytes was twice that observed for lymphocytes from adults age 25-35 years (Evans, 1982). Similar studies in mice demonstrate the SCE frequency in newborns to be 2-3 fold higher than in adults. Thus, in a typical bioassay in which 4-6 week-old animals are administered daily doses throughout their lifetime, the SCE frequency may begin at some elevated level, drop to a mature level, and then increase slowly with time in the dosed animals.

Finally, the question remains as to whether the blood is a good indicator of damage that is occurring in other and, presumably, target tissues. To optimize the chances of producing a response in the blood, the test chemical chosen could

be one that affects the bone marrow. If it were an anti-leukemic agent, such as is used in cancer chemotherapy, the potential to compare the time-course and level of response in humans with that of test animals would be attractive, as discussed earlier.

Advantages and Disadvantages

The advantages and disadvantages of the SCE assay have been recently reviewed by the U.S. E.P.A. Gene-Tox Program (Latt et al., 1983). The following is a brief summary of their evaluation.

The major advantages of the SCE assay are:

1. It is efficient at detecting chemicals that alkylate DNA and produce DNA adducts;
2. It gives few false positives with nonmutagenic compounds;
3. It is very sensitive, detecting effects of many compounds at much lower (10-100 fold) concentrations than are required to produce chromosome aberrations;
4. It is capable of detecting both direct-acting compounds and those that need metabolic activation;
5. Although the data are limited, the assay can detect agents that are known or believed to act as promotors (e.g., phorbol esters, saccharin; as well as inorganic agents, asbestos);
6. It is a relatively easy and rapid method, especially using an in vitro system;
7. It is a mammalian test system; it can be used both in vivo and in vitro, enabling comparison of the two approaches. In vivo studies can be performed on multiple tissues, including germ cells, permitting comparisons of tissue sensitivities;
8. It can be used to make comparisons between the responsiveness of humans

and rodents, of *in vivo* and *in vitro* systems, or of different tissues to exposures to test agents; and

9. It can be used for human population monitoring.

The major disadvantages of the SCE assay are:

1. It is relatively insensitive to compounds (e.g., bleomycin) that induce double-strand DNA breaks;
2. The mechanism of SCE formation is unknown; it may represent only a very small part of the damage that has occurred (i.e., the number of mutagenic events may be much larger than the number of SCEs expressed) and this proportion may vary between agents. Thus, a positive SCE response has greater significance than a negative response;
3. A false negative response may result from inadequate activation or the inactivation of the substance before reaching the target tissue; and
4. *In vitro* systems are sensitive to subtle effects, not necessarily occurring *in vivo*, that might induce SCEs (e.g., visible light on some antibiotics, or pH indicators).

Recommendations

1. Continue to employ the SCE *in vitro* assay using Chinese hamster ovary cells in the NTP standard battery of short term tests; and
2. Include an assay for SCEs in the proposed *in vivo* test scheme. Information sought on SCE frequencies would include:
    a. Intra-individual variation over time;
    b. Inter-individual variation within a dose group;
    c. Differences in mean values between dose groups, and the dose-response relationship; and
    d. Levels in tumor-bearing animals compared to levels in tumor-free animals within a dose group.

Chromosomal Aberrations

Background

Definitions

Chromosomal aberrations, probably the clearest cytogenetic manifestation of genotoxicity, belong to the category of "chromosomal effects" which includes sister chromatid exchange, micronuclei, and dominant-lethal embryonic events. Chromosomal aberrations can in turn be broken down into: changes in chromosome number (polyploidy and aneuploidy) and changes in chromosome structure (e.g., exchanges between chromosomes, rearrangements of DNA, or DNA fragmentation). (See IARC, 1980 and 1983; Evans, 1982 and 1983; Bloom (ed.) 1981; Dunkel, 1983; NAS, 1983; and Preston et al., 1981 for review.) Numerical changes may be an important event with respect to carcinogenesis and teratogenesis. NTP has contracted research in yeast and drosophila aimed at answering the questions: "Do chemicals induce aneuploidy and can it be measured reproducibly?" (Tennant, 1983).

Structural chromosomal aberrations, a consequence of misreplication at sites where damage was sustained to the DNA, are observed in metaphase cells. Aberrations are classified as chromosome type if DNA damage occurred in G_1 cells and was translated into breakage/exchange events prior to chromosome replication. Chromosome-type aberrations include gaps, terminal and interstitial deletions, inversions, rings and interchanges. Chromatid type aberrations are somewhat more varied but generally similar to chromosome type. However, they may result from damage in single chromatids sustained at any stage in the cell cycle, but which must be translated into a structural aberration during or after the replication of chromosomes in the S phase of the cycle. The majority of agents to which humans might be exposed are S-dependent; that is, they will induce chromatid type aberrations. This is a problem because, as will be

discussed later, a large percentage of the damage will be repaired before the cell reaches the S phase and hence will not be reflected in breakage/exchange vents (Bloom, ed., 1981).

The in vitro chromosomal aberration studies as routinely conducted use human peripheral lymphocytes (PBLs) or human or rodent (Chinese Hamster) fibroblasts; in vivo studies are generally conducted with lymphocytes and bone marrow cells which are cultured following exposure in vivo (see Bloom, ed., 1981, IARC, 1980, and Preston et al., 1981 for protocols and methods). Several hundred compounds have been tested in the human PBL in vitro system while the Chinese hamster fibroblast in vitro system has been applied to 400 compounds so far (IARC, 1980). The most frequently used system in vivo is with mitogen-stimulated lymphocytes which, because of the long lifespan of T lymphocytes, makes it possible to detect accumulated damage (Vainio et al., 1983). According to a recent IARC review (1980), chromosomal analysis of bone marrow in vivo may lack adequate sensitivity; moreover, this tissue is not readily obtainable for purposes of human studies.

Significance of Chromosomal Aberrations

The rationale for measuring chromosomal aberrations as an indicator of potential carcinogenic risk is discussed by Evans (1983) and IARC (1980). It is based on factors such as:

(1) the finding that a substantial portion of carcinogens are extremely potent in inducing chromosomal damage which can be detected cytologically;

(2) the increasing evidence that chromosomal rearrangement in somatic cells is important in neoplasia; specifically, the fact that certain chromosomal translocations are associated with lymphoproliferative cancers (CML and lymphomas)(Evans, 1983; IARC, 1980) and evidence that chromosomal rearrangement may play a role in oncogene activation;

(3) the relevance of chromosomal mutation to inheritance of genetic defects in germ cells; and

(4) the observation that substances that produce chromosomal structural changes also produce point mutation; and few chemical mutagens produce no aberrations (Evans, 1983).

Animal and Human Data: Summary

Correlation data are limited but encouraging. Preston et al., (1981) reviewed 177 papers regarding six assays for clastogenic effects on mammalian cells and concluded that in general there was good agreement between clastogenicity and animal carcinogenicity. Of 53 agents reviewed, 25 had some evidence of animal carcinogenicity. Twenty of these (or 80%) had positive results in the mammalian cytogenetic leukocyte or lymphocyte assay, while one had mixed results (+/-). For the remaining 28 agents, there was either insufficient or no information available regarding animal carcinogenicity. Thirty-seven compounds were evaluated in more than one assay reviewed in the report; all results correlated well with the reported animal carcinogenicity response (17 of 18 cases).

Review of the results in the in vitro Chinese hamster fibroblast system for 400 compounds reveals a good correlation between the capacity to cause chromosomal aberrations and mutagenicity as well as animal carcinogenicity (IARC, 1980).

The human in vivo data largely relate to ionizing radiation exposures but there are increasing reports regarding chemical exposures. Analysis of chromosomal aberrations is a widely used and accepted radiation dosimetry technique; human studies have illustrated a dose-reponse, a decrease over time following exposure, and persistence of a percentage of chromosomal aberrations up to 28 years following exposure (Evans, 1982). Cigarette smoking is also associated with increased chromosomal damage in PBLs in humans. Table 6 summarizes the positive results that have been obtained in human studies of six different agents or chemical substances. Among the industrial populations exposed to chemicals listed in Table 6, the clearest dose-response was observed in a prospective

TABLE 6

SOME OCCUPATIONALLY EXPOSED POPULATIONS SHOWING CLEARLY INCREASED ABERRATION FREQUENCIES, RELATIVE TO CONTROLS, IN PERIPHERAL BLOOD LYMPHOCYTES*

AGENT	POPULATION SIZE (>)
X OR y-RAYS	1000
BENZENE	190
EPICHLORHYDRIN	100
STYRENE	50
VINYL CHLORIDE	500
ETHYLENE OXIDE	100

*Positive results in more than one study

[Source: Evans 1982, as modified]

---

study by Kucerova et al., (1977), in which workers were sampled prior to exposure to epichlorhydrin, then one and two years afterwards. Other surveys listed were retrospective and provided little or no dose-response data (Evans, 1982).

B. Advantages/Disadvantages

Advantages

1. There appears to be a strong rationale for using these assays as indicators of potential carcinogenicity/genetic toxicity and indicators of biologically effective dose and/or genetic response to chemical agents. (See discussion of "Significance" above).

2. There is considerable experience with the assays in both laboratory and human studies in vitro and in vivo.

3. Chromosomal aberrations have been shown to be a reliable indicator of exposure to agents that break chromosomes directly.

4. Many technical problems have been identified and solved.

Disadvantages

Disadvantages or limitations of short term tests for chromosomal aberrations include the following:

1. Cell cycle dependency is probably the main problem (Preston et al., 1981; NAS, 1983; Evans, 1982; Bloom (ed.), 1981). The majority of chemicals do not break chromosomes directly. Rather, they produce lesions in DNA which are "registered" only after cell replication. If repair of DNA lesions occurs before the cells replicate, the assay will underestimate the true effect. There will also be differences in cellular sensitivity in different stages of the cycle (NAS, 1983; Evans, 1983). All of the above factors create a problem in monitoring low-level exposures to chemicals. In addition, loss of aberrant cells (e.g. with acentric fragments, or exchange aberrations) will decrease the sensitivity of the assay (Bloom, ed., 1981).

2. The assays are influenced by variations in laboratory procedures (a generic problem); the techniques are fairly expensive, time-consuming (scoring of 300 cells per point is necessary), and require supervision by an experienced cytogeneticist (NAS, 1983).

3. As for all assays considered in this report, there is no direct evidence that individuals with elevated chromosomal aberrations are at higher risk of cancer; therefore they cannot be used to predict specific health effects (Preston et al., 1981).

4. Factors in chemical clastogenicity can lead to uncontrolled variability in the results of cytogenetic studies (Vainio et al., 1983, Bloom (ed.), 1981). They include:

   (a) mode and specificity of action;
   (b) persistence of damage;
   (c) possible selectiveness of cell subpopulations responding to mitogen stimulation; and
   (d) introduction of extraneous factors such as viruses and contaminants that can modify the results.

5. There is lack of knowledge regarding the normal or "background" levels of chromosomal aberration, spontaneous aberration frequencies, inter-individual variation, and persistence of lesions.

6. Direct detection of damage in vivo can only be accomplished with samples of dividing human cells (e.g., bone marrow or testicular tissue); because of difficulties in obtaining these samples, circulating white blood cells -- that must be artificially stimulated to divide -- must be used as a surrogate. The effect in lymphocytes will be small in relation to bone marrow or spermatagonial cells (Bloom, ed., 1981).

Conclusion and Recommendations

Assays for chromosomal aberrations are available which have been fairly extensively applied both to in vitro systems and in animal and human tissues in

vivo. There is quite wide agreement about the practical utility of using aberrations as indicators of carcinogenic potential in screening of chemicals (IARC, 1980), and about their potential usefulness as biologic dosimeters in animals and humans (Bloom, ed., 1981), and for detecting genetic response in human populations (NAS, 1983).

The use of lymphocytes is recommended for parallel testing in *in vitro* systems, laboratory animals, and humans since they are a readily available tissue. Furthermore, a significant proportion of lymphocytes is long-lived; therefore the cytogenetic effects of a dose accumulated from chronic exposures is likely to be measurable. In particular, NTP should retain the CHO *in vitro* assay in its genetic toxicity pre-chronic battery and should include chromosomal aberrations among the short term tests applied in the proposed *in vivo* animal study of model compounds.

Specific Recommendations:

- Conduct *in vitro* experiments prior to *in vivo* testing to determine S-stage specificity of the test substance in order to avoid the possibility of false negative results (Preston *et al.*, 1981).
- Analyze cells in their first post-treatment mitotic division over a complete cell cycle to mitigate cell cycle dependency.
- Select an adequate sample size; randomize both control and study cultures; and conduct a blind analysis of slides.
- For human studies, carefully select control and exposed groups and account for confounding variables such as drugs, radiation exposure, and cigarette smoking, etc.
- Conduct experiments over a range of dose concentrations to obtain dose-response data.

Human Somatic Cell Mutation Assays

Background

There are a variety of test systems which measure mutation frequencies at a specific locus of a mammalian genome. Several of these use the HGPRT locus and are based on the positive selection provided to HGPRT mutants, which can grow in

the presence of several purine analogs that are toxic to normal cells. Such tests have been developed using human or rodent cells from a number of tissues.

Blood is the tissue most accessible for monitoring genetic damage in humans, and considerable progress has been made in the development of a specific locus (HGPRT) mutagenicity test that uses the T-lymphocytes from peripheral blood of rodents or humans as a test tissue (Albertini, 1982; Strauss, 1982). Three assays that are under development are discussed below.

1. Specific Locus Mutation Assay in T-Lymphocytes

Mutations at the HGPRT locus in peripheral blood lymphocytes confer resistance to the cytotoxic effects of the purine analogs, 8 azaguanine (AG) and thioguanine (TG), and affords a positive selection method for the HGPRT$^-$ mutant cells. This system uses the frequency of HGPRT$^-$ peripheral T-lymphocytes as an indication of the frequency of somatic cell mutation that is occurring *in vivo*. Thus, it is proposed as an *in vivo* monitoring assay for mutations at a specific locus in humans. (For reviews, see Albertini, 1980, and 1982; Strauss, 1982; and Thilly *et al.*, 1982.)

In the original method, mutant cells that were resistant to the cytotoxic analogs were detected as viable cells by the uptake of ^{3}H-thymidine into newly synthesized DNA and subsequent autoradiography. With the recent availability of T-cell and B-cell growth factors, it is now possible to select for these mutant cells by plating them on a medium containing the purine analogs and the growth factors. The mutant cells grow in a clonal manner until they are visible as distinct colonies. This improvement should enable mutant frequencies to be assayed using colony counter techniques and is likely to make the assay much faster and more reproducible (Strauss, 1983).

The success of the early system was based on the assumption that all peripheral blood lymphocytes were in G_0 and were not in S. However, it became apparent that this was not the case, and the so called cycling cells which

synthesized DNA during the period of treatment with AG or TG were a major source of interference. These "phenocopies" would be scored as mutant cells and would make the mutation frequency in the cell population artifactually high. The contribution of cycling cells has been reduced by freezing the samples prior to incubation with ^{3}H-thymidine, which moves the cells out of S phase during the labeling period of the normal assay procedure (Albertini, 1982; Strauss, 1982).

2. Glycophorin Assay

Another scheme to monitor for mutations in erythrocytes of rodents or humans is based on a null mutation assay in which the loss of the membrane surface protein, glycophorin, is measured. Glycophorin A is the dominant glycophorin and determines the M and N blood groups. There are two alleles of this protein present in equal frequency in the population, so that roughly half of all people are heterozygous. Monoclonal antibodies have been developed specifically for the M and N forms which are labeled with different fluorophores (red, green). All cells from heterozygous subjects are treated with both antibodies simultaneously, and all normal heterozygous cells should present both colors in equal amounts. Rare cells will present only one color, indicating the absence of one glycophorin gene product. Because the system requires a cell to be labeled with at least one fluorophore before it is counted as a mutant cell, it is hoped that artifactual quenching or masking of one fluorophore by glycosylation of only one glycophorin is unlikely because the glycosylation patterns for both glycophorins are the same. As with the mutant hemoglobin assay, it is likely that this method could assay frozen, stored blood samples (Branscombe, 1983). This system is still at an early stage of development.

3. Human Mutant Hemoglobin Assay

There appear to be several limitations of this system which measures mutations by assaying for mutant species of hemoglobin in erythrocytes. With this approach, fluorescent antibodies are made to various species of hemoglobin, and

the frequency of these mutant species is measured using flow cytophotometry.

At issue is whether the mutation frequency in hemoglobin will be sufficient to produce a favorable signal/noise ratio and permit the practical detection of mutations by this method. The fluorescence activated cell sorter has had trouble distinguishing between erythrocytes that stain for the mutant hemoglobin and fluorescent particles that have been stained non-specifically. Fluorescent particle artifacts are present in the sample preparation at the same frequency (one cell in 10^5 - 10^6 cells) as would be expected for mutant hemoglobin-containing erythrocytes. Thus, the flow system is very sensitive, but it has poor ability to discriminate between stained cells and other particles of similar scatter and fluorescence. More or less rare cells apparently stain with antibody for reasons that cannot be attributed to their carrying a mutant hemoglobin. Attempts to prevent artifactual staining using standard techniques have been unsuccessful (Branscombe, 1983).

Advantages and Disadvantages

These assays are not sufficiently developed with standard protocols to permit an assessment of their advantages or disadvantages. The theoretical advantages are:

1. They are adaptable to both human and rodent tissues, *in vivo* and *in vitro*, and would permit *in vivo*/*in vitro* as well as inter-species comparison of responses.
2. The ability to clone mutant cells gives potentially greater power to the assay.

The disadvantages of assays such as the HGPRT mutant T-lymphocyte assays are:

1. The systems are still under development.
2. The use of heterogeneous erythrocyte cell population makes the interpretation of the results complex.

Recommendations

Because these methods can, in theory, assay for mutations *in vivo* in rodents

or humans using accessible tissues (erythrocytes or lymphocytes), they are very attractive and their development should be encouraged. However, such tests are insufficiently developed at present to be included into either the NTP standard battery of short term tests or into the in vivo testing scheme proposed here. They are in need of further validation and should be considered as promising candidates for inclusion into either test battery in the future.

Sperm Abnormalities

Background

Assays of semen quality (sperm count, motility, or morphology) have long been used in the diagnosis of fertility in humans and domesticated animals. Recent developments indicate that the shape of sperm is genetically determined, and that abnormal forms of sperm may arise from mutational events in the germ plasm (Wyrobek, 1983). Thus, methods that measure changes in sperm morphology in humans or the mouse are assays for genetic damage to the germ plasm, in addition to indicating spermato-toxicity. Deleterious mutations to the germ cell could produce reproductive effects (e.g., decreased fertility or abnormal birth outcome). Agents that cause mutation in germ cells may also induce mutations in somatic cells. The sperm morphology test, therefore, has potential as both an in vivo assay in humans for exposures to mutagens, and an in vivo monitor of mutagenic, reproductive and potential carcinogenic effects. The methods used to assay for sperm morphology in mice and in humans have recently been reviewed by the U.S. E.P.A. Gene-Tox Program (Wyrobek et al., 1983a and 1983b) and will not be described in detail here.

Potential Role of Sperm Morphology Tests in Assessing Carcinogenicity Animals

The mouse sperm morphology test has a low sensitivity (55%) and very high specificity (100%) when assaying 69 carcinogens and 24 non-carcinogens respectively (Wyrobek, 1983a). Thus, the test fails to detect a significant proportion

of the tested carcinogens. With a limited data base of 24 non-carcinogens, the test has a very low "false positive" rate. A positive response in the mouse sperm morphology test, therefore, may be helpful in assessing carcinogenic potential. However, no conclusion as to carcinogenicity can be drawn from a negative result, even when the agent is tested to a high dose level.

Humans

Studies indicate that cancer patients have a higher incidence of abnormally-shaped sperm than do control patients. Eight pre-treatment cancer patients demonstrated an unexpectedly high level of sperm abnormalities (60.6 $\pm$ 5.2% vs. 42.4 $\pm$ 1.8% for the controls; $p < 0.03$). This increase may be due to the illness or confounding factors, medication or diagnostic procedures (Wyrobek, 1980). Other studies indicate that smokers have perhaps 10% more abnormal sperm than do non-smokers (Evans, 1983). Thus, there are limited data for a direct association between abnormal sperm morphology and the risk of cancer in humans.

Role of Sperm Morphology in Assessing Germ Cell Damage

Several recent scientific policy documents have stressed the need to collect more data on the effects of chemicals on the germ cells (Bridges, 1982; Bloom, 1982; Adler, 1983; Streisinger, 1983; Ashby, 1983). The "latency" period before germ cell mutations are expressed as reproductive effects (e.g., fetal loss) may be quite short (1/10th - 1/30th) when compared to the time required for the expression of a somatic cell mutation as cancer. In addition, a far smaller cohort is required to detect significant effects for some of the indicators of germ cell damage (e.g., sperm morphology) than is true for a cancer epidemiologic study. Thus, it is likely that such indicators of germ cell mutation or reproductive outcome could detect the introduction of a mutagenic (or carcinogenic) drug in a human population much earlier than if one waited for cancer to appear.

Therefore, use of tests such as the sperm morphology assay is attractive in spite of their relatively low sensitivity to detect carcinogens. For those chemicals which are known to be carcinogenic, there is an equally valid need to ask if they are likely to have an effect on the germ plasm in humans. It is already known that abnormally-shaped sperm may reflect mutational events, and that the induction of such sperm shape changes is highly correlated with germ-cell mutation as indicated by the positive results of testing chemicals in either the specific locus, dominant lethal, or heritable translocation tests (6/6, 15/15 and 7/7 respectively) (Wyrobek et al., 1983a). There is also evidence that sperm morphology is a more sensitive assay for potential reproductive toxins than either the dominant lethal or heritable translocation tests. For example, there are eight chemicals (acriflavin, aminopterin, 5-bromo-deoxyuridine, ethionine, griseofulvin, hydroxyurea, 5-iodo-deoxyuridine, and vinblastine) that are at present positive in the sperm morphology assay but negative in the dominant lethal test (Wyrobek, 1983a). Thus, the sperm morphology assay in humans or rodents should be considered highly sensitive to germ-cell mutagens.

Human/Animal Extrapolation

There are very limited data on the comparative responses of different species. Such data as are available indicate that different species in general have shown similar responses. Eleven of 16 agents for which sperm morphology was assessed in 2 or more species showed complete agreement in results. Further work is needed on comparisons between mouse and man. Although 154 agents have been studied in the mouse and 89 in humans, it is worth noting that only 6 agents have been studied in both species (Wyrobek et al., 1983a and 1983b).

Recent Developments

Three recent developments make the sperm morphology assay an attractive in vivo system for monitoring genetic and potentially carcinogenic effects in rodents or humans.

1. Person-specific patterns of sperm abnormalities are found to be quite stable, with only slight variations recorded for individuals over several years time. This enables the sample sizes required to detect a significant effect in a longitudinal study to be very small. Consequently, in a longitudinal study where every person serves as his own control, the number of persons in a study population required to demonstrate significant effects on sperm morphology is very small (<10). Mice would be expected to show a similar stability in morphology pattern (Wyrobek, 1982, 1983b, 1984a,b). Furthermore, mice have the advantage that they are inbred.

2. In the mouse there is evidence to indicate that abnormal sperm can arise from mutational events. Treatment of male mice with triethylenemelamine (TEM), an inducer of chromosome abnormalities and abnormal sperm morphology, produces abnormal forms in the male, which are transmitted after mating and persist in the male progeny. This heritable dominant effect has occurred in 100% of the small number of males that have been treated thus far (7/7) (Wyrobek, 1983c).

3. Hamster Egg System. This system fuses capacitated human sperm with denuded hamster eggs. After fusion, the cytoplasm of the egg decondenses the chromatin of the sperm. Some hours later the haploid set of sperm chromosomes is visible in the egg, and a standard karyotype analysis can be performed. By this method, in principle, one can perform cytogenetic analysis of human or rodent sperm (Brandriff et al., 1982, 1983, 1984).

Advantages and Disadvantages

The advantages and disadvantages of the sperm morphology assay have been recently summarized by the U.S. E.P.A. Gene-Tox Program (Wyrobek et al.). The following is a brief summary of their evaluation. The major advantages of the sperm morphology test are:

1. Germ cells are exposed in vivo, and a positive result demonstrates the ability of an agent to affect spermatogenesis.
2. Induction of abnormally-shaped sperm appears to be sensitive to mammalian germ-cell mutagens.
3. Similar assays can be performed on humans and animals and can be followed for reproductive as well as carcinogenic outcomes.

The major disadvantages of the test include:

1. Changes in sperm morphology can, in theory, be caused by agents other than mutagens (i.e., the linkage between abnormal forms and mutation is not exact), although all tested germ cell mutagens are positive in the assay, and all non-carcinogens are negative.
2. The sensitivity to carcinogens is low: an appreciable proportion of carcinogens are negative in the test.
3. It is sometimes difficult to obtain human samples due to social sensitivity.

Recommendations

The mouse sperm morphology test has a low sensitivity to detect carcinogens. In addition, positive responses may arise from spermatotoxic as well as from mutagenic events. The recommendations for the sperm morphology assay are difficult to make because of the mixture of advantages and disadvantages of the test. However, sperm is one of the few candidate tissues which is readily available from humans and rodents and is suitable for in vivo monitoring of genetic damage, and use of this tissue should be encouraged. The mouse sperm morphology test has a reasonable well-developed historical data base of responses to carcinogens and non-carcinogens, and the assay is sensitive to a subset of mutagens that are active in the male germ cells, perhaps more so than existing germ-cell ;mutational assays.

The mouse sperm morphology assay, at present, is in need of further evaluation as a predictor of carcinogenicity. We recommend that, if used, it be employed in a highly selective manner in the NTP prechronic testing, specifically for those chemicals whose chemical structures indicate that they may be potent carcinogens or be active on male germ cells. We recommend that the assay be considered for inclusion into the *in vivo* test scheme that is proposed in this report.

Short Term Tests for Transformation and Promotion*

These tests are relevant to a review of the NTP program although they are not applicable to *in vivo* monitoring -- hence parallel testing. Their relevance stems from the fact that, in theory, they measure endpoints other than the strictly mutagenic/initiating events in carcinogenesis and hence extend the range of any short term test battery. In addition, the NTP genetic toxicity "battery" now includes cell transformation assays, so that a brief assessment of their usefulness is appropriate. On the other hand, the absence of tests for promotion in the NTP program represents a major gap in this and all other short term testing programs; it is therefore an area for future emphasis by NTP's test development and validation program.

Cell Transformation

Background

A variety of cell culture systems have been developed with applications for the study and detection of chemical carcinogens (for recent reviews see Heidelberger *et al.*, 1983; Meyer, 1983). The majority of these cell

---

* Promotion is the process by which an agent, applied repeatedly after a single subcarcinogenic dose of an initiator, results in tumors. By contrast, cocarcinogens can be defined as the production of a greater than additive tumorigenic effect when two agents are given simultaneously. Cocarcinogenesis can occur with a carcinogen and a noncarcinogen (e.g., BaP and pyrene) as well as two carcinogens (e.g., DMBA and BaP). Although in theory a cocarcinogenic effect could result when two promoting agents are given at the same time, this phenomenon has been little studied. The term largely refers, therefore, to the initiating stage of the carcinogenic process.

culture systems utilize fibroblast cultures derived from embryonic tissues. The systems in common use employ either permanent cell lines (e.g., Balb/c 3T3 or C3H/10T1/2 cells) or Syrian hamster embryo cells with limited proliferative potential. Single treatments with carcinogens are frequently sufficient to produce transformed cells which are easily recognized on the basis of distinct changes in morphology and growth control. The relative ease with which transformants can be generated and recognized is an advantage. However, there has been some debate over the question of quantification of carcinogen induced effects (Kennedy, 1984).

The cellular mechanisms responsible for the morphological transformation of cells in culture are not understood. Since such systems appear to assay the ability of chemicals to convert nontumorigenic cells to a tumorigenic state, or normal cells to a preneoplastic state, it is generally assumed that these systems mimic central aspects of the carcinogenic process. For this reason, cell transformation systems may provide a biologically relevant means of detecting chemical carcinogens and may be sensitive to substances which pose a carcinogenic risk by mechanisms (e.g., promotion or cocarcinogenesis) which will not elicit a response in other assays for genotoxicity.

Several other systems are in use today which assess the effects of chemical treatment upon transformation in cells infected with oncogenic viruses. Of these, the two most commonly used systems assess tranformation in leukemia virus infected rat embryo cells and the ability of chemicals to enhance transformation by adenovirus SA7 in Syrian hamster embryo cells. The mechanistic relationship between transformation in these systems and those previously described is uncertain.

System Advantages and Disadvantages

The advantages and disadvantages of commonly used cell transformation systems recently have been reviewed by the U.S. EPA Gene-Tox program (Heidelberger et al., 1983). The following capsule summaries are based, in

large part, upon this extensive evaluation.

a. Systems Using Cell Lines (Balb/c 3T3 and C3H/10T1/2)

1) Advantages: These cell transformation systems are generally reproducible, may be easy to quantitate, in use in a large number of laboratories and provide satisfactory dose-dependent responses to a number of direct-acting carcinogens and procarcinogens. Levels of spontaneous transformation are usually low and are lowest in the C3H/10T1/2 system. Both systems also show responsiveness to tumor promoters in a fashion analogous to multi-stage carcinogenesis in vivo. This phenomenon has been most extensively studied in C3H/10T1/2 cells.

2) Disadvantages: Both cell lines are aneuploid and considered by many to be preneoplastic. The actual mechanistic significance of positive responses is thus unclear. Numerous classes of procarcinogens cannot be metabolized by either cell type, resulting in a large number of false negative test results. Exogenous metabolic activation systems for routine use with either system are not yet available. The C3H/10T1/2 system has also demonstrated relative insensitivity to a number of direct-acting agents. Recent assay modifications (Frazelle et al., 1983; Nesnow et al., 1982; Oshiro et al., 1981) may permit this short-coming to be overcome, however, and even increase sensitivity of C3H/10T1/2 cells to procarcinogens.

b. Syrian Hamster Embryo Cells

1) Advantages: This system permits unusually rapid identification of numerous carcinogens. The cells are capable of metabolizing diverse classes of procarcinogens, and exogenous metabolic activation systems have been successfully employed. Furthermore, this

system uses normal diploid cells and has extremely low levels of spontaneous transformation.

2) Disadvantages: The Syrian hamster embryo assay is exceedingly complex and difficult to conduct. Scoring of the assay is more subjective than in other systems, and dose-response relationships are sometimes difficult to produce. Although a large number of chemicals have been evaluated in this system with excellent results, interlaboratory verification of both system specificity and sensitivity is lacking.

c. SHE/SA7 Transformation

1) Advantages: This system yields reproducible results for a wide variety of carcinogens. The cells used are diploid and capable of metabolizing a diverse range of procarcinogens.

2) Disadvantages: The assay measures the chemical-mediated enhancement of transformation by adenovirus SA7. Whether the mechanism of this effect is necessarily indicative of transforming potential, or is instead an indication of DNA damage or effects upon other factors such as viral infectivity, remains to be established.

d. Leukemia-Virus Infected Cells

1) Advantages: Studies conducted in accordance with the original protocols of Freeman et al., (1973) have found this system to produce satisfactory results with a large number of compounds. However, the utility of this system has been limited by its labor-intensive nature. Serial passaging of cultures has been required for the conduct of the assay, prohibiting quantitative evaluation of results on a per cell basis. Modification of the original assay has been proposed (Traul et al., 1981; Suk et al., submitted)

and may permit execution of experiments within a relatively short time frame (≤ 2 weeks). This modification entails examination of the ability of chemicals to enhance the survival of cells placed into an anchorage independent state and should permit the generation of quantitative results.

2) Disadvantages: The original assay protocols are not suited for quantitative studies and are thus of limited use for carcinogen detection. Modifications to the assay may overcome this deficiency, but are still somewhat labor intensive and remain to be validated. Moreover, the precise relationship between the assay endpoint (survival in an anchorage independent state) and oncogenic transformation remains to be determined.

Short Term Recommendations

a. It is evident from NTP validation studies (and other studies in the literature) that an inability to metabolize procarcinogens is a serious and limiting deficiency of the Balb/c 3T3 system (and probably cell transformation using C3H/10T1/2 cells as well). Use of the system has identified few carcinogens which go undetected by other more rapid and inexpensive assays for genotoxicity. Routine screening or further extensive validation with this system, except as noted below, would not appear to be a wise allocation of resources until such time as an exogenous system for metabolic activation is available for routine use. Funding of research by NTP to develop such an activation system should be given a high priority. Validation studies (Sivak and Tu, 1982) have demonstrated the Balb/c system to be an accurate predictor of carcinogenicity for direct-acting carcinogens,

polycyclic aromatic hydrocarbons, metals and some aromatic amines. Use of the 3T3 system for screening chemicals which belong to these classes may be indicated.

b. The Balb/c 3T3 system is, of course, potentially sensitive to carcinogens which have weak or neglible activity in other assays for genetic damage. Several such "problem" carcinogens have been identified by other systems (i.e., 5-azacytidine, diethylstilbesterol, asbestos, arsenic and benzene). If such chemicals are not included in ongoing validation studies, they should be. Comparison of the ability of the different cell transformation systems to respond to such chemicals is in order and may succeed in identifying systems with the greatest sensitivity to such substances.

c. Future validation studies in all cell transformation systems should include more "negatives" and also suspected human carcinogens.

d. Further validation of the SHE/SA7 system is in order, with emphasis on the following considerations:

   1) How do results from this system compare with results obtained from assays for DNA damage.

   2) How responsive is the system to "nonmutagenic" carcinogens?

   3) What is the specificity of the assay? (The assay responded to all three "negatives" in NTP validation studies: R. Tennant; personal communication.)

e. The C3H/10T½ system appears to have limitations similar to Balb/c with respect to metabolism. Unless these can be overcome, use of this system should probably be limited except as noted in points 2 and 3.

f. The SHE cell transformation system has amassed an impressive literature performance. NTP should determine whether or not the extreme technical difficulties associated with this system can be overcome and

the system employed in a reproducible fashion by multiple laboratories.

g. The modified RLV/RE cell transformation system currently being evaluated by NTP shows promise and deserves additional validation and/or development. However, the biological significance of the modified assay's endpoint for the carcinogenic process is unclear. Efforts should be made to determine the relationship of this assay to the process of oncogenic transformation and the role of the retrovirus in generating chemically-induced responses. These latter efforts should be assigned higher priority than routine screening or extensive validation.

Long-Term Recommendations

a. Tests should be conducted under optimal conditions. Failure to do so limits their utility, may lead to marginal low-level responses of uncertain significance, and may limit reproducibility between laboratories.

b. Resources devoted to the use of cell transformation systems as screening tools have far exceeded those allocated to understanding the cellular mechanism(s) of transformation. As a result the mechanism of transformation is unknown and may vary from system to system. It is thus impossible to state which system provides the most relevant indicator of carcinogenic potential. Clearly, much work is needed to more carefully standardize and optimize the conditions under which the assay is conducted. Many of the questions which need to be addressed can only be answered through basic research on mechanisms.

c. Cell transformation systems can be used to study and evaluate the ability of chemicals to influence multiple aspects of the complex process of carcinogenesis (i.e., cocarcinogenesis, promotion,

progression). The principal strength of these systems may be in this potential, which unfortunately, has not been exploited. Better use of cell transformation systems to explore these aspects of carcinogenesis should be encouraged (IARC, 1983).

d. Finally, NTP should recognize that the more commonly used cell transformation systems do not utilize cells with tissue or species origins of greatest concern to chemical carcinogenesis (i.e., cells of epithelial and/or human origin). It is possible that transformation systems using human epithelial cells will be available in 10 years.

Promotion

The recognized multi-stage theory of cancer includes promotion; yet little is known about the actual mechanisms involved. (See Bohrman, 1983; Yamasaki and Weinstein, 1983; IARC, 1983; Trosko et al., 1983; Troll, 1983; Boutwell, 1974; and Yuspa et al., 1983 for review.) A factor greatly complicating the task of identifying critical mechanisms and then developing definitive tests to detect and measure promoters is their observed wide range of effects. This is well illustrated by the phorbol esters and specifically by TPA, probably the most studied of all the promoters. TPA induces a host of cellular effects including: cell membrane changes, induction of plasminogen activator, changes in growth properties, effects on membrane receptors, inhibition of cell-to-cell communication, modulation of cell differentiation, gene amplification, strand breaks and SCE. (See Yamasaki and Weinstein, 1983; Trosko et al., 1983; Dorman et al., 1983; IARC, 1980 and 1983 for review.) Another complication is the possibility that the same chemical may act primarily as a promoter in one tissue and as an initiating agent in another (e.g., 2-AAF, asbestos). For example, as mentioned in the introduction of the Report of the Subpanel on the Design of Chronic Studies, 2-AAF appears to have initiating activity in mouse liver and promoting activity in mouse bladder; while epidemiological data on asbestos suggest that it is an initiator for induction of mesothelioma and a promoter in

the development of lung cancer. (See IARC, 1983.) The result is that short term tests for promotion are now insufficiently developed for inclusion in a short term test battery (IARC, 1983). More studies are needed on the mechanisms of tumor promotion so that reliable short term tests can be developed.

This is a major gap in short term testing because of the estimated wide extent of human exposure to promotors or late-stage carcinogens (e.g., cigarette smoke, asbestos) and the observations that certain promoters (TCDD, TPA) have induced tumors by themselves at low doses. There is a lack of evidence that epigenetic agents such as promotors necessarily pose a lesser cancer risk to humans than do genotoxic agents (Weinstein, 1983).

Development and validation of tests for promotion should therefore be a priority with NTP. Potential candidates, listed in Table 7, are reviewed by Bohrman (1983), Yamaski and Weinstein (1983) and Sivak (1982). (See also the preceding section on cell transformation.) So far, cell transformation tests have been largely used to study mechanisms of transformation by initiators and promoters rather than as screens for promoters.

TABLE 7
EXAMPLES OF PROMISING IN VITRO AND IN VIVO MODELS FOR SCREENING TUMOR PROMOTORS

In vitro:	o cell transformation models (e.g., C_3H, 10T1/2, T3; hamster embryo cells; mouse epidermal cells; human foreskin cells) o cell adhesion o metabolic cooperation (cell-to-cell communication)
In vivo:	o 2-stage animal models (e.g., mouse skin; mouse lung; Fischer F 344 rat liver)
Source:	Bohrman (1983); Yamasaki and Weinstein (1983); Sivak (1982)

The cell-to-cell communication system is the subject of an NTP research contract at this time. The results should shed some light on the present uncertainties and controversy surrounding the assay. (See Dorman et al., 1983; Trosko et al., 1983.)

Induction of ornithine decarboxylase (ODC) is an early event in all growth systems studied to date. Although elevated ODC activity has also been implicated in certain types of carcinogenesis (Raunio et al., 1983), more studies are needed to determine if it is a useful marker of exposure to tumor promotors.

Promising in vivo systems for promotion are limited rodent bioassays in which small initiating doses of a genotoxic carcinogen active at a specific target organ (skin, liver, lung, pancreas, breast, colon, urinary bladder) are administered prior to the suspected promotor. A number of promotors are active as such only at one site. However, there is evidence that an agent can act as a promotor at one site and an initiator at another (IARC, 1983). This means that it may be desirable to test potential promotors in vivo in several organs (in addition to several animal species and strains) to account for the possibility of organ-specific (species and strain-specific) effects.

E. General Summary Regarding Short Term Tests

1. The present short term test battery under evaluation by NTP is a reasonable one. (See p. 30-33 for discussion.) In order to ensure that this test battery generates the most comprehesive and relevant possible data base, we recommend the following measures:
   a. In general, test 2-year bioassay compounds in the full battery in order to develop comprehensive genetic toxicity data. This would establish a valuable data base for correlation and cross-validation between the various types of assays. (See p. 30.)
   b. Continually reassess the usefulness of individual tests now included in the battery -- giving special consideration to the goal of measuring the widest possible array of endpoints.
   c. Give high priority to the development and validation of in vitro and in vivo tests that identify tumor promotors. The aim should be to add these tests to the standard battery of genetic toxicity assays as soon as possible.

d. Periodically evaluate new short term test methodologies for possible inclusion in either the standard test battery or in parallel in vivo testing studies described in 2 and 3. These include not only tests to detect agents capable of causing initiation, promotion, and cell transformation but also agents capable of interaction with oncogenes and related control genes.

2. In order to validate selected in vivo short term tests as indicators of internal activated dose, biologically effective dose or genetic response to achieve the goals described in the Introduction, NTP should initiate laboratory (rodent) studies on the effects of a few selected model compounds as follows:

   a. The model compounds would be substances that have been tested in the NCI/NTP bioassay program and to which identified human populations are currently exposed through medical treatment, occupation, or environmental pollution (e.g., ethylene oxide, or a cancer chemotherapeutic agent, etc.). The model compounds could include:

      1) chemicals clearly carcinogenic in both the mouse and the rat in studies which have shown a good dose-response relationship. (The Subpanel recognizes, however, that the dose-response pattern can vary significantly between sex and species for the same chemical.) One chemical might be a site-specific carcinogen; another might be active at multiple sites;

      2) chemicals clearly without carcinogenic effect in repeated bioassays; and

      3) chemicals with marginally positive bioassay data.

   b. NTP could conduct the in vivo short term studies on the mouse and/or the rat. The advantage of using the rat is the relatively large amount of blood available for study.

   c. The candidate short term tests listed below could provide in vivo evidence of internal dose, biologically effective dose, and/or genetic response to the chemicals tested. The Subpanel has grouped

available tests in two categories: those that, in its view, are now immediately applicable to parallel studies (A) and those that are less immediately applicable (B). This general ranking is based on present assessment of the state-of-the-art, feasibility, state of validation, and relevance to carcinogenesis of the endpoint measured. However, given the rapid progress in this field, even this very broad classification will no doubt need revision in the near future. (Of course, as mentioned in II, "Background," there are chemical - and chemical class - specific considerations that would determine the selection of a battery of tests for individual test chemicals. For example, covalent adduct assays are only available for a limited number of compounds.)

Group A

SCE

Chromosomal aberrations

Protein-, DNA-, RNA- carcinogen adducts

Detection of mutagens in body fluids

Group B

Micronucleus

Sperm morphology

DNA damage and repair

Somatic cell mutation assays

3. NTP could undertake cooperative studies with NIEHS and NIOSH using the same short term tests in human populations exposed to the selected model compounds in order to compare and calibrate human responses with rodent data. In these studies, rigorous control of potentially confounding variables will be necessary. For example, in studies of mutagenicity of urine one must take into account concurrent smoking, dietary habits, and other exposures to mutagenic agents.

4. In designing parallel short term in vivo experiments described in 2 and 3, NTP should devise an optimal sampling schedule which takes into account the formation and persistence of the lesion(s) or short term effect(s) studied.

5. In a related effort, NTP might consider the possibility of storing certain biological fluids or samples (e.g., blood, urine, or sperm) from a subset of animals undergoing testing with a limited number of additional selected chemicals. This approach could allow retrospective analysis of samples using selected short term tests (excluding only cytogenetic and mutagenic assays which cannot be performed on frozen samples e.g.; all but cytogenetic and some mutagenesis assays) that were used in the parallel laboratory animal/human studies described in 2 and 3. In addition, NTP could retrieve specific information pertinent to dose-response observed in the carcinogenicity bioassay, such as the relationship between the administered dose and the biologically effective dose of the agent. Prior to undertaking such studies NTP should consider:

   a. Whether these samples should be stored from animals undergoing prechronic, carcinogenicity or other testing;

   b. Whether serial sampling (repeated sampling in the same animal or group) would be advantageous;

   c. What are the various cost benefit considerations in determining the type and extent of sampling and storage;

   d. What are the practical problems with storing samples for future analysis? Will freezing and thawing affect the particular lesion to be measured? Should NTP conduct a preliminary test with a number of chemicals and types of samples to answer these questions? and

e. The specific questions to be asked of the samples.

6. NTP might also consider directly incorporating one or more short term test(s) into an ongoing bioassay in order to obtain prospective data on the relationship between the various indices of dose and genetic response (including "dose-response", persistence, and individual variation) and tumorigenesis. For this purpose, NTP could perform the assays on a subset of the bioassay test animals themselves or on an additional (parallel) group of animals -- depending on the impact of sample collection on the viability of the test animals (morbidity and mortality). This prospective approach would allow NTP to correlate levels of short term effects with tumor incidence in the same animals or in animals housed and dosed under identical conditions. It would be particularly desirable to know if those animals which developed tumors had experienced increased levels of short term effects.

7. NTP should establish a goal of better understanding mechanisms by developing a battery of short term tests that measures the widest possible number of endpoints (including promotion, transformation and chemical interaction with oncogenes) developing a comprehensive data base through testing, and then using the results to study and elucidate possible mechanisms. This would help NTP to understand discrepancies in their data base for specific chemicals (e.g., between those with negative short term test data and positive cancer bioassay results).

8. NTP should consider both positive and negative short term test data in evaluating animal bioassay data for a particular chemical. For purposes of interpreting animal data, negative short term test data do not, by themselves, provide a basis for discounting positive bioassay results. This is because of factors such as the limitations of the battery to detect all possible mechanisms involved in carcinogenesis, and the

uncertainty as to which mechanism is operative in the carcinogenic process for a particular chemical. We recommend that NTP develop criteria for assessing the adequacy of negative data from short term tests.

9. NTP's preliminary analysis of genetic toxicity and animal data indicate that there may be a correlation between carcinogenicity in multiple species and both sexes and multiple positive results in the short term battery (Tennant et al., 1983). NTP should pursue this line of research.

10. With respect to the present NTP short term standard battery, it is recommended that NTP:

    a) continue to evaluate the rat hepatocyte UDS test using metabolically competent cells; test chemicals known to require metabolism by intestinal flora in the in vivo/in vitro unscheduled DNA synthesis (UDS) assay; test in the in vivo/in vitro assay all chemicals that induce hepatocellular tumors in the bioassay but are negative in the in vitro UDS test: the SCE in vitro assay using CHO (Chinese Hamster Ovary) cells; and the CHO in vitro assay for chromosomal aberrations.

    b) remedy the deficiency of the Balb/c 3T3 cell transformation test system (inability to metabolize procarcinogens) by developing an exogenous activation system; in the meantime, consider using the 3T3 system only for screening direct-acting carcinogens;

    c) continue to assess the micronucleus test as a secondary or tertiary screen as appropriate following the NTP gene-tox battery; evaluate chemical leukemogens with this technique; support the development of automated methods; and

    d) use the mouse sperm morphology assay in a selective manner in the NTP prechronic testing, specifically for those chemicals whose chemical structures indicate that they may be potent carcinogens or be active on male germ cells.

11. With respect to possible future additions to the NTP short term test program, it is recommended that:

a) NTP explore the possibility of developing short term assays to detect carcinogen-oncogene interraction.

b) somatic cell mutation tests (such as the specific locus mutation assay in T-lymphocytes and the glycophorin assay) be further developed and validated. At some future time these tests could be considered for inclusion in either the NTP standard battery of short term tests or in the proposed *in vivo* testing scheme.

c) The mouse sperm morphology assay may be considered for future inclusion into the proposed *in vivo* test scheme. Further validation and evaluation of this assay as a predictor of carcinogenicity is needed.

REFERENCES

Adler, I.D., "Comparison of Types of Chemically Induced Genetic Changes in Mammals," Mutation Res., 115:292-321, ICPEMC Working Paper 4, 1983.

Albertini, R.J., "Drug-Resistant Lymphocytes in Man as Indicators of Somatic Cell Mutation," Teratogenesis, Carcinogenesis, and Mutagenesis, 1:25-48, 1980.

Albertini, R.J., "Studies with T-Lymphocytes: An Approach to Human Mutagenicity Monitoring," Banbury Report No. 13, Indicators of Genotoxic Exposure, Eds. B.A. Bridges, B.E. Butterworth, and I.B. Weinstein, Cold Spring Harbor Laboratory, pp. 393-412, 1982.

Armuth, V. and Berenblum, I., "Systemic Promoting Action of Phorbol in Liver and Lung Carcinogenesis in AKR mice," Cancer Res. 32:2259-2262, 1972.

Ashby, J., "The Unique Role of Rodents in the Detection of Possible Human Carcinogens and Mutagens," Mutation Res., 115:177-213, ICPEMC Working Paper 1, 1983.

Ashby, J., de Serres, F.J., Draper, M., Ishidate, M., Jr., Mutter, B.E., Shelby, M.D. (eds), "Evaluation of Short-Term Tests for Carcinogens: Report of the International Program on Chemical Safety Collaborative Study on In Vitro Assays," Elsevier, North Holland, Amsterdam, 1984.

Banbury Report, "Indicators of Genotoxic Exposure", Eds. B.A. Bridges, B.E. Butterworth, I.B. Weinstein, Report No. 13, Cold Spring Harbor Laboratory, 1982.

Banscombe, W., personal communication, 1983.

Bennett, R.A., Essigmann, J.M., and Wogan, G.N., "Excretion of an Aflatoxin-guanine Adduct in the Urine of Aflatoxin B_1-Treated Rats," Cancer Res., 41:650-654, 1981.

Berenblum, I., "A Speculative Review: The Probable Nature of Promoting Action and its Significance in the Understanding of the Mechanism of Carcinogenesis," Cancer Res., 14:471, 1954.

Bloom, A.D., "Guidelines for Studies of Human Populations Exposed to Mutagenic and Reproductive Hazards," March of Dimes Birth Defects Foundation, 1981.

Bohrman, J.S., "Identification and Assessment of Tumor-Promoting and Co-carcinogenic Agents: State of the Art in vitro Methods," CRC Crit. Rev. in Toxicol., 11:121-167, 1983.

Boutwell, R.K., "The Function and Mechanism of Promoters of Carcinogenesis," Crit. Rev. in Toxicol., 2:419-443, 1974.

Brandriff, B., Gordon, L., Ashworth, L., Watchmaker, G., Carrano, A., and Wyrobek, A.J., "Chromosomal Aberrations in Human Sperm; Comparisons Among Four Healthy Men," American Journal of Human Genetics, 66:193-201, 1984.

Brandriff, B., Gordon, L., Ashworth, L., Watchmaker, G., Summers, L., Wyrobek, A.J., Carrano, A., "Individuals Differ Significantly in the Incidence of Sperm Chromosomal Aberrations", American Journal of Human Genetics, 35:126a, 1983.

Brandriff, B., Gordon, L., Watchmaker, G., Summers, L., Wyrobek, A.J. "Gamate Interaction and Developmental Progression to the Two-cell Stage in the Human-Sperm/Hamster-Egg Interspecific Cross," J. Cell. Biol. 95:154a, 1982.

Bridges, B.A., "Indicators of Genotoxic Exposure: Status and Prospects," Banbury Report, No. 13, pp. 555-558, 1982.

Butterworth, B.E., Doolittle, D.J., Working, P.K., Strom, S.C., Jirtle, R.L. and Michalopoulos, G., "Chemically-Induced DNA Repair in Rodent and Human Cells," Banbury Report No. 13, pp. 101-114, 1982.

Calleman, C.J., Ehrenberg, L., Jansson, B., Osterman-Golkar, S., Segerback, D., Svensson, K., and Wachtmeister, C.A., "Monitoring and Risk Assessment by Means of Alkyl Groups in Hemoglobin in Persons Occupationally Exposed to Ethylene Oxide," J. Environ. Path. and Toxic., 2:427-442, 1978.

Carrano, A.V., "Sister Chromatid Exchange as an Indicator of Human Exposure," Banbury Report No. 13, Eds. B.A. Bridges, B.E. Butterworth, and I.B. Weinstein, pp. 307-318, 1982.

Carrano, A.V., Minkler, J.L., Stetka, D.A., and Moore, D.H., "Variation in the Baseline Sister Chromatid Exchange Frequency in Human Lymphocytes," Environ. Mutagenesis, 2:235-337, 1980.

Carrano, A.V., and Thompson, L.H., "Sister Chromatid Exchange and Single-Gene Mutation," Sister Chromatid Exchange, Eds. S. Wolff and J. Wiley, pp. 59-66, 1982.

Cheng, M. and Conner, M.K., "Comparison of Sister Chromatid Exchange Induction and Known Carcinogenic Activities of Vinyl and Allyl Carbamates," Cancer Res., 42:2165-2167, 1982.

Cheng, M., Conner, M.K. and Alarie, Y., "Potency of Some Carbamates as Multiple Tissue Sister Chromatid Exchange Inducers and Comparison with Known Carcinogenic Activities," Cancer Res., 41:4489-4492, 1981.

Conner, M.K. and Cheng, M., "Persistence of Ethyl Carbamate-Induced DNA Damage in vivo as Indicated by Sister Chromatid Exchange Analysis," Cancer Res., 43:965-971, 1983.

Croy, R.G., Wogan, G.N., "Quantitative Comparison of Covalent Aflatoxin - DNA Adducts Found in Rat and Mouse Livers and Kidneys," JNCI 66:761-768, 1981.

de Serres, F.J., Ashby, J., "Evaluation of Short-Term Tests for Carcinogens: Report of the International Collaborative Program." Progress in Mutation Research, Vol. I., North Holland, Elsevier, 1981.

Donahue, P.R., Essigmann, J.M., and Wogan, G.N., "Aflatoxin-DNA Adducts: Detection in Urine as a Dosimeter of Exposure," Eds. B.A. Bridges, B.E. Butterworth and I.B. Weinstein, Banbury Report No. 13: Indicators of Genotoxic Exposure, Cold Spring Harbor Laboratory, pp. 221-229, 1982.

Dooley, K.L., Beland, S.A., Bucci, T.J., Kadlubar, S.S., "Local Carcinogenicity Rates of Absorption, Extent and Persistence of Macromolecular Binding, and Acute Histopathology Effects of N-hydroxy-1-naphthylamine and N-hydroxy-2-naphthylamine, Cancer Res , 44:172-177, 1984.

Dorman, B.H., Butterworth, B.E., and Boreiko, C.J., "Role of Intercellular Communication in the Promotion of C3H/10T1/2 Cell Transformation," Carcinogenesis, 4:1109-1115, 1983.

Dunkel, V., "Biological Significance of Endpoints," Ann. N.Y.A.S., 407:34-41, 1983.

Ehrenberg, L., and Osterman-Golkar, S., "Alkylation of Macromolecules for Detecting Mutagenic Agents," Teratogenesis, Carcinogenesis, and Mutagenesis, 1:105-127, 1980.

Erexson, G.L., Wilmer, J.L., Kligerman, A.D., "Analyses of Sister-Chromatid Exchange and Cell-Cycle Kinetics in Mouse T- and B-Lymphocytes from Peripheral Blood Cultures," Mutation Res., 109:271-281, 1983.

Evans, H.J., et al., "Cytogenetic Damage as an Endpoint in Short-Term Assay Systems for Detecting Environmental Carcinogens," In IARC Long-Term and Short-Term Screening Assays for Carcinogens: A Critical Appraisal, Supplement 2, pp. 227-243, 1980.

Evans, H.J., "Cytogenetic Studies on Industrial Populations Exposed to Mutagens," Banbury Report No. 13: Indicators of Genotoxic Exposure, Eds. B.A. Bridges, B.E. Butterworth, and I.B. Weinstein, Cold Spring Harbor Laboratory, 1982a; 325-336, 1982.

Evans, H.J., "Sister Chromatid Exchanges and Disease States in Man," Sister Chromatid Exchange, Eds. S. Wolff, J. Wiley, 1982b; Chap. 1; 183-228, 1982.

Evans, H.J., "Cytogenic Methods for Detecting Effects of Chemical Mutagens," Ann. N.Y.A.S., 407:131-141, 1983.

Evans, H.J., "Sperm Morphology in Cigarette Smokers," Banbury Report No. 13, Indicators of Genotoxic Exposure," Eds. B.A. Bridges, B.E. Butterworth, and I.B. Weinstein, Cold Spring Harbor Laboratory, pp. 543-552, 1982.

Farber, E., "Chemical Carcinogenesis," Am. J. Pathol., 106:271-296, 1982.

Farmer, P.B., Bailey, E., Lamb, J.H., and Connors, T.A., "Approach to the Quantitation of Alkylated Amino Acids in Hemoglobin by Gas Chromatography-Mass Spectrometry," Biomed. Mass Spectrom., 7:41, 1980.

Farmer, P.B., Gorf, S.M., and Bailey, E., "Determination of Hydroxypropylhistidine in Hemoglobin as a Measure of Exposure to Propylene Oxide Using High Resolution Gas Chromatography-Mass Spectrometry," Biomed. Mass Spectrom., 9:69, 1982.

Frazelle, J.H., Abernethy, D.J., and Boreiko, C.J., "Factors Influencing the Promotion of Transformation in Chemically-Initiated C3H/10T1/2 Cl 8 Mouse Embryo Fibroblasts," Carcinogenesis, 4:709-715, 1983.

Freeman, A.E., Weisburger, E.K., Weisburger, J.H., Wolford, R.G., Maryah, J.M. and Huebner, R.J., "Transformation of Cell Cultures as an Indication of the Carcinogenic Potential of Chemicals," J. Natl. Cancer Inst. 51:799-808, 1973

Garattini, S., "Notes on Xenobiotic Metabolism," Ann. N.Y.A.S., 407:1-25, 1983.

Garry, V.F., Hoyler, J., Jacobs, D., Wade, R.L., and Gray, D.G., "Ethylene Oxide: Evidence of Human Chromosomal Affects," Environ. Mutagenesis, 1:375-382, 1979.

Garcia, H.D. and Crouch, G., "Detection of Phenotypically Variant (Thioguanine-Resistant) Rat Spleen Lymphocytes," Banbury Report No. 13, Indicators of Genotoxic Exposure. Eds. B.A. Bridges, B.E. Butterworth, and I.B. Weinstein, Cold Spring Harbor Laboratory, pp. 443-452.

Gilbert, F., "Chromosomes, Genes and Cancer: A Classification of Chromosome Abnormalities in Cancer," J. Nat. Cancer Inst., 71:1107-1114, 1983.

Gupta, R.C., Reddy, M.V., and Randerath, K., "^{32}P-Postlabeling Analysis of Non-Radioactive Aromatic Carcinogen-DNA Adducts," Carcinogenesis, 3:1081-1092, 1982.

Gupta, R.C., Dighe, N.R., "Formation and Removal of DNA Adducts in Rat Liver Treated with N-hydroxy Derivatives of 2-acetylaminofluorene, 4-acetylaminobiphenyl, and 2-acetylaminophenathrene, Carcinogenesis, Vol. 5, No. 3, pp. 343-349, 1984.

Haglund, U., Hayder, S. and Zech, L., "Sister Chromatid Exchanges and Chromosome Aberrations in Children After Treatment for Malignant Lymphoma," Cancer Res.,

Haseltine, W.A., Franklin, W., and Lippke, J.A., "New Methods for Detection of Low Levels of DNA Damage in Human Population," Environ. Health Perspectives, 48:29-41, 1983.

Hayashi, M., Sofuni, T. and Ishidate, M., Jr., "An Application of Acridine Orange Fluorescent Staining to the Micronucleus Test," Mutation Res., 120:241-247, 1983.

Heddle, J.A. and Bruce, W.R., "Comparison of the Micronucleus and Sperm Assay for Mutagenicity with the Carcinogenic Activities of Sixty-one Different Agents," Eds. H.H. Hiatt, J.D. Watson, and J.A. Winsten, Origins of Human Cancer, Vol. 4, Cold Spring Harbor Conferences, 1977.

Heddle, J.A., Hite, M., Kirkhart, B., Mavournin, K., MacGregor, J.T., Newell, G.W. and Salamone, M.F., "The Induction of Micronuclei as a Measure of Genotoxicity: A Report of the U.S. Environmental Protection Agency Gene-Tox Program," Mutation Res., (in press).

Heidelberger, C., Freeman, A.E., Pienta, R.J., Sivak, A., Bertram, J.S., Casto, Bruce, C., Dunkel, V.C., Francis, M.W., Kakunaga, T., Little, J.B. and Schechtman, L.M., "Cell Transformation by Chemical Agents: A Review and Analysis of the Literature," Mutation Res., 114:283-385, 1983.

Hemminki, K., "Dimethylnitroasmine Adducts Excreted in Rat Urine," Chem. Biol. Interact., 39:139-148, 1982.

Hemminki, K., Sorsa, M. and Vainio, H., "Genetic Risks Caused by Occupational Chemicals," Scand. J. Work Environ. and Health, 5:307-327, 1979.

Hemminki, K., and Vainio, H., "Dimethylnitrosamine Adducts Excreted in Rat Urine," Eds. Snyder, Parke, Kocsis, Jallow Gibson, and Whitmer Advances in Experimental Medicine and Biology, pp. 1149-1156, Plenum Press, New York, 1982.

Hiasa Y., Kitahori Y., Ohshima, M., Fujita, T., Yuasa, T., Konishi, N. and Miyashiro, A., "Promoting Effects of Phenobarbital and Barbital on Development of Thyroid Tumors in Rats Treated with n-bis(2-hydroxypropyl)nitrosamine," Carcinogenesis, 3:1187-1190, 1982a.

Hiasa, Y., Ohshima M., Kitahori, Y., Konishi, N., Fujita, T. and Yuasa, T., "B-Cyclo-Dextrin: Promoting Effect on the Development of Renal Tubular Cell Tumors in Rats Treated with n-ethyl-n-hydroxyethylnitrosamine," J. Natl. Cancer Inst., 69:963-967, 1982b.

Hicks, R.M., Chowaniec, J. and Wakefield, J.S.J., "Experimental Induction of Bladder Tumors by a Two-stage System," Eds. Slaga, T.J., Sivak, A. and Boutwell, R.K., Carcinogenesis, Vol. 2, Mechanisms of Tumor Promotion and Cocarcinogenesis, pp. 475-489. Raven Press, New York, 1978.

IARC, "Long-Term and Short-Term Screening Assays for Carcinogens: A Critical Appraisal," Supplement 2, International Agency for Research on Cancer, Lyon, France, 1980.

IARC, "Approaches to Classifying Chemical Carcinogens According to Mechanism of Action," Technical Report No. 83/001, International Agency for Research on Cancer, Lyon, France, 1983.

IARC/IPCS Working Group Report, "Development and Possible Use of Immunological Techniques to Detect Individual Exposure to Carcinogens," Cancer Res., 42:5236-5239, 1982.

ICPEMC, International Commission for Protection Against Environmental Mutagens and Carcinogens, Working Paper 4/4, Mutation Res., 123:121182, 1983.

International Agency for Research on Cancer, Technical Report No. 83/001, "Approaches to Classifying Chemical Carcinogens According to Mechanism of Action," 1983.

International Agency for Research on Cancer Monographs, "Long-term and Short-term Screening Assays for Carcinogens: A Critical Appraisal," Supplement 2, Lyon, France, 1980.

Jenssen, D. and Ramel, C., "The Micronucleus Test as Part of a Short-term Mutagenicity Test Program for the Prediction of Carcinogenicity Evaluated by 143 Agents Tested," Mutation Res., 75:191-202, 1980.

Karstadt, M., "A Survey of Epidemiologic Data on Humans Exposed to Animal Carcinogens," Eds. R. Peto and M. Schneiderman, Quantification of Occupational Cancer, Cold Spring Harbor Laboratory, Cold Spring Harbor, Cold Spring Harbor, New York, pp. 223-245, 1981.

Kennedy, A.R., Cairns, J. and Little, J.B., "Timing of the Steps in Transformation of C3H 10T1/2 Cells by X-irradiation," Nature, 307:85-86, 1984.

Kligerman, A.D., Wilimer, J.L., and Erexson, G.L., "The Use of Rat and Mouse Lymphocytes to Study Cytogenetic Damage After in vivo Exposure to Genotoxic Agents," Banbury Report No. 13, Indicators of Genotoxic Exposure, Eds., B.A. Bridges, B.E. Butterworth, and I.B. Weinstein, Cold Spring Harbor Laboratory, pp. 277-281, 1982.

Kligerman, A.D., Wilimer, J.L., and Erexson, G.L., "Characerization of a Rat Lymphocyte Culture System for Assessing Sister Chromatid Exchange: After in vitro Exposure to Genotoxic Agents," Environ. Mutagenesis, 3:531-543, 1981.

Kligerman, A.J., Strom, S.C., and Michalopoulos, G., "Sister Chromatid Exchange Studies in Human Fibroblast-Rat Hegatocyte Cultures: A New in vitro System to Study SCEs," Environ. Mutagenesis, 2:157-165, 1980.

Kucerova, M., Zhurkov, V.S., Polivkova, Z., and Ivanova, S.E., "Mutagenic Effect of Epichlorohydrin II, Analysis of Chromosomal Aberrations in Lymphocytes of Persons Occupationally Exposed to Epichlorohydrin," Mutation Res., 48:355-360, 1977.

Kuroki, T., "Mutagenesis Assays with Mammalian Cells," In IARC Long-Term and Short-Term Screening Assays for Carcinogens: A Critical Appraisal, Supplement 2, 1980.

Lambert, B., Hindblad, A., Holmberg, K., and Francesconi, D., "The Use of Sister Chromatid Exchange to Monitor Human Populations for Exposure to Toxicologically Harmful Agents," Sister Chromatid Exchange, Eds. S. Wolff, J. Wiley, pp. 149-182, 1982.

Larsen, K.H., Brash, D., Cleaver, J.E., Hart, R.W., Maher, V.M., Painter, R.D. and Sega, G.A., "DNA Repair Assays as Test for Environmental Mutagens: A Report of the U.S. Environmental Protection Agency Gene-Tox Program," Mutation Res., 98:287-318, 1982.

Latt, S.A., Allen, J., Bloom, S.E., Carrano, A., Falke, E., Kram, D., Schneider, E., Schreck, R., Tice, R., Whitfield, B., and Wolff, S., "Sister Chromatid Exchanges: A Report of the Gene-Tox Program," Mutation Res. 87:17-62, 1981.

Legator, M.S., "A Holistic Approach to Monitoring High-Risk Populations by Short-Term Procedures," Quantification of Occupational Cancer, Banbury Report No. 9, Cold Spring Harbor Laboratories, New York, 335-345, 1981.

Lijinsky, W., "Species Specificity in Nitrasomine Carcinogens," Organ and Species Specificity in Chemical Carcinogenesis, Eds., R. Langenbach, S. Mesnow, and J.M. Rice, Plenum Press, N.Y., pp. 63-74, 1983.

Lutz, W.K., "In vivo Covalent Binding of Organic Chemicals to DNA as a Quantitative Indicator in the Process of Chemical Carcinogenesis," Mutation Res., 65, 289-356, 1979.

McCann, J., Kaldor, J., Horn, L., and Siegel, R., "Application of Short-term Tests to Problems in Cancer Risk Assessment: Potency, the Salmonella Test and Inter-species Comparisons," In Proc. of EPRI, Workshop on Extrapolation from Animal Studies to Man, Carmel, Calif., (in press).

MacGregor, J.T., Wehr, C.M. and Langlois, R.G., "A Simple Fluoresencent Staining Procedure for Micronuclei and RNA in Erythrocytes Using Hoechst 33258 and Pyronin Y," Mutation Res., 120:269-275, 1983.

Meyer, A.L., "In vitro Transformation Assays for Chemical Carcinogens," Mutation Res., 115:323-338, 1983.

Mitchell, A.D., Casciano, D.A., Meltz, M.L., Robinson, D.E., San, R.H.C., Williams, G.M. and Von Halle, E.S., "Unscheduled DNA Synthesis Test: A Report from the U.S. Environmental Protection Agency Gene-Tox program," Mutation Res., (in press).

Mirsalis, J.C., Tyson, C.K. and Butterworth, B.E., "Detection of Genotoxic Carcinogens in the in vivo - in vitro Hepatocytes DNA Repair Assay," Environ. Mutagenesis, 4:553-562, 1982.

Muller, R., and Rajewsky, M.F., "Antibodies Specific for DNA Components Structually Modified by Chemical Carcinogens," J. Cancer Res. and Clinical Oncology, 102:99-113, 1981.

Narisawa T., Magadia, N.E., Weisburger, J.H., Wynder, E.L., "Promoting Effect of Bile Acids on Colon Carcinogenesis after Intrarectal Installation of N-methyl-N-nitro-n-nitrosoguanidine in Rats," J. Natl. Cancer Inst., 53:1093-1097, 1974.

NAS, "Identifying and Estimating the Genetic Impact of Chemical Mutagens," National Academy of Sciences, Washington, D.C.

NAS Briefing Panel, "Report of the Research Briefing Panel on Human Health Effects of Hazardous Chemical Exposures," National Academy Press, Washington, D.C.

Nesnow, S., Garland, H., and Curtis, G., "Improved Transformation of C3H/10T1/2 Cl 8 Cells by Direct- and Indirect-Acting Carcinogens," Carcinogenesis, 3:377-380, 1982.

National Toxicology Program Annual Plan, Fiscal Year 1984, U. S. Dept. of Health and Human Services, Public Health Service, February 1984.

Oglesby, L., Hix, C., Snow, L., MacNair, P., Seig, M. and Langenbach, R., "Bovine Bladder Urothelial Cell Activation of Carcinogens to Metabolites Mutagenic to Chinese Hamster V79 Cells and Salmonella Typhimurium," Cancer Research, 43:5194-5199, 1983.

Ohtsuru, M., Ishii, Takai, S., Higashi, H., and Kosaki, G., "Sister Chromatid Exchanges in Lymphocytes of Cancer Patients Receiving Mitomycin C Treatment," Cancer Res., 40:477-480, 1980.

Oshiro, Y., Balwierz, P.S., and Molinary, S.V., "morphological Transformation of C3H/10T1/2 Cl 8 Cells by Alkylating Agents," Toxicol. Lett., 9:301-306, 1981.

Osterman-Golkar, S., Ehrenberg, L., Segerback, D., and Hallstrom, I., "Evaluation of Genetic Risks of Alkylating Agents II: Hemoglobin as a dose monitor," Mutation Res., 34:1-10, 1976.

Osterman-Golkar, S., Hultmark, D., Segerback, D., Calleman, C.J., Gothe, R., Ehrenberg, L., and Wachtmeister, C.A., "Alkylation of DNA and Proteins in Mice Exposed to Vinyl Chloride," Biochem. Biophys. Res. Commun., 76:259-266, 1977.

Pelling, J.C., Slaga, T.J., DiGiovanni, J., "Formation and Persistence of DNA/RNA and Protein Adducts in Mouse Skin Exposed to Pur Optical Enantiomers of 7 beta, a alphadihydroxy-9 alpha, 10 alpha-epoxy-7,8,9,10-tetrahydrobenzo(alpha)pyrene in vivo" Cancer Res., 44:1081-1086, 1984.

Peraino C., Fry, R.J.M. and Staffeldt, E., "Reduction and Enhancement by Phenobarbital of Hepatocarcinogenesis Induced in the Rat by 2-acetylaminofluorene," Cancer Res. 31:1506-1512, 1971.

Peraino, C., Staffeldt, E.F., Haugen, D.A., Lombard, L.S., Stevens, F.J. and Fry, R.J.M., "Effects of Varying the Dietary Concentration of Phenobarbital on its Enhancement of 2-acetylaminofluorene-induced Hepatic Tumorigenesis," Cancer Res., 40:3268-3273, 1980.

Pereira, M.A., and Chang, L.W., "Binding of Chemical Carcinogens and Mutagens to Rat Hemoglobin," Chem. Biol. Interact., 33:301-305, 1981.

Perera, F.P., and Weinstein, I.B., "Molecular Epidemiology and Carcinogen-DNA Adduct Detection: New Approaches to Studies of Human Cancer Causation," J. Chronic. Dis., 35, 581-600, 1982.

Perry, P.E., "Chemical Mutagens and Sister Chromatid Exchange," Sister Chromatid Exchange, Eds. S. Wolff, J. Wiley, pp. 1-39, 1982.

Phillips, D.H., Grover, P.L. and Sims, P., "A Quantitative Determination of the Covalent Binding of a Series of Polycylic Hydrocarbons to DNA in Mouse Skin," Int. J. of Cancer, 23:201-208, 1979.

Pitot, H.C., Goldsworthy T., Moran S., Sirica, A.E. and Weeks, J., "Properties of Incomplete Carcinogens and Promoters in Hepatocarcinogenesis," Carcinogenesis, 7:85-98, 1982.

Pitot, H.C. and Sirica, A.E., "The Stages of Initiation and Promotion in Hepatocarcinogenesis, Biochem. Biophys. Acta., 605:191-215, 1980.

Poirier, M.C., "Antibodies to Carcinogen-DNA Adducts," J. Nat. Cancer Inst., 67:515-519, 1981.

Preston, R.J., Au, W., Bender, M.A., Brewen, J.G., Carrano, A.V., Heddle, J.A., McFee, A.F., Wolff, S., and Wassom, J.S., "Mammalian in vivo and in vitro Cytogenetic Assays: A Report of the U.S. EPA's Gene-Tox Program," Mutation Res., 87:143-188, 1981.

Putzrath, R.M., Langley, D., and Eisenstadt, E., "Analysis of Mutagenic Activity in Cigarette Smokers' Urine by High Performance Liquid Chromatography," Mutation Res., 85:97-108, 1981.

Radman, M., Geggo, P., and Wagner, R., "Chromosomal Rearrangement and Carcinogenesis," Mutation Research, 98:249-264, 1982.

Randerath, E., Agrawal, H.P., Reddy, M.V., Randerath, K., "Highly Persistent Polycyclic Aromatic Hydrocarbons-DNA Adducts in Mouse Skin: Detection by ^{32}p-post labeling Analysis", Cancer Letters, 20:109-114, 1983.

Raunio, H., Vahakangas, K., Saarni, H., and Pelkonen, O., "Effects of Cigarette Smoke on Rat Lung and Liver Ornithine Decarboxylase and Aryl Hydrocarbon Hydroxylase Activities and Lung benzo(a)pyrene Metabolism," Acta. Pharmacol. et Toxicol., 52:168-174, 1983.

Sega, G.A., "DNA Repair in Spermatocytes and Spermatids of the Mouse," Banbury Report No. 13, Indicators of Genotoxic Exposure., Eds. B.A. Bridges, B.E. Butterworth, and I.B. Weinstein, Cold Spring Harbor Laboratory, pp. 503-513, 1982.

Segerback, D., Calleman, C.J., Ehrenberg, L., Lofroth, G., and Osterman-Golkar, S., "Evaluation of Genetic Risks of Alkylating Agents IV: Quantitative Determination of Alkylated Amino Acids in Hemoglobin as a Measure of the Dose after Treatment of Mice with Methyl Methanesulfonate," Mutation Res., 48:71-82, 1978.

Shelby, M.D., and Stasiewicz, S., "Chemicals Showing No Evidence of Carcinogenicity in Long Term Two Species Rodent Studies: Their Use in Short Term Test Evaluations," Cellular and Toxicology Branch, National Toxicology Program, 1983.

Sina, J.F., Bean, C.L., Dysart, G.R., Taylor, V.I. and Bradley, M.O., "Evaluation of the Alkalyne Elution Rat Hepatocyte Assay as a Predictor of Carcinogenic Mutagenic Potential," Mut. Res., No. 5, 113:357-392, 1983.

Singer, B. and Grunberger, D., "Molecular Biology of Mutagens and Carcinogens," Plenum Press, N.Y., N.Y., 1983.

Sivak, A. and Tu, A., "Development of Detailed Methods and Protocols for Carcinogenesis Screening Using Cell Culture Assays," Final report to National Institute of Environmental Health Sciences, December 31, 1982.

Slaga T.J., Sivak, A. and Boutwell, R.K., Eds., "Mechanisms of Tumor Promotion and Cocarcinogenesis, Carcinogenesis, 2:558, 1978.

Slaga, T.J., Fisher, S.M., Nelson, K. and Gleason, G.L., "Studies on the Mechanism of Skin Tumor Promotion: Evidence for Several Stages in Promotion, Proc. Natl. Acad. Sci. USA, 77:659-3, 663, 1980.

Sorsa, M., Hemaninki, K., and Vainio, H., "Biologic Monitoring of Exposure to Chemical Mutagens in the Occupational Environment," Teratogenesis, Carcinogenesis, and Mutagenesis, 2:137-150, 1982.

Sorsa, M., Norppa, H., Vainio, H., "Induction of Sister Chromatid Exchanges Among Nurses Handling Cytostatic Drugs," Banbury Report No. 13, Indicators of Genotoxic Exposure, Eds. B.A. Bridges, B.E. Butterworth, and I.B. Weinstein, Cold Spring Harbor Laboratory, pp. 341-354, 1982.

Stetka, D.G., Minkler, J. and Carrano, A.V., "Induction of Long-Lived Chromosome Damage, as Manifested by Sister Chromatid Exchange, in Lymphocytes of Animals Exposed to Mitomycin-C," Mutation Res., 51:383-396, 1978.

Strauss, G.H.S., "Direct Mutagenicity Testing: The Development of a Clonal Assay to Detect and Quantitate Mutant Lymphocytes Arising in vivo," Banbury Report No. 13, Indicators of Genotoxic Exposure, Eds. B.A. Bridges, B.E. Butterworth, and I.B. Weinstein, Cold Spring Harbor Laboratory, 1982.

Streisinger, G., "Extrapolations From Species and From Various Cell Types in Assessing Risks from Chemical Mutagens," Mutation Res., 114:93-105, 1983.

Suk, W.A., Poiley, J.A., Raineri, R., Stever, A.F., and Tennant, R.W., "Chemical Enhancement of Anchorage Independent Survival of Retrovirus-infected Cells as an Indicator of Neoplastic Transformation." Submitted.

Swenberg, J.A., Dyroff, M.C., Bedell, M.A., Popp, J.A., Huh, N., Kirstein, U., Rjewsky, M.S., "O^4-ethyldeoxythymidine, but not O^6-ethyldeoxyguanosine, Accumulates in Hepatocyte DNA of Rats Exposed Continuously to Diethylnitrosamine", Proceedings of National Academy of Sciences, 81:1692-1695, 1984.

Taylor, J.H., "Sister Chromatid Exchange: Early Experiments with Autoradiography," Sister Chromatid Exchange, Eds. S. Wolff, J. Wiley, pp. 1-16, 1982.

Tennant, R.R., Personal Communication, 1983.

Tennant, R.W., Spalding, J.W. and Dunnick, J.K., "An Evaluation of Rodent Chemical Carcinogens and Noncarcinogens: The Relationship of Genetic Toxicity," 1983, (submitted for publication).

Thilly, W.G., Heong, P.M., and Skopek, T.R., "Potential of Mutational Spectra for Diagnosing the Cause of Genetic Change in Human Cell Populations," Banbury Report No. 13, Indicators of Genotoxic Exposure, Eds. B.A. Bridges, B.E. Butterworth, and I.B. Weinstein, Cold Spring Harbor Laboratory, pp. 453-470, 1982.

Thilly, W.G., deLuca, J.G., Furth, E.E., Hoppe, H., Kaden, D.A., Krolewski, J.J., Liber, H.L., Skopek, T.R., Slapikoff, S.A., Tizard, R.J., and Penman, B.W., "Gene-Locus Mutation Assays in Diploid Human Lymphoblast Lines," Chemical Mutagens, Principles and Methods for Their Detection, Vol. 6, Eds. F.J. deSerres and A. Hollaender, pp. 331-364, 1980.

Timourian, H., Bishop, J.B., Wyrobek, A.J., "Sperm Shape Abnormalities in the Offspring of Male Mice Treated with Triethylene Melamine (TEM)," Environ. Mut. 5, 90, 1983c.

Traul, K.A., Takayama, K., Kaschevsky, V., Hink, R.J. and Wolff, J.S. "A Rapid in vitro Assay for Carcinogenicity of Chemical Substances in Mamalian Cells Utilizing an Attachment-Independent Endpoint." J. Appl. Tox. 1:190-195, 1981.

Troll, W., "Effects of Tumor Promoters," Ann. N.Y.A.S., 407:301-302, 1983.

Trosko, J.E., Jone, C. and Chang, C-C., "The Role of Tumor Promoters on Phenotypic Alterations Affecting Intercellular Communication and Tumorigenesis," Ann. N.Y.A.S., 407:316-327, 1983.

Tsuchimoto, T. and Matter, B.E., "In vivo Cytogenetic Screening Methods for Mutagens, with Special Reference to the Micronucleus Test," Arc. Toxicol., 42:239-248, 1979.

Vainio, H., Sorsa, M., and Herminki, K., "Biological Monitoring in Surveillance of Exposure to Genotoxicants," Am. J. Ind. Med., 4:87-103, 1983.

Waters, M.D., Allen, J.W., Claxton, L.D., Garrett, N.E., Huang, S.L., Moore, M.M., Sharief, Y., and Strauss, G.H., "Evaluation of Genotoxic Effects in Human Populations," F.J. deSerres, and R.W. Pero (eds.), In Individual Susceptibility to Genotoxic Agents in the Human Population, Topics in Chemical Mutagenesis Series, Plenum Press, New York, 1983a, (to be published).

Waters, M.D., Garrett, N.E., Covone-deSerres, C.M., Howard, B.E., and Stack, H.F., "Genetic Toxicology of Some Known or Suspected Human Carcinogens," Eds. F.J. deSerres, and A. Hollaender, Chemical Mutagens: Principles and Methods for Their Detection, Plenum Press, New York, 1983b, (to be published).

Weinstein, I.B., "The Monitoring of DNA Adducts as an Approach to Carcinogen Detection," Ann. Rev. Pub. H., 4:409-413, 1983.

Weinstein, I.B., letter to Science, 219:794-796, 1983.

Weisbuger, J.H. and Williams, G.M., "Carcinogen Testing: Current Problems and New Approaches," Science 214:401-407, 1981.

Welsch, C.U. and Nagasawa H., "Prolactin and Murine Mammary Tumorigenesis, A Review," Cancer Res., 37:951-963, 1977.

Wolff, S., "Sister Chromatid Exchange," Annual Rev. General, 11:183-201, 1977.

Wyrobek, A.J., Gordon, L., Berkhart, J.G., Francis, M.C., Kapp, R.W., Letz, D., Malling, H., Topham, J.C., Whorton, M.D., "An Evaluation of Mouse Sperm Morphology Test and Other Sperm Tests in Non-Human Animals," A Report for the US EPA GENETOX Program. Mutation Res., 115:1-72, 1983a.

Wyrobek, A.J., Gordon, L., Berkhart, J.G., Francis, M.C., Kapp, R.W., Letz, G., Malling, H., Topham, J.C., Whorton, M.D., "An Evaluation of Human Sperm as Indicators of Chemically Induced Alterations of Spermatogenetic Functions", A Report for the US EPA GENETOX Program. Mutation Res. 115:73-148, 1983b.

Wyrobek, A.J., Gordon, L., Watchmaker, G., Moore, D., "Human Sperm Morphology Testing Description of a Reliable Method and Its Statistical Power." In: Banbury Report 13. Indicators of Genotoxic Exposure (Eds.: B.A. Bridges, B.E. Butterworth and I.B. Weinstein), Cold Spring Harbor Laboratories, pp. 527-541, 1982.

Wyrobek, A.J., Gordon, L., Watchmaker, G., Moore, D., "Genetic Control of Human Sperm Shaping," Human Genetics, 5:123a, 1984a.

Wyrobek, A.J., Moore, D., Watchmaker, D., Gordon, L., "Optimal Design and Cohart Size Requirements for Human Studies of Induced Sperm Concentration and Morphology Effect," J. of Occupational Medicine, in press.

Wyrobek, A.J., Watchmaker, G. and Gordon, L., "Sperm Morphology Testing on Mice." In: Handbook of Mutagenicity Test Procedures. Elsevier Scientific Publishing Co., N.Y. (Eds.: B.J. Kilbey, M. Legator, W. Nichols and C. Ramel), pp739-750, 1984b.

Yamasaki, H. and Weinstein, I.B., "Cellular and Molecular Mechanisms of Tumor Promotors and Their Implications with Respect to Risk Assessment." SGOMSEC Workshop on Quantitative Estimation of Risk to Human Health and from Chemicals," Rome, 1982.

Yunis, J.L., "The Chromosomal Basis of Human Neoplasia," Science 221:227-236, 1983.

Yuspa, S.H., Hennings, H., Lichti, U. and Kulesz-Martin, M., "Organ Specificity and Tumor Promotion," Basic Life Sci, 24:157-171, 1983.

Report of the Subpanel on Subchronic Studies and Related Issues

A. Introduction and Charge to Subpanel

As a result of the discussions held at the public meeting of the Ad Hoc Panel on May 17, 1983, on the agenda items for that meeting as well as additional issues that were raised, a Subpanel to consider Data Required from Prechronic Studies was established with Andrew Sivak, Arthur D. Little, Inc., as chairman and Richard Adamson, National Cancer Institute, and Perry J. Gehring, Dow Chemical Company as members.

The prechronic studies in the context of their use in the carcinogenesis bioassay program of the National Toxicology Program (NTP) should be examined in terms of the general use and design of this type of study as a toxicological tool. As the information in Table 1 reveals, the general protocols and endpoints for prechronic studies as described by several agencies, including one external to the United States, exhibit a remarkable similarity. Moreover, the use of the data by agencies as part of the overall toxicological dossier of a chemical suggests that NTP should consider the prechronic study as it now is formulated or could be modified to provide direct information on toxic hazards of chemicals, in addition to its initially stated purpose.

The paragraph quoted below from the "Guidelines for Carcinogen Bioassay in Small Rodents" sets forth this purpose as stated in 1976 (Sontag et al., 1976):

> "VII.C.1. Need. The purpose of the subchronic study is to predict the toxic effects which may occur in the animals during the chronic administration of the test agent. Based on the subchronic observations, a prediction is made of the maximum tolerated dose (MTD) that can be given to the animals in the chronic study without producing unwanted side effects."

TABLE 1

Prechronic Study Guidelines

	EPA Pesticide Assessment Guidelines (1982)	FDA "Red Book" (1982)	FDA IND/NDA Pharmacology Review Guidelines (1981)	OECD (1981)	EPA Health Effects Test Guidelines (1982)	NTP (1976)
Purpose	Pesticide registration support No-observed-effect-level	Food additives and color additives Safety assessment No observed adverse effects, no-effect-level	IND/NDA Pharmacology review guidelines (1981) Characterize pharmacology, toxicology, pharmacokinetics and metabolism of drugs for precautionary clinical decisions	Assessment and evaluation of toxic characteristics Select chronic dose levels Useful information and permissible human exposure	Select chronic dose levels Establish safety criteria for human exposure No-observed-effect-level	Predict dose range for chronic study
Species	Rat (20)*, Dog (8)	Rat (40), Dog (8)	Rat/Mouse/Other Rodents; Dog/Monkey/ Other non-rodents	Rat (20), Dog (8)	Rats (20), Dogs (8)	Fischer 344 Rats (100), $B6C3F_1$ Mice (120)
Doses	3 dose levels	3 dose levels	3 dose levels	3 dose levels	3 dose levels	5 dose levels
End Points	Ophthalmology Hematology Clinical Chemistry Histopathology (29)** Rat (30), Dog (33) Target Organs	Ophthalmology Hematology Clinical Chemistry Histopathology Rat (41) Target Organs	Ophthalmology Hematology Clinical Chemistry Histopathology Target Organs Behavioral and Pharmacological Effects	Ophthalmology Hematology Clinical Chemistry Histopathology (41) Target Organs	Ophthalmology Hematology Clinical Chemistry Histology (41) Target Organs	Weight Loss Histopathology Target Organs

* (X) Number of animals per test group

** (X) Number of tissues per animal

The relationship between the MTD and "unwanted side effects" was described as follows:

> "VII.C.7. MTD Determination. A MAXIMUM TOLERATED DOSE (MTD) should be selected for each sex of each strain to be used in the chronic study. The MTD is defined as the highest dose of the test agent given during the chronic study that can be predicted not to alter the animals' normal longevity from effects other than carcinogenicity. The MTD is estimated after a review of the subchronic data. Since these data may not always be easily interpretable, a degree of judgment is often necessary in estimating the MTD. The MTD should be the highest dose that causes no more than a 10% weight decrement*, as compared to the appropriate control groups; and does not produce mortality, clinical signs of toxicity, or pathologic lesions (other than those that may be related to a neoplastic response) that would be predicted to shorten the animal's natural life span. Other measurements (see VII.C.6) also may be used to aid in predicting the MTD."

Thus, the information base from the prechronic studies initially designed only to select the doses to be used in the chronic study was limited, with considerable emphasis on weight gain decrement as the controlling factor. As the program matured, it became apparent that the dose selection process could be refined by use of a broader range of information obtained directly from the prechronic study as well as information from other research programs (chemical disposition, genotoxicity) being carried out by NTP and others (Schwetz, 1983). A more detailed presentation of present procedures in using data from the prechronic studies is presented in Section V.

*"Although a depressed weight gain is a clinical sign of toxicity, this particular effect is acceptable when estimating the MTD."

The Food Safety Council (1980) has presented in its Proposed System for Food Safety Assessment a detailed discussion of subchronic studies which addresses many of the issues of concern to this panel. The Department of Health and Social Security (1982) of the United Kingdom has recently published a set of "Guidelines for the Testing of Chemicals for Carcinogenicity" which also considers the factors associated with design and interpretation of carcinogenicity studies.

In addition to the dose selection issue, a number of additional operational factors were identified that had major impacts on the prechronic studies and chronic bioassay and were perceived to be appropriate for reevaluation and possible modification to improve the interpretability of the results and their relationship to assessments of toxic effects in humans.

As a consequence of discussions among members of the Ad Hoc Panel and representatives of NTP who provided technical guidance, the following issues were reviewed:

1. Chemical selection process.
2. Suitability of continued use of F344 rat and B6C3F_1 mouse.
3. Establishing dose range, number of doses, administration route and vehicle:
    a. Use of toxicological parameters (target organs, weight loss, organ function, histopathology, clinical chemistry, hematology, other toxic signs);
    b. Use of chemical disposition and metabolism data, and human exposure information;
    c. Factors affecting dose route and vehicle; and
    d. Use of other supporting information (short term test data).

The process used was to develop a position paper on each item that would serve as a basis for the recommendations offered at the end of each section.

B. The Chemical Selection Process

Recommendations

1. The development by NTP of a set of definitive criteria for each of the selection elements would provide a more uniform and justifiable process for chemical selection.
2. NTP should consider a methodology to make appropriate use of the exposure factor in the chemical selection process given its important role in assessment of risk.
3. The NTP Board of Scientific Counselors should ensure that the opportunity is maintained and enhanced for participation of interested parties in the review of decisions made between the subchronic test and the chronic bioassay.

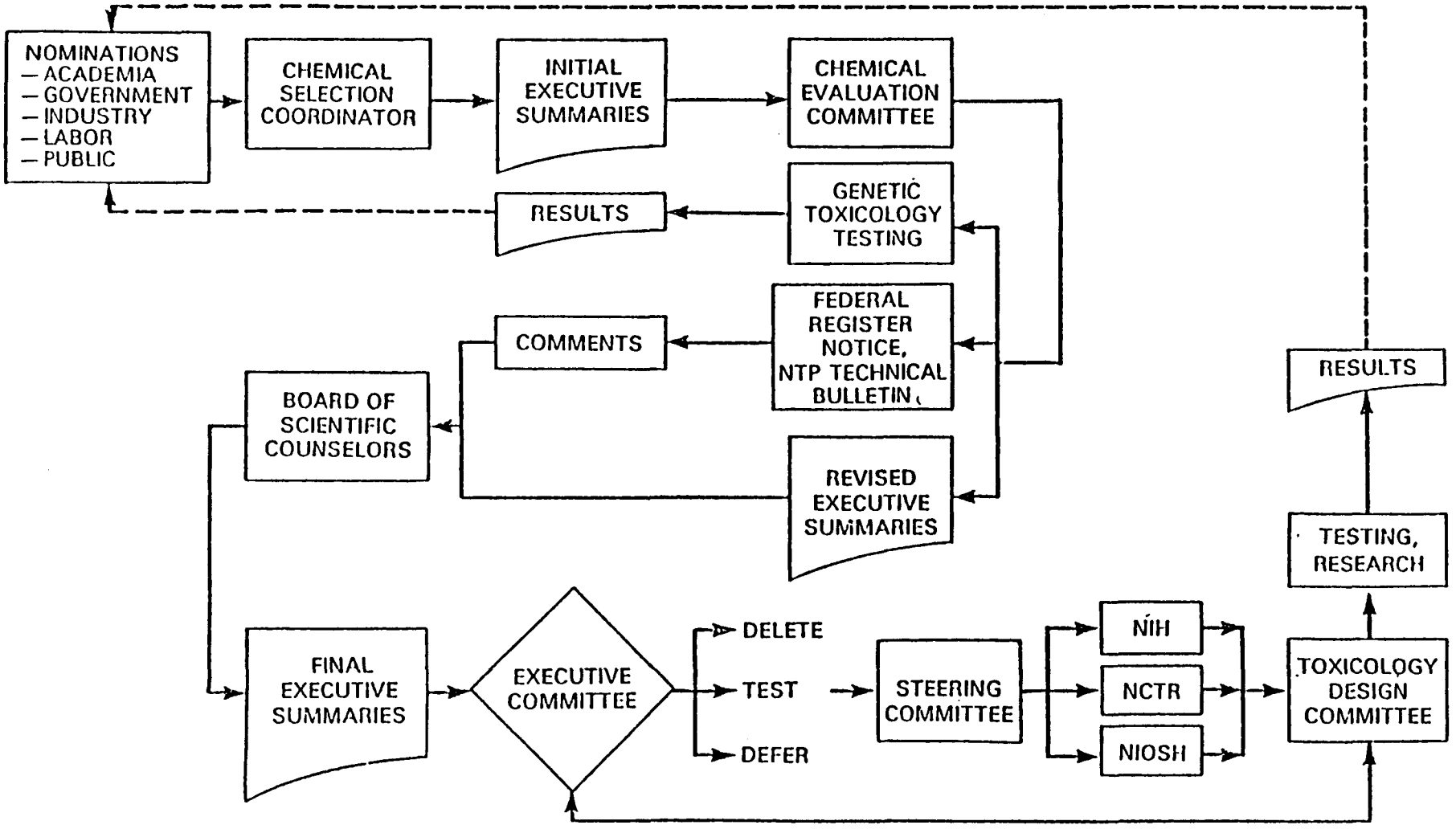
Background

Although the chemical selection process for NTP is not directly part of the prechronic studies area, its important early role in determination of which chemicals will be placed in the assay stream indicates that this process should be evaluated. We are aware of the National Academy of Sciences "Testing Needs Study" and one of their primary goals to "integrate the intensity of selected toxicity data elements . . . . to set priorities." In light of public comments that have been made with respect to the chemical selection process, the prechronic Subpanel included its consideration in our evaluation.

The chemical nomination and selection process and the results of the efforts in this area over the last year are set out in some detail in the NTP 1983 Annual Plan. Figure 1 displays a flow chart that shows a number of key elements in the process including (1) access to the nomination process by any interested party, (2) multiple review steps and (3) participation

FIGURE 1

NTP CHEMICAL NOMINATION AND SELECTION PROCESS



in this review by members of appropriate Federal agencies. The lineage of the process, though somewhat cumbersome, does appear to address the needs of the different constituencies involved. Several issues were raised by commenters from the public sector. The first of these dealt with the quality of the executive summaries prepared for the chemical evaluation committee. According to the commenters, the quality of these summaries was highly variable and in some cases incomplete and/or inaccurate. Concern was expressed that key decisions on setting priorities for entry of chemicals into the bioassay process may have resulted in the bioassay of chemicals of lesser urgency than those that were deferred or not tested at all.

A second related issue that was raised by a number of commenters related to external comment and peer review at the prechronic/chronic interface. In earlier times, the operation of the testing program was such that opportunities for comment at the prechronic/chronic interface were limited. As the program matured, information relating to testing plans and schedules for the components of the bioassay have been publically announced in the Federal Register and the NTP Annual Plan and contain invitation for public input. There are examples of public participation in the design of recent chronic studies (NTP TR 249 - amosite asbestos; also on-going studies with vinylidene fluoride, p-nitroaniline, trichloroethylene, hexachloro-1,3-butadiene and 1,3-butadiene). Thus, there exist provisions for a regular and consistent opportunity for public input. The balance between public participation in the decision process and the need to carry out an efficient program without undue delay appears to be the crux of the debate. The availability of a matrix of criteria that could be applied impartially and without undue influence of vested interests regardless of origin could assist in guiding the design factors in a chronic bioassay.

The development and judicious application of such a matrix based on the best data available (weight gain, organ specific toxicity, clinical chemistry, pharmacokinetics) at the time would tend to reduce the concerns of those external to the program and obviate to some degree the need for multiple peer reviews. An important issue in the use of criteria is that they be used with flexibility recognizing the imperfect nature of the data that will be used and the basic scientific problems in experimental design of carcinogenesis studies.

One important segment of information that has not been publicly available in bioassay reports until relatively recently is a disclosure of the scientific rationale for selection of the doses in the chronic study. Even in recent reports, the description of the manner in which the available data from the subchronic studies and other data are used to arrive at the chronic doses could be explained in more detail and made clearer and more definitive.

The third issue raised addressed the criteria used in application of the chemical nomination elements (Table 2). Although the elements span the range of information suitable for setting priorities for chemical selection, there are no defined criteria against which the data elements can be tested and weighted in the nomination and selection process. The development and application of a set of criteria would aid in providing a more rigorous basis for chemical selection. Among the chemical nomination elements, there is no specific identification of human exposure in either

TABLE 2

NTP Chemical Nomination Elements

I. Chemical Identification
   A. Chemical Abstracts Service (CAS) preferred name
   B. Common or generic name and synonyms
   C. CAS Registry Number
   D. Chemical class and related compounds
   E. Physical and chemical properties
      1. Physical description
      2. Structural and molecular formula and molecular weight
      3. Melting and boiling points
      4. Solubility
      5. Stability and reactivity
      6. Other relevant information
   F. Commercial product(s) composition
   G. References
II. Production, Uses, Occurrences, and Analysis
   A. Production
      1. Source and synthesis, year and pathway of first production
      2. Current production and pathway
   B. Uses
   C. Occurrence in the Environment
      1. Naturally occurring
      2. Air, water, and soil
      3. Occupational
   D. Analysis
   E. References
III. Toxicology
   A. Human data, case reports, and epidemiological studies
   B. Experimental animal information
   C. In vitro and other short-term tests
   D. Other relevant information
   E. References
III. Disposition and Structure-Activity-Relations
   A. Absorption, distribution, metabolism and excretion
   B. Structure-activity correlations and considerations
   C. References
V. Ongoing Toxicological and Environmental Studies in the Government, Industry, and Academia
VI. Rationale for Recommendation and Suggested Studies

magnitude or frequency used as a key factor. Since exposure is a dominant if not controlling factor in the assessment of risk, it would appear that its use in the nomination sequence would also be dominant. While detailed exposure scenarios may not be readily available for a large number of chemicals, especially those now under consideration, an effort in this direction to obtain a more quantitative exposure analysis would add considerable power to the chemical selection process as well as providing guidance in setting a dose range for the chronic bioassay.

C. Suitability of Continued Use of Fischer 344 Rat and B6C3F_1 Mouse

Recommendations

1. For the present, it is recommended that NTP maintain the two species, (Fischer 344 rat and B6C3F_1 mouse) presently used for carcinogenesis bioassays in the NTP program.
2. Based on data available both within the program and without, the NTP Board of Scientific Counselors should explore whether continued use of both a rat and mouse strain is needed for detection of carcinogens given the range of assay tools available to the oncology research community.
3. If a determination is made to maintain a two species bioassay protocol, give serious consideration to replacement of the B6C3F_1 mouse with a strain having an established lower and less variable spontaneous incidence of important tumors that are induced by chemicals. In addition, continued investigation of the use of other species as adjuncts or replacements for the ones now in use should be undertaken.

Historical Rationale for Selection of Animal Strains for Carcinogenicity Studies

The recent description of the history of the carcinogenesis bioassay

program at the National Cancer Institute by Elizabeth Weisburger (1983) provides a modest insight into the rationale for selection of the Fischer 344 rat and $B6C3F_1$ mouse as the animals of choice to carry out chronic carcinogenesis studies.

> "The search for better and more sensitive models for testing led to extensive tests on the development of mammary cancer in virgin female Sprague-Dawley rats. The use of infant mice and rats to increase sensitivity was explored thoroughly and several publications resulted therefrom. The Sprague-Dawley derived rat from Charles River Laboratories was also used in several other exploratory projects and in some bioassays. However, the large size and numerous spontaneous tumors of this strain of rat detracted from its value. In 1970 the F344 rat, which was easy to maintain, sensitive to various types of chemical carcinogens, and had a relatively low spontaneous tumor rate (except for the testicular Leydig cell tumors of aged nonbreeding males), was therefore adopted as the standard rat for the bioassay program. The CD-1 HaM/ICR mouse was also found to have an unacceptable high spontaneous tumor incidence and relatively short survival and was subsequently replaced in the bioassay program by the $B6C3F_1$ mouse."

The issue of selection of appropriate experimental animals for biological tests, specifically cancer studies, has also been commented upon by several members of the carcinogenesis research community.

At a symposium in 1967, Shimkin made several interesting observations that bear recalling. In particular, his views on the necessity for inbred strains of animals.

> "The dictum that 'inbred mice are as essential to the biologist as pure chemicals are for the chemist' needs re-examination. The fact is that

practically all key discoveries in chemical carcinogenesis were made, and continue to be made, on less exacting biological material which is easier and cheaper to procure and which is often more hardy and longer lived. We now have added pathogen-free and bacteria-free and newborn animals, which certainly have intriguing uses. But such animals, just as homozygous strains, are not necessarily more desirable than more prosaic ones for many biological and carcinogenic experiments. It is our hope that official standards will not be promulgated in regard to animal systems. More appropriate is good experimental design, with adequate controls, as selected by the individual investigator."

In contrast, Festing (1979) has, for a number of years, presented detailed arguments supporting the use of inbred strains of rodents for chronic toxicity and carcinogenicity studies. In his view, the use of inbred strains will aid in achieving "uniform experiment material" which will lead to a reduction of variability in the responses. With proper controls of environment, nutrition and infectious disease and allocation of animals to test control groups in an orderly and comparable fashion, the probability of false positive (Type I) and false negative (Type II) errors should be substantially reduced. A similar position supporting the use of inbred strains of test animals has been offered by Haseman and Hoel (1979) and also by Shimkin (1974) who proposed a general set of criteria that could be applied to the selection process for test animals in a carcinogenicity study.

They are:

1. Availability
2. Economy
3. Sensitivity to carcinogens
4. Stable as to response

5. Similarity to man in regard to metabolism
6. Similarity to man in regard to pathology responses

Most regulatory and government advisory bodies have not taken positions with respect to recommending specific strains of animals for use in carcinogenesis or other chronic exposure studies. For example, the British Committee on Carcinogenicity of Chemicals in Food, Consumer Products and the Environment (1982) in its "Guidelines for the Testing of Chemicals for Carcinogenicity" recognized the various competing factors in animal strain selection and concluded, "Extensive background or experience of a particular inbred or outbred strain in a particular laboratory may be a strong reason for choosing it. Where the test substance is chemically related to a known carcinogen, the choice of strains may be influenced by the strain used in the tests of the related substance. Otherwise the Committee felt unable to come down firmly in favor of specific inbred or outbred strains of rats or mice."

The Committee's report continues:

"Many strains of rats and mice have high 'spontaneous' incidences of particular kinds of neoplasm. For instance, particular strains of mice, both inbred and outbred, may be exceptionally prone to the development of adenomatous tumours of the lungs, benign and malignant parenchymal cell neoplasms of the liver, malignant lymphoma or mammary tumours, and some strains of rats may be prone to the development of mammary tumours, interstitial-cell tumours of the testis or pituitary adenomas. Exposure of such strains to test substances may be associated with an increased incidence of the same kind of tumours. Opinion is divided as to the extent to which such increases should be regarded as evidence of carcinogenic potential for man. Similarly, opinion is divided as to whether high-spontaneous tumour incidence

strains should be used in preference to low-spontaneous tumour incidence strains for carcinogenicity testing."

The Food Safety Council (1982) in its "Proposed System for Food Safety Assessment" similarly sets out some general principles in the selection of test species.

It is only considerably after the fact that the xenobiotic metabolic characterization has been examined in the test strains. It should be emphasized that the NTP embarked on acquiring such metabolic data in an expeditious fashion upon acquiring responsibility for the bioassay program.

The overriding issue, by far, in the continued use of the originally selected strains of test animals is the impact of the occurrence of spontaneous hepatic neoplastic tumors in $B6C3F_1$ mice on the interpretation of the effects of chemicals in inducing these kinds of lesions. For this evaluation, each of the test strains will be considered separately, followed by a discussion of overall test animal selection strategies.

The Fischer 344 Rat

At the time of the selection of the Fischer 344 rat for use in the NCI carcinogenesis bioassay studies, there was not a large data base on the sensitivity of this strain to carcinogens or on the metabolic capability of the strain for xenobiotics. Neither was there a detailed understanding of the profile of spontaneous tumors over the planned two year interval of the bioassay. In some of the earlier studies in the NCI Bioassay Program, Osborne-Mendel rats were used; however, this strain has not been used extensively by the NTP in its bioassay studies.

As the data base has grown, it has become evident that this strain of rat is vigorous, with good survival characteristics (Solleveld *et al.*, in

press). Moreover, the spontaneous tumor profile, which is now well documented, does not appear to have any substantial impact on the interpretation of chemically induced tumor response (Haseman, 1983).

Recent analyses of the responses of Fischer 344 rats in the NTP program have revealed some interesting data that bear further study as well as providing information that supports the continued use of this strain of rat as a test animal in carcinogenesis bioassays. Haseman's (1983) evaluation of patterns of tumor incidence among 25 feeding studies in the NTP program revealed a negative correlation between incidences of liver tumors and leukemia/lymphoma. In addition, reduction in the incidence of breast fibroadenomas was associated with decreased weight gain. Evaluation of the spontaneous tumor spectrum reveals only one tumor type (testicular interstitial cell tumor) that reaches a very high incidence (92%) in the 104 week assay interval. Other tumors that exhibit greater than a 20% incidence in controls are mononuclear cell leukemia and adrenal medullar pheochromocytoma in males and breast fibroadenoma, mononuclear cell leukemia and pituitary gland adenoma in females. The comparable survival (median lifespan of 28 months) of males and females and the observation that the variety of neoplastic lesions in animals carried for their lifespan was not greater than in animals killed between 110 and 116 weeks of age, although the overall frequency of these lesions increased markedly as the animals aged beyond the time they would usually be employed in a standard bioassay.

The $B6C3F_1$ Mouse

The use of the $B6C3F_1$ mouse as one of the strains in the carcinogenesis bioassay program has spawned a considerable debate that emanates from the occurrence of spontaneous liver tumors. This characteristic of substantial

hepatic tumor incidence in untreated male mice has raised questions about the interpretation of experiments in which an increased liver tumor incidence is observed associated with chemical exposure. Because of the attention focused on this issue and its apparent importance in the determination of the carcinogenicity of several classes of chemicals, especially the chlorinated hydrocarbons of both the aliphatic and aromatic types, it has been considered by several individual investigators as well as by a number of workshops and symposia (Squire and Levitt, 1975; Society of Toxicologic Pathologists, 1982).

The basic source of the debate appears to center on the identification of proliferative lesions in the livers of mice of the $B6C3F_1$ strain and the fate of those lesions. At one end of the spectrum, the position is espoused that proliferative lesions whether hyperplastic nodules, benign adenomas or carcinomas are all indicative of carcinogenic potential (Ward et al., 1979). In contrast, other arguments have been made that evidence for malignancy is essential before designating a chemical as a carcinogen (Newberne, 1982; Newberne et al., 1982; Vesselinovitch, 1982). In addition, the wide variation in spontaneous liver lesions in male $B6C3F_1$ mice among the various test laboratories, and even within a single laboratory, among those participating in the carcinogenicity bioassay program raises further questions with respect to the interpretation and use of these lesions as markers for direct induction of carcinogenicity by a chemical agent (Nutrition Foundation, 1983). Indeed, the variation in occurrence of hepatic proliferative lesions among mice appears to be a general characteristic of this species (Grasso and Hardy, 1975).

Maronpot and Boorman (1982) summarize the NTP position with respect to liver tumor responses:

"Examination of factors involved in the interpretation of liver tumor responses for the four bioassays discussed leads to some general conclusions regarding the current philosophy of assessing carcinogenicity on the basis of two-year rodent bioassays. First, there is agreement among most scientists that induction of hepatocellular carcinomas early in the course of a bioassay, as in the case of the pentachloroethane bioassay, is indicative of carcinogenicity under bioassay conditions. Furthermore, tumor frequency data must be appropriately analyzed when there is early mortality in the bioassay. The example of the pentachloroethane bioassay also exemplifies this point. Second, the occurrence of hepatocellular tumors in more than one species and/or more than one sex makes judgment regarding carcinogenicity more convincing. This was exemplified by the DEHP bioassay results. Third, neither the practice of combining benign and malignant liver tumors nor analyzing them separately for purposes of statistical determination of carcinogenicity is universally accepted. NTP scientists are in the process of establishing policy relative to combining benign and malignant tumors in the near future. Fourth, consideration of a chemical to be carcinogenic on the basis of an increase in hepatocellular adenomas (mice) or neoplastic nodules (rats) will remain controversial until such time as there are data to clearly indicate the biological behavior and significance of the benign liver tumors."

Because a substantial data base has been developed with the $B6C3F_1$ mouse, there is a natural reluctance to change the test strain to one that might present fewer problems. Yet, if the continued use of the present

strain does not provide a level of discrimination to yield reasonably definitive results, its continued use must be weighed seriously. Important factors that should be considered in changing the test strain of mouse relate to evidence that the strain now used does not provide an adequate level of discrimination and that an alternative strain with superior characteristics in terms of longevity, responsiveness and level of spontaneous tumors has been identified. A compilation of carcinogenicity data from studies using alternative strains would aid in making an appropriate decision in this matter.

One final issue that merits mention is the genetic integrity of the $B6C3F_1$ mouse test strain. Since the occurrence of spontaneous mouse hepatomas is strongly genetically dependent (Grasso and Hardy, 1975), although the source of this genetic influence has not been elaborated, the occurrence of genetic impurity could have dramatic effects on both spontaneous and induced liver tumor incidence, especially related to the contribution of the C3H gene pool to the hybrid. The application of the NTP genetic screening program as described in the 1984 Annual Plan should provide the quality assurance needed to remove genetic variability as a factor in the chronic bioassay. Further discussion of this issue may be found in Section D of the Report of the Subpanel on Design of Chronic Studies.

Beyond the question of the appropriate mouse strain to use in carcinogenicity studies, the merits of using only a rat strain for carcinogenicity studies has been discussed (von Wittenau and Estes, 1983). With the wide range of short term tests that are applicable and a commitment to develop pharmacokinetic data on bioassay compounds, employing the resources used in a long-term bioassay with a second species that may not be needed, and applying these resources to additional studies of mechanisms and/or additional substances is an attractive possibility that bears

further study. There was concern that one species might not adequately detect false negatives and this issue requires further analysis.

Finally, as the program has evolved a variety of strains of rats have been employed for comparison of toxicologic and metabolic response for a select group of chemicals. The rat strains being used other than the Fischer 344 strain are Osborne-Mendel, ACI 9935, August 28807, Marshall, Long-Evans, Sherman and Wistar. For the series of asbestos studies that have been carried out by NTP, the species of choice was Syrian golden hamster and for skin painting studies, CD-1 mice have been generally used.

D. Toxicological and Chemical Disposition Data for Selection of Doses For Chronic Studies

Recommendations

1. The use of the MTD as described in this report (see also Chronic Studies Chapter) should be continued. The rationale for selection of the doses for a chronic study and the procedures relating to utilization of the data from subchronic studies and other sources in this process should be included in all NTP bioassay reports.
2. NTP should continue to develop criteria and methodologies to evaluate and employ toxicological and pharmacological data as well as human exposure estimates where appropriate to select doses for the chronic study.
3. Pharmacokinetic studies should continue to be conducted before or during the prechronic phase of a bioassay so that a data set as complete as possible will be available to aid in the design of the chronic protocol and the interpretation of the results of the chronic study.

Background

As indicated earlier, the conduct of the prechronic studies and the

procedures for design of the chronic study are undergoing a continuing evaluation within the NTP bioassay based on new information that is being developed as part of the research program. A description of the present methodology prepared by Dr. Eugene McConnell of the NTP is included in this introduction to establish the framework for the following discussions and recommendations.

The maximum tolerated dose (MTD) is that dose which, when given for the duration of the chronic study as the highest dose, will not shorten the treated animals longevity from any toxic effects other than the induction of neoplasms (Schwetz, 1983). The MTD should not cause morphologic evidence of toxicity of a severity that would interfere with the interpretation of the study. For example, necrosis of a degree to be associated with a significant amount of regeneration may complicate the interpretation of a neoplastic response in that organ. Thus, toxicity and pathology criteria from the subchronic study are the primary criteria for setting the MTD.

The spread of doses below the MTD in chronic studies is determined by one or more of the following factors.

1. The slope of the dose reponse curve in the prechronic studies.
2. The need for a no effect dose.
3. Doses should rarely be spread by more than a factor of 10.
4. Relevance to exposure guidelines such as a TLV.
5. The pharmacokinetics of the chemical such as saturation and altered metabolism.

The inclusion in the final bioassay report of a description of the process of dose selection for the chronic bioassay, including the data elements that were employed and the rationale for dose selection based on these data as well as any other input from NTP staff or public sources would aid in the evaluation of the results of the bioassay.

The discussions concerning the use of toxicity and pharmacokinetic data in dose selection for the chronic study was prepared by a working group selected and chaired by Dr. Perry Gehring. These discussions in this position paper (Appendix I) are intended to indicate that the design of a chronic study is a complex issue that requires the use of all the supporting information that can be appropriately generated. As indicated in the body of the discussion of Appendix I, the concepts are presented as models of the kinds of procedures that can be employed, and are not meant as final methodological recommendations, but rather as an impetus for further work on this area. Additional discussion related to the design of the chronic study is presented in Section D of the Report of the Subpanel on Design of Chronic Studies.

Subchronic Toxicity Considerations and Guidelines for Using Subchronic Tests Results for Selecting Doses for Chronic Toxicity Tests

Before proceeding with a general discussion of toxicity data available for dose selection, it is appropriate to set down some definitions of what data are considered in the evaluation of prechronic and chronic toxicity.

Prechronic toxicity encompasses those effects observed in experiments in which the test animals are exposed to the chemical agent for a period of six months or less, usually 90 days. Common endpoints or markers of prechronic toxicity include organ specific and/or systemic pathology, body weight and organ weight alterations, mortality, clinical signs (including pharmacological or altered physiological parameters) and clinical laboratory measurements (hematology, urinalysis and clinical chemistry). Beyond these classical toxicity measures, the effects of test chemicals on the immune system or on hormonal balance, factors which could have a significant influence on the carcinogenic response, should be considered in the analysis of the subchronic studies for chronic study dose selection where such data are available.

Chronic toxicity studies elucidate effects in test species exposed for longer than six months, and generally focus on endpoints of organ-specific or systemic pathology, both neoplastic and non neoplastic. In addition, pathophysiology (abnormal function) is usually observed either as a primary or secondary response to treatment.

It is very important to recognize that all observations made in such studies, irrespective of duration of exposure, do not necessarily constitute toxicity per se but commonly span the entire range of effects from altered physiology, through pharmacology to toxicity.

The high dose for chronic toxicity testing is generally based upon endpoints achieved in prechronic tests ranging from 14 to 90 days in duration. The slope of the prechronic dose-response curve may influence the range and number of doses selected for the chronic toxicity test. The high dose for the chronic toxicity test generally lies between the dose level producing a positive endpoint and the dose producing a minimal effect in the three-month subchronic test.

1. Organ-Specific and/or Systemic Pathology Endpoints

The use of organ specific and pathology endpoints in a subchronic study are important criteria to aid in setting doses for a chronic study. Although no formal documentation for these criteria are presently available, the experience of toxicologists has involved some operational guidelines. An example of how such guidelines may be used is presented below. It should be considered as a model and not a formal mandate for the described procedure.

Evidence of moderate to marked necrosis, (3+ to 4+), degeneration, irritation, inflammation, or atrophy is generally considered a positive effect (endpoint) in prechronic studies. Such positive effects should be

found in at least 30% (3/10) of the test animals at a given dose to judge that dose to yield positive toxicity. If the degree of necrosis, degeneration, irritation, inflammation, or atrophy is minimal to mild (1+ to 2+), then such an effect must be present in at least 60% (6/10) of the animals to judge that dose to have positive toxicity. Additional judgments regarding positive effects must be made on the basis of the specific lesion and be influenced by the presence or absence of a dose-response for the endpoint in question and whether similar lesions are found in some of the vehicle or untreated controls. Of course, in rendering such judgments it is assumed that animals are free of complicating disease and that there is consensus agreement on the quality of the pathologic findings. In practice, the positive dose is often different for males and females of a given species; the same is true for a given sex of two species. Finally, the judgment must be made with respect to whether the pathologic lesion will affect the lifespan in a detrimental way.

An example of pathological findings obtained in a 90 day dermal study is depicted in Table 3. Although tempting, it is likely inappropriate to select the doses to be used in a chronic study from only the information set forth in this table. Prior to selection of the top dose it must be rationalized whether the purpose of the study is to reveal dermal or systemic toxicity or both. In any case, the pharmacokinetics of dermally applied doses should be ascertained so data resulting from a chronic study can ultimately be related to human exposure, both dermally and systemically.

TABLE 3

Pathological Alterations Observed in a 90 Day Dermal Study

Lesion	Dose 0	1	2	4	8	16
Lung, congestion	2/10	6/10	4/10	4/10	10/10*	9/10*
Thymus, atrophy	0/10	0/10	0/10	0/10	0/10	7/07
Liver, cytomegaly (cellular hypertrophy)	0/10	0/10	4/10	6/10*	10/10*	10/10*
Testes, tubular atrophy	0/10	0/10	0/10	0/10	5/10*	10/10*
Subcutaneous tissue, inflammation	0/10	0/10	0/10	0/10	9/10*	7/10*
Epidermis, hyperplasia	0/10	0/10	0/10	0/10	8/10*	3/10

* indicates $P < 0.05$ (2-tailed)

Comments on Example:

1. Lung congestion is not a sufficiently definitive lesion for determining an endpoint of toxicity.

2. Diagnosis of thymic atrophy should support thymic organ weight findings; modest reductions in organ weight in the absence of histological evidence of atrophy is too non-specific.

3. Liver cellular hypertrophy (cytomegaly) should support organ weight findings. The endpoint for this specific lesion should be determined in conjunction with pharmacokinetic data since hepatocellular cytomegaly may represent enzyme induction.

4. Although not life-threatening, testicular atrophy is a definitive endpoint of toxicity.

5. Subcutaneous inflammation and epidermal hyperplasia are target-organ-specific endpoints in dermal studies.

In Table 4, specific lesions from various subchronic toxicity studies are listed. These lesions provide rational endpoints upon which dose selection for chronic studies may be based.

In certain instances a constellation of pathologic effects may be present which are secondary to a generalized perturbation which is induced from the primary lesion. For example, the observation of splenic hemosiderosis, centrilobular fatty changes in the liver, gross presence of icterus, myocardial fiber degeneration and necrosis, and bone marrow hyperplasia are secondary effects of hemolytic anemia. Another example might be generalized metastatic calcification (heart, blood vessels in the brain and lung, gastric mucosa and submucosa), myelofibrosis of bone marrow, ulcers in the oral cavity, parathyroid hyperplasia, and renal tubular necrosis. In such a case, the majority of lesions are secondary to the uremia resulting from renal failure.

Judgments regarding selection of a toxic dose in the subchronic study should be influenced by the primary insult more than the secondary or systemic lesion responses.

2. Body Weight and Organ Weight Endpoints

Body weight and organ weight alternations are usually sensitive, quantitative indicators of toxicity (Weil and McCollister, 1963). Decreases in the weight of thymus or testes can often be detected in the absence of clear-cut lesions in these organs. The same is true for increases in liver or kidney weights. Body weight gain data are even more sensitive indicators of toxicity since they can potentially reflect effects not otherwise apparent. Body weight evaluations of necessity must be made in light of food or water consumption to help rule out decreased consumption due to poor palatability. Both absolute and relative organ weights should be evaluated in establishing dose effects.

TABLE 4

Selected Examples of Specific Toxic Lesions that are Acceptable Positive Endpoints in Subchronic (3-month) Studies

Lesion	Chemical
Myocardial necrosis	isoproterenol adriamycin
Neuronal degeneration	neuroleptics
Brain necrosis	mercury
Status spongiosis	hexachlorophene
Peripheral neuropathy	acrylamide triethyl tin
*Thyroid hyperplasia	PBB
Forestomach ulceration with *epithelial hyperplasia	2,4-diamionphenol
Hepatocellular necrosis	bromobenzene cycasin
Alveolar edema	paraquat alpha-naphthylthiourea
*Nasal cavity epithelial necrosis	formaldehyde acrolein
Testicular atrophy	cadmium DES DDT
*Renal tubular karyomegaly	trichloroethylene TRIS
*Urinary bladder-transitional cell hyperplasia	saccharin t-butanol
Thymic lymphoid depletion	TCDD DES

* These specific lesions may represent pre-neoplastic changes. In selecting a high dose (MTD) for a two-year study, it is appropriate to select a dose known to produce a potential pre-neoplastic lesion to maximize the likelihood of eliciting a carcinogenic effect.

Since body weights and organ weights are continuous variables, estimation of toxic endpoints is best achieved by the use of parametric statistical evaluation. Employing the fiducial limit of 0.05 (2-tailed tests), body weight or organ weight changes can be identified. Organ weight changes must be considered in light of other findings (e.g., pathologic alterations, clinical signs, reduced food consumption). Judgments regarding body weight in mice must be tempered by the profound effects that small changes have on measures of group average weights. In specific situations body weight and organ weight alterations can occur in 90-day studies at doses that do not produce pathologic alterations.

3. Clinical Laboratory Measurements

Hematologic, clinical chemistry, and/or urinary measurements are frequently performed in subchronic 90-day toxicity studies, usually only at the end. Routine hematologic measurements are justified on the basis of providing the most reasonable means for assessing toxic effects on the hematopoietic system. Routine clinical chemistry has generally been less sensitive than organ weights and histopathology in identifying metabolic and functional organ effects in rodents. Certain urinanalyses (volume, specific gravity, and microscopic evaluation of urine sediment) are of benefit when renal toxicity is present while other urinalyses procedures (dip-stick qualitative measurements) are rarely more sensitive than histopathology.

Because present day technology permits high precision in clinical laboratory measurements, there are frequent instances where statistically significant effects are flagged in the absence of clinical significance. For example, animals treated with the high dose in a 90-day subchronic toxicity test quite frequently have RBC counts that are 0.5 to 0.8 million

cells/mm^3 less than controls. While statistically significant at alpha <0.05, there is no apparent relationship to specific toxic effects of the chemical. This sort of change is probably more a reflection of general alteration of body physiology, possibly secondary to decreased food consumption and growth, than a true reflection of direct toxicity. Consequently, it is recommended that results from clinical laboratory studies be considered statistically significant at a fiducial limit of alpha <0.05, but that all results be evaluated for clinical significance in assessments of treatment-related endpoints.

While the potential utility of hematologic parameters can be intuitively appreciated as useful measures of toxicity to the hematopoietic system, experience with routine clinical chemistry and urinalyses have been less satisfactory. These latter measurements rarely identify toxic endpoints that are not specifically reflected in organ weight and histomorphologic alterations. If used at all, it is recommended that clinical laboratory tests be prudently selected and timed in 90-day as well as two-year study designs. Selective timing versus routine performance at a predetermined time interval may enhance the utility of clinical laboratory studies thereby permitting detection of initial organ insult with subsequent compensation. Such information would theoretically be useful in characterizing mechanisms of toxicity.

At the present time it is recommended that clinically significant changes in hematologic, clinical chemistry, or urinary measurements be used as supporting information to be correlated with more definitive positive toxicity endpoints. With the possible exception of hematologic results, clinical laboratory measurements as currently obtained are not generally sufficient as the sole reason for identifying toxic dose endpoints in

prechronic studies, although with certain chemicals they may have specific value in identifying a toxic response.

Indeed, the need for routine collection of complete clinical laboratory data is highly questionable and it is recommended that plans to collect such data should be implemented only when the results are judged to be worthy of collection.

4. Statistics

Some general thoughts on the use of statistical analyses of data are appropriate here. Statistical evaluation is a tool to assist in the determination of a positive lesion endpoint. However, it is neither a substitute for making informed judgments nor should statistical tests be inappropriately applied in evaluating toxicological data. Many toxicology studies utilize an exploratory strategy which involves measurements of as many as 30 different parameters in a control group versus groups treated with several different dose levels of a chemical. Although some toxicologists may use the t-test, ANOVA with multiple comparisons are typically used for each parameter. The data for each parameter are, however, from the same subjects; thus, the measurements are correlated to an unknown degree and the ability to define the true level of type I or type II error for the experiment as a whole is difficult to assess.

One possible solution to this problem is to use statistical methods such as t-tests and ANOVA with multiple comparisons as screening devices to indicate those variables that show statistical evidence of biological effects. The methods section should acknowledge the multiple comparisons issue and give some indication of the number of significant effects that might be expected by chance alone. The toxicologists must then examine the data, use all available information, and make a judgmental decision on a

case-by-case basis about the toxicologic significance of the various statistically signficant effects observed.

5. Mortality

Mortality is a highly definitive endpoint of toxicity in 90-day studies. Assuming that the cause of death is clearly associated with treatment, then high dose selection for the two-year study should be set lower than the lowest 90-day study dose with mortality. An example of mortality data which may be useful in selection of the top dose is given in Table 5. In the absence of attendant pathology, the top dose selected for a chronic study should be between the 4 and 8 dosage rates.

6. Clinical Signs

Generally clinical signs will be correlated with and support other endpoints of toxicity. However, in special situations clinical signs seen in prechronic studies will preclude using high doses for the two-year study which cause definitive pathological effects (e.g., anesthesia produced by inhaling high doses of an aerosol, prolonged "intoxication" associated with ingestion of alcohols).

7. Example of a Composite Data Set

Table 6 presents a typical composite data set from a subchronic evaluation. Assuming that the indicated doses are useful for assessment of the human risk, the top dose selected for a chronic dose should not exceed "2"; in many cases selection of "1" will be appropriate. The latter dose will in most cases significantly exceed allowable human exposure and still elucidate any insidious chronic toxicity affecting the liver and other tissues as well.

TABLE 5

Example - Mortality Data

Dose Group	0	1	2	4	8	16
Mortality	0/10	0/10	0/10	0/10	1/10	10/10

Comments on example:

Dose group 8 with mortality constitutes a toxic endpoint. Consequently, the high dose for the two-year study should be lower than 8.

Mortality = number dead/number treated.

TABLE 6

Example (90-Day Dermal Study) - Composite of Study Findings

Parameter	Doses					
	0	1	2	3	8	16
Liver/BW		alpha <0.01	alpha <0.01	alpha <0.01	p <0.01	died
Thymus/BW					p <0.01	died
Kidney/BW					p <0.01	died
Reticulocyte count					C.S.	died
Methemoglobin					C.S.	died
Lung, congestion	2/10	6/10	4/10	4/10	10/10	9/10
Thymus, atrophy	0/10	0/10	0/10	0/10	0/10	7/7
Liver, cytomegaly	0/10	0/10	4/10	6/10	10/10	10/10
Testes, atrophy	0/10	0/10	0/10	0/10	5/10	10/10
Subcutaneous inflammation	0/10	0/10	0/10	0/10	9/10	7/10
Epidermal hyperplasia	0/10	0/10	0/10	0/10	8/10	3/10
Mortality	0/10	0/10	0/10	0/10	1/10	10/10

C.S. = clinically significant

Pharmacokinetic Studies (Chemical Disposition)

In a very general sense, pharmacokinetics is the study of the dynamics of the fate of chemicals in the body. Perhaps the most significant contribution of pharmacokinetics in the design and/or interpretation of toxicology studies is the concept of dose-dependent or non-linear kinetics. Many of the physiological and biochemical processes which affect the disposition of chemicals in the body are capacity limited (i.e., saturable). When any of the rate-limiting processes involving absorption, distribution, metabolism, or excretion of a chemical become saturated, the internal concentration of chemical and/or activated metabolites may not be directly proportional to the administered dose. Instead, disproportionate increases or decreases may be observed, with concurrent effects upon toxicity. Levy (1968) has described some of the criteria which suggest that such saturation may be occurring:

1. The decline of levels of chemicals in the body does not follow an exponential time curve.
2. The biological half-life increases with increasing dose.
3. The area under the plasma concentration versus time curve (AUC) is not proportional to doses.
4. The composition of excretory products may be changed both qualitatively and quantitatively with increasing dose.
5. Competitive inhibition by other chemicals metabolized by the same enzymatic system(s) is likely.
6. Dose response curves show unusually large increases in response with increasing dose, starting with the dose level where saturation effects first become evident.

Another important area of pharmacokinetic studies is the ability to predict bioaccumulation. If a material is eliminated slowly, repeated

dosing may result in the introduction of a new chemical into the body before the last dose is gone. Consequently, the eventual concentration of this material in the test organism during a chronic study may be many times higher than anticipated. Polybrominated biphenyls (PBB) are examples of such chemicals. Determination of pharmacokinetic parameters allows the prediction of the level in the body *in toto* or tissues subsequent to any number of repetitive doses as well as the ultimate steady state levels resulting from continuous or interrupted exposures. For a more detailed discussion of the application of pharmacokinetics, the investigator should consult the following references: Gehring *et al.*, (1976), Tsuchiya and Levy (1972), O'Flaherty (1981), Gibaldi and Perrier (1982), Dayton and Sanders (1983).

Appendix I to this chapter describes a sample methodology that should be useful to aid in the dose setting process for the chronic bioassay. As indicated above, the establishment of the doses in a chronic study should be the result of an analysis of all the information available in the subchronic study such as 1) variations in weight gain, 2) physiological disturbances, 3) histopathology, 4) dose response of mortality and 5) pharmacokinetics. The presentation of this pharmacokinetic methodology (in Appendix I) per se is not intended to serve as a specific recommendation for its use in the NTP program. However, this or similar technology should be utilized to obtain data to aid dose selection and interpretation of results from chronic bioassays. NTP should encourage and support further work to refine and develop the application of pharmacokinetics.

Indeed, as indicated in Appendix I, additional methods are available to obtain more detailed information about pharmacokinetics, physiological disposition and metabolism of xenobiotics. Some of these methods are the

use of body compartment distribution models, target tissue binding, metabolic profile and whole body autoradiography.

As a companion piece to this discussion of dose selection in chronic studies, another review prepared by a group of experienced toxicologists should be consulted for details relating to this critical issue (Grice et al., 1984).

E. Factors Affecting Dose Route and Vehicle

These factors were also considered by the Subpanel on Design of Chronic Studies with the development of sets of recommendations that are consistent with those of this section.

Dose Route

Recommendations

1. **Exposure to compounds in the carcinogenesis bioassay should reflect the predominant human exposure route where possible taking into account the need to achieve adequate dose levels for an appropriate study.**
2. Should special circumstances indicate that an alternative exposure route may be necessary, an appropriate supporting rationale should be provided including the development of pharmacokinetic and toxicologic data associated with the surrogate exposure route.
3. Since inhalalation exposure represents a significant route of human exposure to environmental chemicals, NTP should develop a data base for a selected group of substances tested by several routes to aid in determining whether a single exposure route is adequate to assess carcinogenicity.

Background

Although the general principle of using a dosing route for chronic studies that was comparable to the exposure route in humans has been known and accepted in the toxicological community, there was also general acceptance of use of gavage techniques. The rationales for the use of intubation procedures included a more exact measurement of dosage than available by diet or drinking water procedures and/or ease of delivery of materials whose properties (solubility, palatability, corrosiveness, volatility) made them difficult to administer by the usual feeding or drinking modes of dosing (Weisburger and Weisburger, 1967).

In recent years however, the practice of using gavage administration in corn oil has come under considerable criticism. Although the fundamental reasons for this criticism are not easy to discern, those that have been offered include: 1) bolus administration results in highly atypical pharmacokinetic patterns compared to usual human exposures, except where the human exposure is in fact a bolus dose, 2) the test animals are subjected to stress on a daily basis with some likelihood of fatality resulting from faulty intubation techniques, 3) substantial increase in personnel resources to perform dosing as compared to diet or drinking water exposure, and 4) necessity to use vehicles (corn oil) that could have a substantial confounding influence on the final outcome of the bioassay. This issue of vehicle for gavage studies, especially as it relates to the use of corn oil, will be considered separately.

In connection with the use of diet and drinking water exposures to model oral intake in humans and the use of skin exposure procedures to model exposure by that route, Page (1977) has discussed extensively the methodology and problems associated with these procedures. The protocols

as they are laid out appear to cover the necessary details. Adherence to the protocols requires a rigorous quality assurance program to ensure that the dosage consumed by or applied to the test animals is consistent over the interval of the study.

The inhalation route is the source of exposure of many materials in the environment. Yet, because of the substantial resources in terms of cost and trained personnel associated with the thorough conduct of such bioassays, they have only had limited use in comparison to the numbers of compounds that would be desirable to test using this route of exposure.

Since a substantial number of volatiles have been identified that should be examined for carcinogenicity, and the resources have not been provided to test them by the most applicable route (inhalation), many of them have been tested by the gavage route using corn oil as the vehicle. Thus, the decision to gain carcinogenicity data for some chemicals by procedures using surrogate exposure routes has resulted in a substantial controversy as described above.

Beyond these policy decisions, there is the difficulty in inhalation exposure with rodents as models for humans. This problem is especially acute for particulate materials which display a substantial difference in disposition patterns that result from the small diameter, highly turbinated nasal passages of rodents although definitive documentary support for this view does not appear to be available. For all materials, the pattern of exposure is likely to be considerably different in obligate nose breathers such as rats and mice and humans. Based on an examination of the tumor types likely to be induced in experimental animals, the rat has been identified as the species that most closely meets the needs as an appropriate animal model (Laskin and Sellakumar, 1974). For special studies

intratracheal exposure in Syrian hamsters has been employed with considerable success by applying materials directly to the respiratory tissue (Saffiotti, 1968). Further consideration of this issue can be found in Section E of the Report of the Subpanel in the Design of Chronic Studies.

Vehicle

Recommendations

1. For those studies where a gavage route is deemed to be necessary, a search should be made to select a vehicle and/or dose delivery systems that would have little or no impact on the final carcinogenesis outcome. A discussion of the technical details relevant to this is presented in Section F and Appendix I of the Report of the Subpanel on the Design of Chronic Studies.
2. Regardless of the final decision with respect to the use of corn oil as a vehicle in gavage studies, it is necessary to examine the body of data within the bioassay program and elsewhere to determine the mechanisms by which corn oil has induced the effects that have been documented, and to resolve inconsistencies between studies with respect to gavage related effects.

Background

The gavage route of exposure appears to be the only procedure for which the vehicle has become an important issue. With appropriate assurances of test compound stability in food and drinking water and in atmospheres generated in inhalation chambers, there has been no major concern over the dosing vehicle.

Nonetheless, it has become clear that the use of corn oil as a vehicle in chronic gavage studies is not without some impact of its own on the test

animals in a number of studies. However, this finding is not consistent in the bioassay program. Three primary issues have emerged. The first, and possibly most critical, of these is the finding in some gavage studies that pancreatic tumors are enhanced in rats receiving corn oil alone. This observation is enigmatic in that the occurrence of pancreatic tumors in association with corn oil gavage has not been consistently observed over the interval of carcinogenesis bioassay program. Because the incidence of such tumors was further increased in test groups receiving test agents (methylene chloride, benzyl acetate), the problem is further complicated.

The second issue is the one of altered pharmacokinetics associated with administration of test chemicals as a bolus in a lipid solvent. There is substantial evidence that the pattern of metabolism varies considerably with respect to both rate and qualitative nature of the biochemical profile when dosing by gavage in corn oil and by inclusion in drinking water are compared (Newberne et al., 1979; Iritani and Ikeda 1982; Wade and Norred, 1976; Williams et al., 1983.

The third issue is the nutritional and physiological impact of the corn oil dosing. In particular, the lipid nutritional profile has been shown to affect dramatically the outcome of carcinogenicity studies (Rogers and Newberne, 1980; Newberne et al., 1982; Reddy et al., 1980; Rogers, 1983).

F. Use of Other Supporting Data (Short Term Tests)

Following a consideration of the applicability of short term test data for use in the design of a chronic bioassay protocol, it was determined that this information would provide only limited guidance. The primary uses for these tests at the present time appear to be in assisting the chemical selection process and in providing information to aid in clarifying the responses observed in a carcinogenesis bioassay. For the future, the possibilities for their wider utility in the decision process

to identify carcinogens has been considered by the Subpanel on Short Term Tests. In particular, the prechronic study can provide a resource of treated animals over a wide dose range that is now not employed to obtain valuable data on a variety of genetic endpoints. The Subpanel on Prechronic Studies concurs with the statements of the Short Term Tests Subpanel regarding the value of the prechronic and chronic studies for validation of the short term assay as a source of material for further study, as a means of developing parallel data and as a possible lead to work in humans.

REFERENCES

Bischoff, K.B., Dedrick, R.L., Zaharko, D.S., Longstreth, J.A., "Methotrexat Pharmacokinetics," J. Pharm. Sci., 60:1128-1133, 1971.

Dayton, P.G., Sanders, J.E., "Dose-Dependent Pharmacokinetics: Emphasis on Phase 1 Metabolism," Drug Metab. Rev., pp. 347-405, 1983.

Dedrick, R.L., "Animal Scale-Up," J. Pharmacokin. Biopharm., 1:435-461, 1973.

Department of Health and Social Security, "Guidelines for the Testing of Chemicals for Carcinogenicity," Report on Health and Social Subjects by the Committee on Carcinogenicity of Chemicals in Food, Consumer Products and the Environment, London, United Kingdom, 1982.

Eustis, S.L., Boorman, G., "Proliferative Lesions of the Exocrine Pancreas: Relationship to Corn Oil Gavage in the NTP Carcinogenesis Testing Program," In: 1983 NTP Working Paper, 1983.

Festing, M.F.W., "Inbred Strains and the Factorial Experimental Design in Toxicological Screening," ICLA Symposium, Utrecht, Holland, 1979.

Fiserova-Bergerova, V., Holaday, D.A., "Uptake and Clearance of Inhalation Anesthetics in Man," Drug Metab. Rev. 9:43-60, 1979.

Food Safety Council, "Proposed System for Food Safety Assessment," Final Report of the Scientific Committee of the Food Safety Council, Washington, DC, 1980.

Gehring, P.J., Watanabe, P.G., Blau, G.E., "Pharmacokinetic Studies in Evaluation of the Toxicological and Environmental Hazard of Chemicals," In: New Concepts in Safety Evaluation, Eds. M.A. Mehlman, R.E. Shapiro, H. Blumenthal, pp. 195-270, Wiley & Sons, New York, 1976.

Gibaldi, M., Perrier, D., Pharmacokinetics, 2nd Ed., Marcel Dekker, New York, 1982.

Grice, H.C., Arnold, D.L., Blumenthal, H., Emmerson, J.L., and Krewski, D. "The Selection of Doses in Chronic Toxicity/Carcinogenicity Studies," Chronic Toxicity/Carcinogenicity Studies," In: Current Issues in Toxicology, Ed. H.C. Grice, Springer Verlag, New York, pp. 6-49, 1984.

Grasso, P., Hardy, J., "Strain Difference in Natural Incidence and Response to Carcinogenesis," In: Mouse Hepatic Neoplasia, Eds. W.H. Butler, P.M. Newberne, Elsevier Press, Amsterdam, pp. 111-132, 1975.

Haseman, J.K., Hoel, D.G., "Statistical Design of Toxicity Assays: Role of Genetic Structure of Test Animal Population," J. Tox. Environ. Hlth., 5:89-101, 1979.

Haseman, J.K., "Patterns of Tumor Incidence in Two-Year Cancer Bioassay Feeding Studies in Fischer 344 rats," Fund. Appl. Toxicol., 3:1-9, 1983.

Iritani, N., Ikeda, Y., "Activation of Catalase and Other Enzymes by Corn Oil Intake," J. Nutr., 112:2235-2239, 1982.

Laskin and Sellakumar, "Models in Chemical Respiratory Carcinogenesis," In: Experimental Lung Cancer: Carcinogenesis and Bioassays, Eds. E. Karbe, J.F. Park, New York: Springer-Verlag, 1974.

Levy, G., "Dose-Dependent Effects in Pharmacokinetics," In: Importance of Fundamental Principles in Drug Evaluation, Eds. D.H. Tedeschi, R.E. Tedeschi, pp. 141-172, Raven Press, New York, 1968.

Maronpot, R.R., Boorman, G.A., "Interpretation of Rodent Hepatocellular Proliferative Alterations and Heptocellular Tumors in Chemical Safety Assessment," Toxicolog. Pathol., 10:71-79, 1982.

Mattsson, J.L., Park, C.N., Fundam. Appl. Toxicol., 2:16a, 1982.

National Toxicology Program, Annual Plan for Fiscal Year 1983, Research Triangle Park, NC, 1983.

Newberne, P.M., "Assessment of the Hepatocarcinogenic Potential of Chemicals: Response of the Liver," In: Toxicology of the Liver, Eds. G. Plaa, W.R. Hewitt, Raven Press, New York, 1982.

Newberne, P.M., Weigert, J., Kula, N., "Effects of Dietary Fat on Hepatic Mixed-Function Oxidases and Hepatocellular Carcinoma Induced by Aflatoxin B_1 in Rats," Cancer Res., 39:3986-3991, 1979.

Newberne, P.M., de Camargo, J.L.V., Clark, A.J., "Choline Deficiency, Partial Hepatectomy, and Liver Tumors in Rats and Mice," Toxicolog. Pathol., 10:95-109, 1982.

The Nutrition Foundation, "Report of the Ad Hoc Working Group on Oil Gavage in Toxicology," July 14-15, 1983, Arlington, VA, 1983.

The Nutrition Foundation, "The Relevance of Mouse Liver Hepatoma to Carcinogenic Risk," Report of the International Expert Advisory Committee, September, 1983.

O'Flaherty, E.J., Toxicants and Drugs: Kinetics and Dynamics, Wiley & Sons, New York, 1981.

Organization for Economic Cooperation and Development, "OECD Guidelines for Testing of Chemicals," Paris, France, 1981.

Page, N.P., "Concepts of a Bioassay Program in Environmental Carcinogens," In: Advances in Modern Toxicology, Vol. 3, Eds. H.F. Kraybill, M.A. Mehlman, Hemisphere Publ., Washington, DC, 1977.

Ramsey, J.C., Anderson, M.E., "A Physiologically Based Description of the Inhalation Pharmacokinetics of Styrene in Rats and Humans," Toxicol. Appl. Pharmacol., (submitted), 1983.

Reddy, B.S., Cohen, L.A., McCoy, G.D., et al., "Nutrition and it's Relationship to Cancer," In: Advances in Cancer Research, Vol. 32., Academic Press, Inc., 1980.

Rogers, A.E., "Influence of Dietary Content of Lipids and Lipotropic Nutrients on Chemical Carcinogenesis in Rats," Cancer Res., Suppl., 43:2477s-2484s, 1983.

Rogers, A.E., Newberne, P.M., "Lipotrope Deficiency in Experimental Carcinogenesis," Nutr. Cancer, 2:104, 1980.

Saffiotti, U., Cefis, F. Kolb, L.H., "A Method for the Experimental Induction of Bronchogenic Carcinoma," Cancer Res., Vol. 28, pp. 104-124, 1968.

Schwetz, B.A., "Prechronic Toxicological Tests - Scope and Use," In: NTP Working Paper, 1983.

Shimkin, M, Triolo, V.A., "History of Chemical Carcinogenesis: Some Prospective Remarks," In: Progr. Exp. Tumor Res., Vol. 11, pp. 1-20, Karger, Basel/New York, 1969.

Shimkin, M.B., "Section 2: Report of Discussion Group No. 1; Species and Strain Selection," In: Carcinogenesis Testing of Chemicals, from the Proceedings of the Conference on Carcinogenesis Testing in the Development of New Drugs, May 23-25, 1973, Washington, DC, 1974.

Society of Toxicologic Pathologists, Proceedings of the International Symposium of the Society of Toxicologic Pathologists: "Rodent Liver Nodules: Significance to Human Cancer Risk," Toxicologic Path., Vol. 10, Number 2, 1982.

Solleveld, H.A., Haseman, J., McConnell, E.E., "The Natural History of Body Weight Gain, Survival and Neoplasia in the Fischer 344 Rat, JNCI 72:929-940, 1984.

Sontag, J.M., Page, N.P., Saffiotti, U., "Guidelines for Carcinogen Bioassay in Small Rodents," Report to the Carcinogenesis Program, Division of Cancer Cause and Prevention, National Cancer Institute, Bethesda, MD, 1976.

Squire, R.A., Levitt, M.H., "Report of a Workshop on Classification of Specific Hepatocellular Lesions in Rats," Cancer Res., 35:3214-3223, 1975.

Tsuchiya, T., Levy, G., "Relationship Between Dose and Plateau Level of Drugs Eliminated by Parallel First-Order and Capacity-Limited Kinetics," J. Pharm. Sci., 61:541-544, 1972.

U.S. Environmental Protection Agency, "Health Effects Test Guidelines and Support Documents," Washington, DC, 1982.

U.S. Environmental Protection Agency, Pesticide Assessment Guidelines, Sub-Division F, "Hazard Evaluation: Human and Domestic Animals, Washington, DC, 1982.

U.S. Food and Drug Administration, IND/NDA Pharmacology Review Guidelines, Washington, DC, 1981.

U.S. Food and Drug Administration, "Toxicological Principles for the Safety Assessment of Direct Food Additives and Color Additives Used In Food," Washington, DC, 1982.

Vesselinovitch, S.D., "Liver Tumor Induction," Toxicol. Pathol., 10:110-118, 1982.

von Wittenau, M.S., Estes, P.C., "The Redundancy of Mouse Carcinogenicity Bioassays," Fundam. Appl. Toxicol., (in press), 1983.

Wade, A.E., Norred, W.P., "Effect of Dietary Lipid on Drug-Metabolizing Enzymes," Fed. Proc., 35:2475-2479, 1976.

Ward, J.M., Griesemer, R.A., Weisburger, E.K., Toxicol. Appl. Pharmacol., 51:389-397, 1979.

Weil, C.S., McCollister, D.D., "Relationship Between Short- and Long-Term Feeding Studies in Designing an Effective Toxicity Test," J. Agr. Food Chem. 11:486-491, 1963.

Weisburger, E.K., "History of the Bioassay Program of the National Cancer Institute." Prog. Exp. Tumor Res., 26:187-201, 1983.

Weisburger, J.H., Weisburger, E.K., "Tests for Chemical Carcinogens," In: Methods in Cancer Research, Ed. H. Busch, Academic Press, New York, 1967.

Williams, G.M., Furukawa, K., Maeura, Y., Croci, T., "Induction of Hepatic Enzymes by Corn Oil Gavage," Science, (in press), 1983.

Authors of Appendix I

Dr. Melvin E. Andersen
AFAMRL/THT
Wright-Patterson AFB, OH 45433

Dr. George D. Divincenzo
Health, Safety & Human Facility Lab
Eastman Kodak Company
Rochester, NY 14650

Dr. Perry J. Gehring
Vice President of Agricultural Products R & D and
Director of Health & Environmental Science
Dow Chemical Company U.S.A.
Midland, MI 48640

Dr. Robert R. Maronpot
NIEHS
P.O. Box 27709
Research Triangle Park, NC 27709

Dr. John C. Ramsey
Health & Environmental Sciences
Dow Chemical U.S.A.
1803 Building
Midland, MI 48640

Dr. Richard H. Reitz
Health & Environmental Sciences
Dow Chemical U.S.A.
1803 Building
Midland, MI 48640

APPENDIX I

Position Paper on Pharmacokinetics

1. Acquisition of the Minimum of Pharmacokinetic Data

No single procedure will be appropriate for all compounds because of the wide diversities in physical properties, routes of exposures, analytical sensitivities, and feasibility of radiochemical synthesis. In some cases it may not be feasible to gather pharmacokinetic data at all for technical reasons. Nevertheless, a general approach which will provide useful information is outlined below:

a. An analytical method for quantitating the parent material within a biological matrix should be developed. If sufficient sensitivity cannot be obtained by direct analysis, it may be desirable to use a radioactive form of the test material, coupled with one or more separation techniques to characterize the chemical nature of the radioactivity being measured (i.e., to see if it is still parent material).

b. The test material should be administered to animals of the appropriate species by the intravenous route if feasible. Initially, one sex should be studied. The time course of elimination of parent material from blood should be determined for a dose just beneath that needed to cause demonstrable toxicity and another which is the lowest feasible as determined by analytical sensitivity. It is desirable for the lowest dose for pharmacokinetic study to be in the range of anticipated human exposure for which risk assessment may be attempted. If comparison of the pharmacokinetic parameters for these two doses reveals nonlinearity, additional intermediate doses should be administered to

allow estimation of the magnitude of doses which will accomplish 80 to 95% saturation.

c. Animals should be administered the test compound by the proposed route of administration for the chronic study. Again a series of single doses should be employed, spanning the range from near lethality to the level of analytical sensitivity. Generally these studies may be done in one sex unless there is evidence from toxicity or other studies for or other reason to believe that significant sex differences are likely to occur. If appropriate, comparisons of different routes of administration should be conducted.

d. If radioactive forms of the chemical are available, the major routes of elimination, and the percentage of label eliminated by each route, may be determined for a series of doses in a single sex. Partial characterization of metabolites (e.g., urinary metabolites) might be useful. Elucidation of metabolic pathways is not the objective here; we are looking primarily for evidence of dose-dependency.

Once these data are obtained, the investigator can assess the rates of absorption and elimination of the test chemical. By plotting peak blood concentrations or, better, area under the blood concentration/time curves against administered dose the point at which non-linearity (if present) occurs can be estimated. Evaluation of elimination rates over a range of doses gives indications of saturable or capacity-limited elimination or metabolism of the chemical as well as an indication of potential accumulation with repeated exposure. Before proceeding with dose selection for the chronic study, some additional information is desirable:

a. Selected parameters should be monitored in the other sex of the species to be tested. It is not anticipated that every parameter studied above will be investigated in the other sex.

b. The effects of multiple dosing should be investigated by conducting limited pharmacokinetic studies in animals dosed for approximately two weeks. This time period is a compromise between the need to investigate the potential for changes in pharmacokinetic parameters in repetitively-dosed animals and practical considerations of the difficulty of obtaining such animals. It may be possible to coordinate subchronic toxicity evaluations with pharmacokinetic studies. This allows verification of any predictions about bioaccumulation, as well as providing time for induction/repression of enzyme systems and readjustment of metabolic pools (i.e., glutathione). Often the validity of using pharmacokinetic parameters determined from a single dose for chronically treated animals may be established by determining the steady state blood or tissue levels of the chemical in animals from the subchronic study. Using the pharmacokinetic parameters from the single dose study, the levels attained in tissues of animals used in the subchronic study should be reasonably predictable.

c. For some agents, differentials in pharmacokinetic parameters between trace and toxic doses may be caused by toxicity rather than saturable processes. Such information is important in dose selection; however, whether this differential is attributed to saturation or toxicity may be important in hypothesizing potential mechanisms of toxicity.

2. Acquisition of Additional Valuable Pharmacokinetic Data

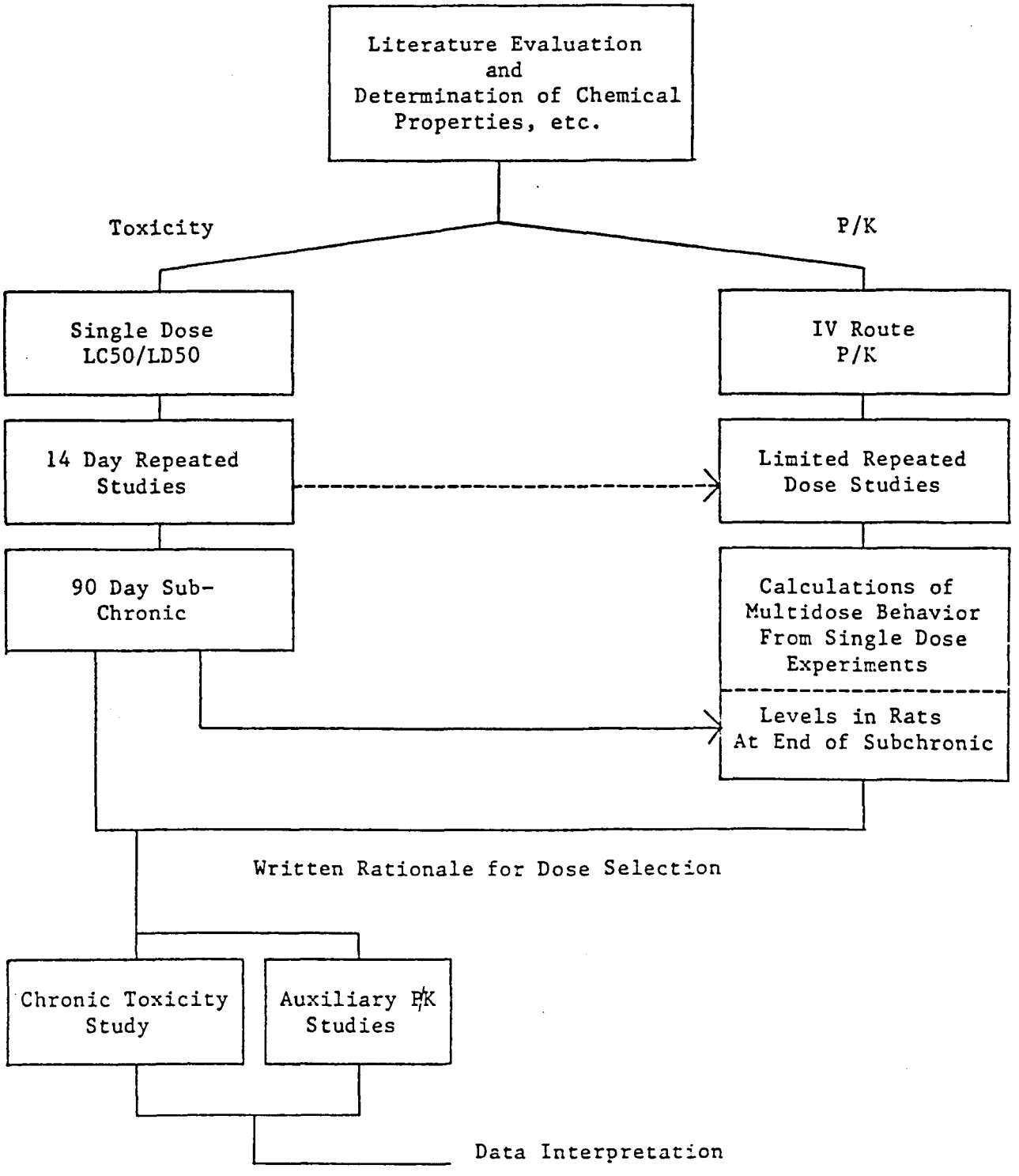
Although pharmacokinetic studies may be very sophisticated, the pharmacokinetic studies proposed in the foregoing for use in dose-selection are easily performed and relatively straightforward. They require only minimal analysis of the curves, mostly by visual inspection or multiexponential analysis with standard feathering techniques. They attempt in a general way to answer only two broad questions: Are there non-linearities in the pharmacokinetic behavior, and at what doses or concentrations do these non-linearities occur? With the variety of analytical procedures now available for determining chemicals in blood, the approach should be readily implemented for most test chemicals and interpretation of the pharmacokinetic studies would not require specialized pharmacokinetic expertise on the part of the chemical managers who will be responsible for defending the rationale for the dose levels selected.

Potential ambiguities may still arise in establishing the non-linear dose region when comparing an experimental route of administration (single dose gavage) with an intended chronic route (chemical mixed with diet). For example, non-linear absorption may be seen for a chemical given via gavage. Similar doses delivered in the feed may result in an entirely linear system since the localized concentration in the gastrointestinal tract is lower even though the daily dose is equivalent. Problems of this type can be evaluated by more detailed pharmacokinetic work which might be conducted during the 90-day subchronic study or concurrently with the chronic toxicity study.

The idealized attainment of pharmacokinetic and toxicity data shown in Figure A-1 will allow the former to be utilized in selection of doses to be used in the 90-day study. Having basic single dose pharmacokinetic data in hand and using standard equations, the achieved body burdens after multiple

FIGURE A-1

Dose Selection/Data Interpretation
Integration of Toxicity Data and Pharmacokinetics (P/K) Data



doses can be predicted. As indicated previously, reasonable predictability constitutes evidence that pharmacokinetic parameters do not change with repetitive exposure. Other pharmacokinetic studies may also be conducted during the two-year chronic toxicity test to address the question of whether age influences pharmacokinetic parameters. Other studies of value include (1) metabolite identification and evaluation of their pharmacokinetics, (2) analysis of glutathione depletion or covalent binding, and (3) determination of the effect of diet, disease, or altered physiology on basic pharmacokinetic parameters in test animals. These studies though valuable are incidental to the main objectives in the foregoing proposal. Such sophistication needs to be reserved for those materials warranting the best feasible estimation of risk; i.e., materials for which there is a small difference between human exposure and exposure which cause toxicity in animals.

Modeling efforts may be pursued by conventional compartmental analysis to understand the disposition of chemical in the test animal and the rate at which chemicals move between the various subcompartments. Physiological modeling with computer simulation is an area with exceptional promise for pharmacokinetic analysis (Bischoff et al., 1971). This approach may help explain the underlying basis of non-linearities and in some instances the biochemical and physiological factors involved in the non-linear processes. Ramsey and Anderson (1983) have described the non-linearity in metabolism of inhaled styrene in the rat in relation to the maximum rate of metabolism and organ blood flow with a physiological model. This approach showed that the inhaled concentration at which non-linear behavior would be observed should be the same from species to species. This extrapolation of physiological pharmacokinetic results from one mammalian species to another has been called animal scale-up (Dedrick, 1973). The styrene

model was also readily adapted to evaluate the enzyme induction seen after repeated exposures of styrene, a phenomenon of potential consequence for long-term studies.

It appears likely that predictive pharmacokinetic models can be developed based on careful measurement of physical and biochemical constants for the test compound. This approach has already been used with anesthetic gases and inhaled gases and vapors (Fiserova-Bergerova and Holaday, 1979). Because physiological factors which control pulmonary absorption of vapors are fairly well understood, a general description of lung function can be obtained which together with physical constants for the vapor becomes predictive of pharmacokinetic behavior by the pulmonary route of administration. The processes of xenobiotic absorption from the gastrointestinal tract and skin uptake across the skin are not as well understood so a priori predictive models are not yet available for these routes. One prospective use of a pharmacokinetic data base associated with a "national chronic toxicity" program would be to develop an improved understanding of the basic physiological processes which determine the kinetic parameters of absorption, distribution, and elimination. This would enhance the predictive powers of pharmacokinetic models in the future.

Implementation of predictive modeling would reduce the scope of experimentation necessary in pharmacokinetic data collection and reduce somewhat the numbers of animals required for toxicity testing. Limited, 'critical' results in rats, consistent with a general model could support simulation of expected behavior in other proposed test species, and in humans as well. Predicted behavior would be validated by appropriate but limited work in other species. The ability to predict human kinetics would be extremely useful for risk assessment since it would tell how administered dose and

internal target tissue concentrations are related in the species of interest - man. This would remove one of the uncertainties in species extrapolation - i.e., does the human handle the test chemical similarly to the test species? The technology to drive these developments now exists but a commitment by the NTP program to incorporate pharmacokinetic considerations would accelerate progress in this area.

3. Choosing the Doses

No matter what objectives are selected for a chronic toxicity evaluation, it is essential to document the dose selection process in detail. This documentation needs to address: What toxic effects were considered in establishing the highest dose? Why were the lower doses selected? Should three or four doses be given in addition to controls? Are there non-linearities in pharmacokinetics and how did those relate to dose selection? How do the doses selected relate to anticipated human exposure? What results from the study may be anticipated to relate to human risk assessment? Are there any which may not?

To some degree two objectives of a chronic toxicity study may be in conflict - evaluation of data for human risk assessment versus evolution of data for elucidation of toxicity. Albeit infrequent, this conflict occurs when there is a large differential between doses needed to elicit toxicity and potential human exposure excepting abusive use. The first approach for dose selection set forth below assumes that the overwhelming objective is that of obtaining data for human risk assessment.

If human exposures are in the range where linear pharmacokinetics predominate, then the chronic toxicity study must contain at least two dose levels where linear pharmacokinetics predominate if extrapolation is to be possible; however, utilization of less than three doses is to be avoided

except in rare, well documented instances. Conversely, if human exposures are near or above the area of non-linear pharmacokinetics, then the chronic toxicity evaluation should place emphasis upon defining the dose response curve in this area. Dose selection thus is influenced by three variables: (1) pathological observations in subchronic studies, (2) pharmacokinetic studies, and (3) expected human exposure levels.

This approach may be illustrated by reference to Table A-1. A test material is placed into one of the categories listed in the table in accordance with these three criteria and dose levels are selected based on principles appropriate for that category. This approach is a model for how a dose setting regime may be employed.

In Table A-1, "X" stands for the highest allowable dose on pathology criteria alone, "S" stands for the approximate dose where non-linear pharmacokinetic behavior becomes apparent, and "H" stands for the estimated human dose.

When pharmacokinetic studies have not revealed any non-linear behavior (Category A), or have shown such behavior only at doses well above the maximum judged allowable on pathological criteria (Category B), then pharmacokinetics will play no further role in the selection of doses. Three doses would normally be employed in such chronic toxicity evaluations. Geometric spacing will be employed for selecting these doses; i.e., the doses will be X, X/Y, and $X/(Y)^2$. A good rule of thumb in such cases would be to set Y equal to the square root of ten so as to span one order of magnitude in the selected doses.

Depending upon the pathological endpoint used in setting doses, it may be desirable to have an additional dose which is likely to cause a lesser effect than the selected maximum dose, while still retaining two doses

TABLE A-1

Dose Selection Parameters

Category	Pathology Top Dose	Non-linear Pharmacokinetics	Human Pharmacokinetics	Human Exposure (Estimated)
A1	X	No	- -	H>.1X
A2	X	No	- -	.1X>H>.001X
A3	X	No	- -	.001X>H
B1	X	S > X	- -	H>.1X
B2	X	S > X	- -	.1X>H>.001X
B3	X	S > X	- -	.001X>H
C1	X	S = X	- -	H>.1X
C2.1	X	S = X	No	.1X>H>.001X
C2.2	X	S = X	Yes	.1X>H>.001X
C3	X	S = X	- -	.001X>H
D1	X	S < X	- -	H>.1X
D2.1	X	S < X	No	.1X>H>.001X
D2.2	X	S < X	Yes	.1X>H>.001X
D3	X	S < X	- -	.001X>H

X = Highest allowable dose on pathology criteria.

S = Approximate dose where non-linear pharmacokinetic behavior is apparent.

H = Estimate of dose received by humans.

anticipated to have no effect. For example, if 3X causes 20-30% increase in liver weight, and X causes approximately 10% increase in liver weight, then four doses might be employed: 3X, X, X/Y and X/Y^2.

The value of Y will be influenced to some degree by the steepness of the pathology dose response curve and the results of *in vitro* tests of genotoxicity. A shallow pathology dose response curve would suggest a larger value of Y than a steep dose response curve, (increase dose spacing) while steep dose response curves would have the opposite effect. Similarly, indication of potential genotoxicity in short term tests would reduce the value of Y and increase the number of doses selected for the chronic toxicity test.

In Category C, pharmacokinetic non-linearities occur in the same region of doses as those identified by pathological criteria. In such a case, categories C1 and C2.1 would be treated differently than C3 and C2.2.

Materials in category C1 have an expected human exposure close to that identified by pathological criteria (greater than or equal to 0.1X). Materials in category C2.1 have an estimated human exposure between 0.1X and 0.001X and lack human pharmacokinetic data to indicate that the test animals handle the material like man. In these cases, the top dose should be selected so as to produce an estimated 80 to 95% saturation, assuming the pathology associated with such a dose is not likely to be life shortening. Two other doses will be selected as follows: (1) Middle dose to produce an estimated 15 to 30% saturation as determined from the area under the blood or plasma concentration versus time curve (often an 0.3 log unit of the MTD will be suitable), (2) Lowest dose at least a half log unit below the middle dose. If it is anticipated that an even lower dose will be required to attain a no effect level, additional logarithmically spaced doses should be included.

Materials in category C2.2 are those which have an expected human exposure level of between 0.1X and 0.001X with supporting human pharmacokinetic data to suggest that the test animals handle the material like man and materials in category C3 are those where human exposures are less than 0.001X. For these the highest dose is chosen to produce 15 to 30% saturation, with at least two geometrically spaced lower doses.

Materials are placed in category D when non-linear pharmacokinetics can be demonstrated at doses well below those identified by pathological criteria. In such a case, four dose levels are selected for categories D1 and D2.1. The highest dose level employed is X, and the next three dose levels are chosen so as to provide: (1) One dose level causing an estimated 80 to 95% of saturation; (2) One dose level causing an estimated 15 to 30%

saturation; and (3) One dose at least one half log unit below that estimated to be 15 to 30% of the saturation level. This regimen will provide data which may be meaningfully extrapolated if chronic toxicity is incurred.

Materials in category D3 would be tested with three doses. The top dose would be selected so as to give 70-80% saturation with two additional doses selected as for C3.

Materials in category D2.2 may be tested as designated for materials in category D3 if the human pharmacokinetic data is strong and consistent with the absence of saturation at anticipated human exposures and exposure is near the 0.001X value rather than 0.1X. Otherwise chronic toxicity evaluations will normally require four doses as outlined for materials in category D1 and D2.1.

If the primary objective of a chronic toxicity evaluation is to elucidate toxicity regardless of its meaningfulness for assessment of human risk, the foregoing approach remains applicable except for category D. In this case, pathological criteria will be used solely for selection of the top dose, regardless of its relationship to known, allowed, or anticipated human exposure. Materials falling into category D will be evaluated with a minimum of four doses; the top dose based on pathological criteria, a dose equal to or greater than that needed to cause 80 to 95% saturation, a dose estimated to give 15 to 30% saturation and a dose at least one half log unit below the latter.

The foregoing guidelines for dose selection are not intended to address all of the possible combinations of pathological and pharmacokinetic criteria which may be useful in dose selection. Only with sound, considered, scientific judgment may doses be selected for chronic toxicity evaluations

which will ultimately produce a data base which can be meaningfully extrapolated for human risk assessment. Historically, such judgment has all too often been attempted only after the chronic toxicity study is completed and at this point even sound judgment is too often precluded by results which have a poor relationship to human risk assessment. The most important scientific task in conducting chronic studies is that of designing the study and clearly documenting the rationale used in its design.

Report of the Subpanel on Design of Chronic Studies

A. Introduction

The Ad Hoc Panel on Chemical Carcinogenesis Testing and Evaluation quickly realized that its task could be most easily addressed by separating the effort into stages. The Subpanel on Design of Chronic Studies concentrated on that portion of the Panel's charge dealing with "methods that the NTP should use for the detection and evaluation of chemical carcinogens." Through discussion with the full Panel, staff of NTP, communication with other scientists in the field and public input, a list of 40 topics relevant to the Panel charge was compiled. From that list, 12 items judged to be of major importance were selected for incorporation into this report. These items form the main chapter headings of this section and are generally presented in the format of Recommendations, Background, Review of the Literature and Listing of Options.

In preparing this section of the report, the Subpanel had available to it briefing papers or publications prepared by individual members, by NTP staff, by staff of other agencies or from outside sources. The Subpanel wishes to acknowledge at the outset the valuable contributions it has received from interested colleagues.

B. Scientific Basis of Carcinogenicity Testing

Clinical observations, epidemiological investigations and experimental studies have led to the identification of some 40 chemicals, groups of chemicals, industrial processes or risk factors, which are known to be or are strongly suspected of being associated with cancer induction in man (IARC Monograph, Supplement 9, 1982; Doll and Peto, 1981). For 32 chemicals or groups of chemicals, with sufficient data for a valid comparison, a good

qualitative correlation is observed between carcinogenic effects in man and carcinogenic effects in experimental animals (see Wilbourn et al., 1984; Rudnay and Borzsonyi, 1981; Pershagen et al., 1984). In addition for some chemicals (4-aminobiphenyl, aflatoxin B_1 diethystilboestrol, mustard gas and vinyl chloride) the experimental evidence of carcinogenicity preceded the epidemiological evidence. This indicates that findings in long-term carcinogenicity studies in animals may be of value in predicting a similar effect in man and that they should be taken into account in the implementation of public health measures aiming at primary prevention of cancer.

This a posteriori analysis does not necessarily imply that all chemicals found to be carcinogenic in experimental animals will eventually be found to be carcinogenic in man. Epidemiological studies have strong limitations in sensitivity mainly due to sample size, quantification of exposure or confounding factors. Experimental studies that may have been carried out under such extreme conditions make difficult any useful extrapolation of the findings to the human situation. However, positive results from well conducted carcinogenicity studies in animals must be taken as a strong evidence that the compound considered represents a carcinogenic hazard to man.

It should be stressed that these extrapolations from species to species and from rodents to man are generally carried out for public health purposes and are often based on best estimates at that time. They do not necessarily imply extrapolation on a scientific basis in all respects for each species. Such estimates are often required to meet societal needs and are subject to change when further scientific information becomes available.

During recent years new knowledge has been acquired on the mechanism(s) of carcinogenesis, which, although still limited, may permit a more scientific extrapolation of carcinogenicity data from species to species and an improvement in the planning, execution and analysis of the long-term carcinogenicty tests *in vivo* and of short-term tests having endpoints other than cancer.

The term "chemical carcinogenesis" has been defined (IARC Monograph Vol.31, 1983) as "the induction by chemicals of neoplasms that are not usually observed, the earlier induction by chemicals of neoplasms that are commonly observed, and/or the induction by chemicals of more neoplasms than are usually found." There is substantial evidence from epidemiological and experimental studies that carcinogenesis is a multistage process and that each of these stages can be affected by different agents through various mechanisms. These stages have been described as *initiation, promotion and progression* and the natural history of this process has been examined mainly in the mouse skin and rat liver. (Yamasaki and Weinstein, 1984; Pitot, 1984; Armitage and Doll, 1961; Peto, 1977; Foulds, 1969, 1975; Farber, 1982; Weinstein, *et al.*, 1984.)

Substantial evidence exists that the process of carcinogenesis, and in particular *initiation* results from mutational changes occurring in somatic cells. Mutation is used in the broad sense as a change in DNA or chromosome structure that is heritable at the somatic level. It may include specific changes in DNA sequence, i.e., point mutations (base substitution, frame shift), deletions, insertions, gene rearrangement and gene amplification as well as gross chromosomal changes, i.e., translocations, aneuploidy, etc. This is substantiated by the following observations: a) for some tumors monoclonal origin has been demonstrated, b) the ultimate reactive forms of many chemical carcinogens react with cellular DNA and are

mutagenic, c) DNA repair processes can modulate the carcinogenic response, d) many types of experimental and human cancers have chromosomal abnormalities, and e) neoplastic transformation can be associated with activation of proto-oncogenes through a mutational event, (e.g., point mutation, chromosome translocation).

The stage of promotion describes the process by which various agents ("promoters") are capable of increasing the probability of neoplastic development initiated by a previous exposure of the cells to a carcinogen. This process has been extensively studied in mouse skin (Berenblum and Shubik, 1947; Hecker et al., 1982) and it has now been observed in other biological systems (Pitot, 1984).

The term "promoters" includes agents belonging to different chemical classes or types of exposure, and acting probably through different mechanisms that result in an increased incidence of tumors and that do not necessarily comply with the original requirements described by Berenblum, 1974. An increased incidence of tumors seen after sequential administration of various chemical agents could also result from an additive or a multiplicative effect of those agents. Various examples of this phenomenon have been reported in experimental and epidemiological studies and are referred to as co-carcinogenic effects. The classification of carcinogens as 'genotoxic,' 'initiating' or 'epigenetic' or 'promoter' (see below) does not seem justified in an absolute sense, since a carcinogen could act as an initiating agent in one tissue and as a promoter in another. For example, urethane, which is carcinogenic for lung but not for the skin of mice, has only initiating activity in skin (Haran and Berenblum, 1956); 2-acetylaminofluorene, a carcinogen for mouse and rat liver, seems to behave as a promoter in mouse bladder (Day and Brown, 1980).

Epidemiologists (Peto, 1977; Day, 1984) have proposed that carcinogenic agents could be grouped according to their capacity to affect mainly the 'early' or the 'late' stages of carcinogenesis; in this respect, it should be noted that a carcinogen like asbestos seems to affect mainly the late stages of the induction of lung cancer and mainly the early stages in the case of mesothelioma (Peto et al., 1982). A similar situation exists for carcinogenic hormones; thus, diethylstilboestrol acts as an early-stage carcinogen in the transplacental induction of vaginal carcinomas, while oestrogens probably act on the late-stage of endometrial cancer development (Armstrong, 1982).

Although it is true that many carcinogens induce point mutations (see Hollstein et al., 1979), there is also good evidence that carcinogens, including promoting agents, induce more complex genetic changes at the level of the gene or chromosome. These include chromosomal aberrations, aneuploidy, gene amplification, gene rearrangement (translocation, transposition), altered DNA methylation, gene activation, etc. Activation of specific sequences or genes called oncogenes has been directly associated with neoplastic transformation or tumorigenesis. The mechanism of activation may involve point mutation or other genetic change of proto-oncogenes that are normal cellular sequences. As yet there are no methods available to systematically examine potential chemical-oncogene interactions, but such studies will undoubtedly have an impact on understanding the mechanism of chemical induced tumors and in identifying carcinogens. Newbold and Overell (1983) showed that EJC-Ha-ras-1 can transform primary diploid hamster fibroblasts only if those cells have been immortalized previously by treatment with chemical carcinogens. These studies represent important progress in the understanding at the cellular and molecular levels of the various stages of carcinogenesis.

Various mechanisms have been proposed for the activation of these proto-oncogenes and for the tissue specificity of the process which suggests that it results from a genetic alteration (see Weinberg, 1982, 1983; Bishop, 1983). In the near future it will be possible to evaluate the contribution of various chemical carcinogens (initiating and promoting agents) to this process and to determine whether this effect is mediated through direct DNA damage or through alteration of gene function via an epigenetic mechanism (e.g., hypomethylation). It is interesting that various carcinogens, (including some for which there is no evidence of covalent interaction with DNA) induce hypomethylation in DNA (Boehm and Drohovsky, 1983; Wilson and Jones, 1983). A decreased level of DNA methylation at the 5-position of cytosine has been shown to result in the induction of various gene functions (Doerfleer, 1983; Riggs and Jones, 1983); and, recently, it was reported (Feinberg and Vogelstein, 1983) that in some primary human tumours the c-Ha-ras and c-Ki-ras genes were hypomethylated compared to the normal tissues adjacent to the tumour. It might be hypothesized that activation of proto-oncogenes can be induced by chemical carcinogens not only by direct DNA damage but also as the result of stable DNA hypomethylation. The 5-methylcytidine analogue, 5-azacytidine, which is incorporated into DNA but cannot be methylated, activates various genes (see Doerfler, 1983 for review) and Carr et al., (1984) report that 5-azacytidine is carcinogenic to rats, inducing tumours in various tissues after chronic administration and increasing the incidence of liver tumours when given after a single dose of N-nitrosodiethylamine.

It has been proposed (Williams, 1980, 1983; Marshall, 1982) that carcinogens be classified into different categories according to their capacity to modify DNA chemically or to produce genetic damage in various short-term tests, e.g., mutagenicity in bacteria or mammalian cells, and

that the evaluation of risk for agents that lack this capacity (so called 'epigenetic' carcinogens, including promoters) should be less stringent and that this be taken into account in establishing regulatory measures and threshold values (Squire, 1981; Weisburger and Williams, 1981). Recent progress in the understanding at the cellular and molecular level of the multistage carcinogenesis would, however, suggest that such classification of carcinogens into different categories with the almost automatic implication of different degrees of risk for humans is still premature. This problem was recently discussed (IARC, 1983) and, although it was realized that a classification of carcinogens according to their mechanisms of action could make an important contribution to the prevention of human cancer, strong reservations were expressed about the feasibility of such a classification (see also Weinstein, 1983) since present understanding of the mechanisms of carcinogenesis would not justify it.

Although until recently the induction of genetic damage has been implicated in initiation but not promotion, there is some evidence that tumour promoters, like TPA, can produce DNA damage through the formation of free radicals and by other mechanisms (Kinsella and Radman, 1978; Varshovsky, 1981; Cerutti et al., 1983; see also Copeland, 1983) and that some effects of TPA are irreversible (FUrstenberger et al., 1983). As our knowledge of carcinogenesis advances, it is likely that we will find fewer arguments to support the sharp distinction that is now made between initiating and promoting agents (see Montesano and Slaga, 1984).

C. Design Considerations - General Recommendations

1. For routine bioassays use a design employing three test doses plus control; distribute the animals equally among groups and continue to

use 50 animals of each sex and species in each dosage group, including controls.

2. For special studies or special needs the numbers of groups and number and distribution of animals within groups may be altered.

Background and Review

The use of experimental animals (generally rodents) in toxicity experiments to predict possible human health risks and serve as a guide for setting permissable exposure levels for man has been in practice for many years. However, the responses being studied have changed. Whereas originally the studies were used to assess acute toxicity, the need to evaluate possible health effects related to chronic diseases requires that the bioassays also provide data relevant to long-term effects and relative potencies. Ideally, such data would also be sufficiently quantitative that they will be useful in estimating safe exposure levels.

With these changes in purpose have come many changes in the design of the chronic bioassay and the methods used to analyze its results. Extensive research on animal husbandry and pathology led to specialized strains of animals, well controlled exposure conditions and standardized methods of tissue preparation and analysis. However, very little research has been done on bioassay design parameters related to the analysis of the response data, especially with respect to the number of doses to employ, the dosing intervals, the minimum number of animals and how to allocate these animals into the dose groups.

One of the more common cancer bioassay designs is that currently used by the National Toxicology Program (NTP). This NTP bioassay is essentially the same as that developed by the National Cancer Institute's (NCI) carcinogen testing program. Most of the procedures used were developed at NCI

in the early 1970's, particularly the design parameters to be investigated (Page, 1977).

The current NTP cancer bioassay is typically conducted on four groups of animals (male and female rats, male and female mice) and is designed to cover the greater part of the animal's life (NCI, 1976). Each experimental group is composed of animals receiving identical treatment. Typically, each experimental group contains 50 animals. Control groups are used to match each experimental group and are generally untreated or vehicle treated controls. Two dose levels of the test agent are regularly used. The highest dose is the maximum tolerated dose (MTD) as predicted from subchronic data and the remaining dose is generally taken as one-half of the MTD. Thus, the usual NTP bioassay utilizes 150 animals of each sex of two species; 50 animals per experimental group with a control group and two experimental groups that receive doses of MTD/2 and MTD.

In the current NTP bioassay there is an interim and terminal sacrifice. The animals either die naturally during the course of the study or are sacrificed at the predetermined termination dates. The animals are necropsied at death and are subjected to gross and microscopic examination by a pathologist to determine systemic changes. This type of bioassay provides data on the incidence of various neoplastic and non-neoplastic lesions, as well as survival information.

Animal bioassays have two goals: 1) to determine if exposure to the agent under study increases the incidence of some response variable, typically, the age-specific incidence of specific types of cancer; and 2) to provide some information relevant to the potency of the test compound that will be useful for the assessment of potential human risks. The objective of this section is to suggest experimental designs which provide reasonable results with respect to both of these two goals.

Consideration of Options

The optimal design of an experiment depends critically on the goal of the investigation, which for the NTP bioassay includes both screening of separate compounds for carcinogenic activity and more detailed study of the shape of the dose-response curves in an attempt to elucidate carcinogenic mechanisms. If the goal is simply to demonstrate a difference in tumor rates between treated and control animals, greatest statistical power is achieved by dividing the available animals equally between the untreated control group and a single treatment group. The dose for the treated animals is placed at the upper ranges of the dose-response curve so as to maximize the difference in the observed rates. This rationale underlies the current NTP choice of the MTD as the high dose. The lower dose group at 1/2 MTD is included as a type of insurance against the MTD being overestimated and toxicity being so severe that many animals at the high dose die or are otherwise unevaluable for the carcinogenic response. Only limited information about the shape of the dose-response curve is available from these two dose levels.

The scientific judgment that there is a true causal association between the carcinogenic response and the administration of the test compound is strengthened by an observation of increasing tumor production at each of three or four increasing dose levels. A cogent argument for dividing animals among additional dose levels can be made on this basis, even though it may mean a reduction in the power of the formal statistical analysis. Assuming that the response increases roughly linearly over a specified dose range (which may often be achieved by a suitable choice of the dose metameter), the best visual effect is produced by equal spacing of the doses over this range and by use of a sufficient number of animals at each dose level so that the linear increase is not obscured by statistical fluctuations.

Precise elaboration of the shape of the dose-response curve is generally not possible with an experiment of the magnitude of the current NTP bioassay. One might ask, for example, whether the assay could be used to determine if the dose response curve was non-linear or more specifically if there was a sharp break in the slope suggestive of a "hockey-stick" behavior. Such a determination could be made for an agent that dramatically increased the occurrence of rare tumor types. An observed result of 0/50, 0/50, 20/50 and 30/50 tumors at the control, 1/4, 1/2, and full MTD dose levels would indicate a change in slope between the second and third groups. However, because of the increased variability at the lower dose levels, no such inference could be made for tumors with a high spontaneous incidence and where the observed result was 10/50, 10/50, 20/50, and 30/50. It is well known that different parametric statistical models, all based on presumed carcinogenic mechanisms, can be more or less equally consistent with the experimental data in the observable range and yet differ markedly upon extrapolation to low doses (Munro and Krewski, 1981; Portier and Hoel, 1983a).

Different theoretical models and analytical goals can lead to quite different recommendations regarding the relative numbers of animals allocated to the various treatment groups. One model for low dose extrapolation using a four point assay with equally spaced doses suggests that the allocation be made in the ratio 1:2:2:1 (Krewski, Kovar, and Bickis, 1982). Another model leads to the recommendation that the dose levels be in the approximate ratio 0:1/4:1/2:1 and the sample sizes in the approximate ratio 1:1:2:1 or even 1:1/3:2/3:1 (Portier and Hoel, 1983b). More general considerations suggest that, if additional lower doses are to be used in the bioassay, then the number of animals at such doses (and presumably also in the control group) should be increased relative to the number at higher doses so

that the power of the design to detect a significant effect at each dose level is maintained (Statement from Consumer Product Safety Commission). In view of these competing interests, and the fact that balanced and moderately imbalanced designs have roughly comparable power using the linear trend test (Portier and Hoel, 1984), the recommendation made here is that NTP continue to use equal allocation for its routine, general purpose bioassays. Unequal allocation may be preferable for some specific circumstances.

In summary, the stated goal of using the same bioassay both to screen for carcinogenic activity and to explore the dose-response relationship for a known carcinogen leads to somewhat conflicting criteria for optimizing the experimental design. The Panel recommendation that one additional low dose group of 50 animals be added to the current routine NTP design is a compromise solution that provides for slightly increased statistical power and some further information about the dose-response curve, especially if the highest dose group is lost to analysis. An alternative strategy would be to conduct the bioassay in two stages: screening followed by a later exploration of dose-response for chemicals identified as having carcinogenic activity. This would have the advantage that the design of the dose-response study could utilize the information on target organ(s) and tumor rates observed at the screening stage. If only a small fraction of the agents put on test screened positive, moreover, this two stage design would be more cost effective since it would restict the exploration of the dose-response function to chemicals that demonstrated carcinogenic activity in a comparison of control and high doses. However, this two-stage design was ultimately rejected by the Panel, largely on the grounds that a delay of several years between a positive finding of carcinogenicity and the subsequent elaboration of dose-effect relationships could have intolerable consequences for the regulatory process.

D. Selection of Species and Dose

Recommendations

1. The MTD as currently defined in this report should be employed in whole animal bioassay for carcinogenic agents as the highest level administered.
2. NTP should continue to use metabolic and pharmacokinetic studies in selecting doses below the MTD.
3. A continued search for species other than the rat, mouse, and hamster which satisfy reasonable criteria for chronic bioassay studies should be encouraged.
4. Although strongly recommended, chronic bioassay procedures by NTP need not be restricted to inbred or hybrid lines of rats and mice.
5. In order to understand a possible factor in the background incidence of neoplasia, serious consideration should be given to a study by NTP of the effect of restricted dietary feeding on the spontaneous tumor incidence.

Background and Review of the Literature

Despite a variety of systems for assessing the carcinogenicity of chemical compounds that were developed between 1956 and 1964 (cf. Weisburger and Weisburger, 1967), it was not until the Weisburgers (1967) proposed a systematic method for assessing carcinogenicity of chemicals in rodents that a more generally accepted procedure came into being. The testing methods presently in use by various branches of the federal Government, as well as the NTP, have as their basis proposals by these investigators and others.

In the 1960s the predominant species used for the assay of carcinogenicity were mice, rats, and dogs. The hamster was just coming into use in

bioassays at that time. Obviously rodents are relatively inexpensive and have a short life expectancy, whereas the dog represents a longer lived species which is more expensive to maintain. For specific reasons other more exotic species had also been employed but were not in general use (cf. Weisburger and Weisburger, 1967).

Although the use of the maximum tolerated dose (MTD) of the text chemical may not have been initiated by the Weisburgers, they emphasized the importance of the determination of such a factor. As a first approximation they suggested that the LD_{10} might be determined using classical pharmacologic techniques. After several trials with relatively small groups of animals the high dose (MTD) was to be selected such that no more than 10-20% of the animals succumbed to long-term toxicity. At that time it was also suggested that no more than 10% of the weight of the diet should be the test compound in the case of relatively non-toxic compounds. Under most circumstances it was suggested that a second (and lower) dose level allowing extended survival of all of the animals also be used. These workers suggested that this dose should be 50-70% of the highest dose. If only a single dose were used in the test, then the lower dose should be chosen. When more than two dose levels were employed it was suggested to use the ratios 1000 (top dose), 500, 100, 50, 10, 5, and 1. This dosage schedule would only be employed where highly accurate evaluations of dose-responses were sought. This concept is discussed in greater detail in Chapter D and Appendix I of the Report of the Subpanel on Subchronic Studies and Related Issues.

From these historical developments certain conventions have arisen. Paramount among these are the use of rats and/or mice as the principal species in carcinogenicity bioassays in whole animals. Virtually all testing is carried out in inbred or hybrid strains of these animals and the use of

the MTD, although with some modifications indicated below, still remains an apparent necessity for "valid" testing of chemicals for their carcinogenic potential in mammals.

Consideration of Options

Dosage Considerations

As indicated above the maximum tolerated dose (MTD) is still utilized in one form or another as the high dose of test agents in any rodent bioassay. In reality this dose is determined by prechronic studies which aid in the identification of a dose level which, when given for the duration of the chronic study as the highest dose, will not impair the normal longevity of the treated animals from effects other than the induction of neoplasms (Schwetz, 1983). Such doses should not cause morphologic evidence of toxicity in organs other than mild changes such as slight hypertrophy or hyperplasia, inflammation or slight changes in serum enzymes. Still such dose levels have distinct disadvantages as have been shown by studies such as those of Watanabe and Gehring (1976) in the case of vinyl chloride. In this case the pharmacology, pharmacokinetics and metabolism of the compound indicate a distinct difference in the organism's treatment of the material at low as compared with high doses. In view of this it would seem reasonable to suggest that results obtained with the MTD may be more interpretable when appropriate metabolic and pharmacokinetic data are also available. Under other circumstances the lower dose (1/2 to 3/4 MTD) would be more reasonable and such data may not impair the sensitivity of the assay. Furthermore more than one assay has had to be eliminated or redone because of unsuspected toxicity occurring well into the assay, beyond the periods of the prechronic testing.

While extensive considerations have been given to the MTD, relatively little consideration has been given to the dosage schedule, and dose rate in bioassay protocols. Clearly changes in these parameters markedly alter the tumor incidence induced by complete carcinogens such as those of the bladder (McCormick et al., 1981) and for promoting agents such as croton oil and phorbol esters (Boutwell, 1964) and phenobarbital (Goldsworthy et al., 1983). Furthermore in some carcinogenesis studies, the appearance of specific histologic types of neoplasms may depend on the dose rate, while in the same experiment the appearance of other histologic types may show no relationship to dose rate (Peto et al., 1982).

Finally one of the critical factors in dosage consideration is the question of a "no-effect" level or a threshold dose below which no effect is elicited or the more nebulous concept of "virtual safe dose" (VSD). A principal assumption upon which the extrapolation of animal test data to the human situation is based is that carcinogenic agents display no measurable threshold level. This assumption is based largely on extrapolation usually from the 5-10% tumor level since most published bioassays for carcinogenicity do not allow statistical assurance of significant carcinogenicity beyond this level (Shubik and Clayson, 1976). Moreover the alternatives to such extrapolations are, in general, so expensive and time consuming as not to be realistic. Despite the fact that some rather potent carcinogens have been reported to exhibit experimental threshold levels (Arai et al., 1979; Scherer and Emmelot, 1975), other studies either do not support such findings (Peto et al., 1982) or the findings are dismissed on statistical grounds.

The use of mathematical models has become a major factor in predicting the relationship of dosage to tumor response. Some models, such as those

based on the one-hit theory prevalent in radiation carcinogenesis, or the multi-stage model used frequently by regulatory agencies, extrapolate the response to 0 with no observable threshold. On the other hand other models, such as the Weibull model, may take into account thresholds (Carlborg, 1981).

Species Considerations

Sontag (1977) has proposed certain criteria for species selection. On the basis of these criteria he has suggested that at the present time the only species that meet such criteria for large scale, long term carcinogen bioassays are the rat, mouse, and hamster. On the other hand, as pointed out earlier by Weisburger and Weisburger (1967), when specific questions are asked of the bioassay rather than the simple endpoint of the production of an increased number of benign and/or malignant neoplasms, a variety of species can be used. For example, the specific relationship of a highly defined genetic background and the induction of neoplasms in animals harboring few if any known oncogenic viruses may lead to the use of fish (Anders et al., 1981) or insects (Gateff, 1978). Some chemicals may not be carcinogenic in the three species emphasized by Sontag, thus requiring that other species be considered (cf. Purchase, 1980).

Finally one of the more debatable characteristics of the animals utilized in carcinogenicity bioassays is the genetic purity of the strain of animals. As Sontag points out, at the present time inbred, genetically homogeneous strains of rats and mice are preferred because of their predictability, their uniform biological responsiveness, the greater opportunity for comparison of studies between laboratories, and their possible greater desirability for dose response and other studies. Furthermore the deterioration of inbred strains due to genetic drift and other factors may in the future be detrimental to such experiments, while the more uniform development of outbred or partially inbred strains of animals may suggest

their use in place of, or in addition to, the inbred strains currently employed for most whole animal bioassay experiments. Further discussion of this topic is found in Section C of the Report of the Subpanel on Subchronic Studies and Related Issues.

E. Selection of Route

Recommendations

1. Bioassays, in general, should employ a route or routes relevant to anticipated human exposure unless not possible for technical reasons.
2. Bioassays which, for technical considerations, use gavage or other atypical routes should be backed up by pharmacokinetic studies which measure peak blood level, "area under the curve," or other relevant measures of dose for the typical and other routes.

Background and Review of the Literature

A number of recent monographs or reports on the conduct of longterm bioassays have dealt with the issue of selection of route of administration. Consistently, the authors have noted that the route chosen in the experimental studies should be realistic, relevant to anticipated human exposure and take into account the known physical and chemical properties of the test agent (Baron, 1980; FDA Advisory Committee on Protocols for Safety Evaluation, 1971; IARC, 1980; NCI, 1976; OECD, 1981; Roe, 1981; Roe, 1978; TOSCA, 1979; USFDA, 1982; USEPA, 1982; WHO Scientific Group, 1969). None of these reports, however, have dealt in any detail with certain species considerations or the pharmacokinetic relationship of route to dose. Yet the successful conduct of a bioassay should consider not only the relevance of the dose route to anticipated human exposure but also the relevance of that route to the test species. For example, technical considerations (volatility, stability, palatability) may favor oral gavage for a material which is normally a

dietary constituent for humans. Gavage produces a bolus of material and often, but not always, the expected consequence is a rapid rise and fall in blood levels. Rodents with full-time access to feed will spread their feeding activities out over a period of hours. A dietary level equivalent on a mg/kg body weight/day basis may, therefore, produce a much different absorption, blood level and elimination profile.

Unpublished studies of reproductive effects of color additives have shown that dietary admixture failed to produce effects with FD&C Red 2 comparable to those seen when the dose was given by gavage. One of the important roles for prechronic mechanistic studies could be the establishment of equivalence values for various dosing regimens.

Further discussion of this subject is found in Chapter E of the Report of the Subpanel on Subchronic Studies and Related Issues. In particular, mention is made of skin and inhalation routes.

Consideration of Options

The oral route is often chosen because it is easier, more convenient or cheaper than one which is comparable to the human exposure situation. At times the oral route may be justified even if human exposure occurs by another route if there is reason to believe that both routes produce systemic exposure and grossly similar kinetics, and that it would be difficult to mimic the human route in the animal bioassay. There are instances where the exposure to test material may occur in humans via multiple routes. Selection of the proper routes may, therefore, require considerable care.

Consider, for example, an organic solvent used in various industrial applications such as a thinner for coatings, a degreasing agent and a component of adhesives used in food packaging. The occupational exposures could be at the level of 50-100 ppm in air (175-350 mg/m^3). Assuming eight-hour exposure, 5 cu m breathed in eight hours and 50% uptake, the dose to a 70 kg worker is 435-870

mg or about 6-12 mg/kg/day. A residue of 1 ppm in the total diet of the same worker from the food contact use of the solvent, a most unlikely prospect, would yield a dose (based on 1500 g/d dry weight of food eaten) of 1.5 mg or about 20 ug/kg/day. If the agent is administered in a bioassay orally at the MTD, important dose and route considerations may have been overlooked. The bioassay route represents <1% of the anticipated human exposure route. The kinetics of uptake, distribution, conversion and elimination by the oral route at the MTD may in some cases produce a statistically valid but biologically misleading outcome. The temptation is to conduct two studies--one by inhalation at MTD and the second orally at a significant exaggeration of the anticipated intake (but not necessarily the MTD).

This same combination of anticipated human exposures and choice of animal dosing procedures is possible for many compounds likely to be included in the bioassay. There should be careful review of all the options and an awareness of the impact on outcome before the final selection is made.

F. Selection of Vehicle

Recommendations

1. A number of nominally acceptable but unvalidated alternatives to vegetable oil gavage exist. These include the use of aqueous suspending agents, microencapsulation, and highly polar nonvegetable oils. Those should be considered, but since they have not been extensively used and validated, further study is needed.
2. The nature and extent of potential confounding effects from vegetable oil gavage, including altered susceptibility to the test agent, should be evaluated.

Background and Review of the Literature

Edible vegetable oils are frequently used as vehicles or solvents to administer water-insoluble compounds to experimental animals. This technique is also employed for gavage of test agents which are volatile, irritating, of

objectionable taste or odor or unstable and which can be dissolved or dispersed in oil.

A recent report by a work group convened by the Nutrition Foundation, addressed some of the issues associated with vegetable oil vehicles, especially in connection with gavage administration of test agents. This work is presented in some detail in Appendix I.

It has now become important to take into consideration the specific sources of dietary fat especially in long-term toxicity studies. The analytical data for the composition of corn oil per se suggests no basis for toxicity at the levels used in digestibility or metabolic studies in animals, or as normally consumed in human diets. Tests conducted to determine whether or not excessive dietary levels of oils may introduce confounding variables in toxicity tests have suggested that this may be the case.

The physiological and biochemical effects of increased polyunsaturated lipid intake in the diet may be expected to include: 1) stimulation of pancreatic enzyme and bicarbonate secretion; followed by 2) increased pancreatic enzyme synthesis; and 3) increased pH in the upper small bowel; 4) alteration of the quality and quantity of bile acid secretion; 5) alteration of gastrointestinal motility; 6) increased prostaglandin synthesis in the gut mucosa; 7) increased peroxidation of tissue lipids and tissue superoxide and peroxide content (Iritani et al., 1980 a,b); 8) altered fatty acid composition of tissues and consequent changes in membrane fluidity; 9) altered activity of hepatic and other microsomal oxidases; 10) decreased hepatic and adipose tissue lipogenesis (Baker et al., 1981; Herzberg & Rogerson, 1981). The first five of the above effects probably result from increased intake of any fat; the last 5 may be more accentuated by polyunsaturated fat due to its free fatty acid (FFA) content; especially when compared to high fat diets low in FFA or fat-free or low FFA diets.

The effects of vegetable oils in carcinogenesis studies may involve chemical bio-availability, bio-transformation or direct effects on neoplasms.

Consideration of Options

There probably are differences in the effects of lipid delivered in a bolus by gavage and delivered in the diet. The differences have not been studied, and yet they should be if oil gavage is to be used in further testing of chemicals for toxicity and carcinogenicity.

The critical areas for investigation include studies on: 1) extent and rate of absorption of the test compound; 2) distribution and metabolism of the test compound; and, 3) alteration by the lipid bolus of hormonal milieu, cell division and other factors that alter the target organ response to the test substance or its metabolites.

The substances administered in oil are lipophilic and presumably absorbed from the gut by mechanisms common to other lipophilic substances. Absorption requires solubilization in the gut lumen by bile salts within micelles that are composed also of triglycerides and the products of their digestion.

The presence of oil in the intestine and the dietary fat content can alter absorption of a test substance. DMBA absorption from an isolated small bowel loop was measured in rats fed a high-lard diet; absorption was decreased compared to controls. Perfusion of the intestinal loop with 5mM oleic acid, comparable to the concentration after a high fat meal, further decreased absorption (Hollander and Rogers, unpublished).

Metabolism of the test compound may be altered by the vegetable oil bolus since dietary corn oil has been shown to influence hepatic Mixed-Function Oxidases (Newberne et al., 1979) catalase, glutathione peroxidase and superoxide dismutase (Iritani and Ikeda, 1982). The alteration of critical enzymes for activation and detoxification would be expected to have a significant influence

on test compound metabolism, distribution, excretion and tissue effects. The administration of a substance in food may significantly lower peak plasma levels and increase biological half-lives.

Peak levels in the plasma or at target organ sites resulting from bolus dosage by any vehicle may exceed the capacity of the animal to detoxify the substance, as compared with the lower sustained levels resulting from dosage via the diet or drinking water. This hypothesis needs to be tested.

Factors that influence the response to gavage include duration of the pre-dose fasting period; the nature of stomach contents before and after dosage; stress, trauma or accidents due to the technique of administration; the concentration of test material in the vehicle, especially when it may be irritant or inflammatory; the volume of the test doses; and how oil may affect the caloric density and fat composition of the total diet. For example, a daily dose of 1.5 g of an edible oil added to the diet of a 400 g rat, consuming 20 g of a 5% fat diet, will increase the fat intake to 2.5 g, or by 150%. The effect of the diluent (fats in particular) on gastric emptying time, the pharmacokinetics of absorption, and the extent of peak blood levels at target organ sites with specific test agents is worthy of study.

Alternatives to oil gavage include: 1) the use of powdered basal diets in combination with the serving of a small portion of the diet containing the test compound in non-scatter feed cups when the animal normally begins to eat; 2) the use of a pelleted diet in combination with the serving of a diet pellet(s) containing the test compound as in the above alternative; 3) the use of a microencapsulated test compound incorporated in the diet in combination with use of non-scatter feed cups and regular measurement of food intake; and 4) the use of dispersing agent (e.g. Tween 60) or a suspending agent (e.g. carboxymethyl cellulose) with the test compound in an aqueous gavage solvent.

G. Duration of Study

Recommendation

1. In view of the rate of appearance of spontaneous tumors as the test animals age, NTP should carry out studies using both ongoing assays and the recently completed assays to determine the optimum termination point for the bioassay.

Background and Review of the Literature

According to the current standard NTP rodent bioassay protocol, administration of the test substance begins at 6 weeks of age and continues until 110 weeks of age, (~2 years), unless there is excessive mortality prior to that time. A histopathological examination is performed on animals that die spontaneously during the study, on those that are killed when moribund and on those that are killed at the end of the study, as described in other sections of this report. Although for reasons of cost and efficiency it would be desirable to shorten the exposure period and to sacrifice animals earlier, say for example at 12 months, there are reasons why this would not be prudent. Chief among these would be a loss in sensitivity of the assay and a significant increase in false negative results (Solleveld, et al., 1983).

Consideration of Options

The major purpose of these bioassays is to assess the effects of chronic exposure to substances in question. This includes the assessment of carcinogenic effects that might occur only after repeated and prolonged exposure, and only after a considerable period of latency, in part due to the multistep nature of the carcinogenic process.

It is known that the latent period in humans following exposure to radiation and the occurrence of breast cancer, or to cigarette smoking and the occurrence of lung cancer, can occupy a major fraction of the lifespan of the species.

Such adverse health effects would not have been detected if the period of observation was markedly curtailed. Latent periods that occupy a major fraction of the lifespans of rodents are also frequently seen in experimental carcinogenicity studies in rats, mice and hamsters.

Most human tumors occur during the latter third of the human lifespan. It is important, therefore, to allow experimental animals to survive for most of their natural lifespan to examine an equivalent stage of life with respect to tumor incidence. The median lifespan for Fischer 344 rats (both male and female) on a control diet is 28 months (Solleveld et al., 1983).

A recent comprehensive survey of previous carcinogenesis bioassay data indicates that if the period of observation were curtailed from two years to 12 months a significant number of otherwise positive assays would have appeared negative.

However, it would appear that lengthening the rodent bioassays by allowing all animals to live out their lifespan ("lifespan studies") would not appreciably increase the senstitivity of the assay or alter the variety of neoplasms that are seen. The use of a protocol that continues exposure until all the animals have died considerably complicates the conduct of large scale bioassays since animals die at inconvenient times and they may be lost for pathologic studies due to cannabilism, autolysis of tissues, etc. In addition, lifespan studies have a disadvantage over two-year studies because of a higher background incidence of several types of neoplasms, which could increase the chance of obtaining false positives (Solleveld et al., 1983).

In general, it does not seem wise to terminate exposure of the animals to the test substance at a time that is several months prior to sacrifice, since this diminishes the duration of chronic exposure, and may mask a process which would otherwise be progressive. If there is concern that the test substance might be hazardous to individuals processing the rodent tissues (because of

volatility or other reasons), then exposure of the animals can be terminated a few days or weeks prior to the time of sacrifice.

Since previous data obtained with the Fischer 344 rat indicate that most of the common neoplasms were already present between 85 and 97 weeks of age, a study duration of 18 months for the F344 rat may be as appropriate as a two-year study for testing the carcinogenic potential of a given chemical. However, more data, including interim kill evaluations, would be needed to establish that there would be no sacrifice in the sensitivity of the assay.

In specific assays in which species other than rodents are used it would be appropriate to considerably lengthen the period of chronic exposure prior to sacrifice since there is evidence that at equi-doses the latent period for cancer induction is roughly proportional to the life span of the species.

H. Use of an in utero Exposure System

Recommendations

1. The in utero design should be considered where the test agent has reproductive or teratogenic activity.
2. The in utero design should be considered when the pattern of use, nature and degree of exposure, pharmacokinetics or metabolism or other data suggest that this is a more appropriate means of testing.

Background and Review of the Literature

The use of the in utero design for subchronic or chronic studies has been encouraged by regulatory agencies concerned with food and food additive safety. This section of the report is based on considerations developed by the U.S. Food and Drug Administration and presented in greater detail in Appendix II.

Most long term studies in experimental animals utilize weanling animals without prior exposure to the test compound. However, the need to expose animals in a way that more closely parallels human exposure should include the

important phases of exposure that occur during fetal development and suckling of the infant.

The exposures of the test animal can then involve both parent compound and maternal metabolites of the parent compound that can either cross the placental barrier or enter mother's milk, or metabolites formed in the developing embryo which may or may not be the same as those metabolites formed in the maternal system. It is also possible that the concentrations of test material in milk could exceed the MTD during the nursing period.

The end result is to expand the usual screening for potential deleterious effects by providing information on the susceptibility of the embryo, developing fetus, and parental animals to the test compound. The susceptibility of the test animal during these stages of development may provide a particularly sensitive indicator of the toxic effects of the test compound.

Consideration of Options

A number of questions have been raised regarding the relevance of *in utero* studies. These include: 1) differences in structure of placenta in humans and experimental animals and how this relates to the transfer and/or metabolism of the test compound, and 2) distribution of the test compound in the fetuses and the concentration relative to that in the maternal tissue. Considerations of these are important in evaluating the results of *in utero* studies.

The determination of whether the use of the *in utero* design is important for either subchronic or chronic toxicity studies is based on application of the following considerations:

- The *in utero* design should be considered in those instances where the lowest "effect" level is only a small multiple of the expected human exposure level.
- Certain classes of agents, such as nutritive or nonnutritive food additives, should be considered for the *in utero* design.

- The *in utero* protocol should also be considered where pattern of use, nature and degree of exposure, pharmacokinetic data, metabolic findings, differences between *in utero* and conventional designs in preliminary work or other data suggest that this may be the more suitable approach.

I. Husbandry Requirements and Quality Control

Recommendations

1. A proper laboratory animal environment is a combination of facilities, personnel and management. Performance standards bring all of these into operation. Management must assure compliance.
2. A clear statement of goals and objectives in a study will help to define the performance standards appropriate to that study in the specific facility.
3. These performance standards relate to disease control, sanitation, husbandry, cage placement and rotation and environmental control.
4. Each bioassay should include contemporary quality assurance, good laboratory practice and related activities.

Background and Review of the Literature

The major sources of information for this topic are the DHEW (1978) and ILAR publications (1976, 1978). An important collateral source is the Proceedings of a Working Seminar on the Optimal Use of Facilities for Carcinogenicity/Toxicity Testing (Boiling Springs, PA, May 27-29, 1980). The subpanel has drawn on these sources, and especially on the publications and suggestions of Thomas E. Hamm, D.V.M., Ph.D., in compiling this section. Details of this material are in Appendix III.

In the event the desired level of protection from designated pathogens is not achieved during the course of the study, there should be a review to answer the following questions:

Experience with the early NCI bioassays led to the development of program guidelines (Sontag, et al., 1976). In addition to defining such items as species and strain of animals, number of animals per dosage group, duration of study, animal care criteria, environmental factors, etc., the protocol also outlined necropsy procedures and a list of tissues for subsequent histopathologic examiniation. A necropsy examination was conducted on all animals which died during the study and on those which lived to the end of the study (typically 2 years). This was followed by routine histopathologic examination of 32 sections of 25 organs/tissues from every animal on the study. Additional sections were examined based on necropsy findings of tissue masses and other lesions.

As experience with the protocol was gained the need to examine additional tissues became necessary, i.e., nasal cavity neoplasms were found to be associated with inhalation of 1,2-dibromoethane (NTP/TR-210, 1982). As a result, the current protocol now requires the histopathologic examination of 42-44 sections of 31-33 organs/tissues (Table 1). Historically, the standard carcinogenesis 2-year study (2 doses plus control) consisted of 600 animals (300 mice and 300 rats), thus an average 25,000-26,000 sections were evaluated on each test chemical. This represents a minimum of one-half professional man year on the part of the responsible pathologist in addition to the support provided by the histology laboratory. Current NTP, EPA and OECD 2-year study designs often incorporate 3 dose levels plus controls making the projected pathology workload even greater.

The pathology results from 277 NCI/NTP rodent carcinogenesis studies were reviewed to tabulate those organs/tissues reported to show a chemically related increase in neoplasia. In addition, the control groups from these studies were reviewed to tabulate those tissues/organs reported to have a spontaneous incidence of >1% of neoplasia. All neoplasms were evaluated to determine those that

1. Is the study compromised as a result of the findings?
2. Was the incident reasonably avoidable and, if so, what actions should be taken?

Quality Control is an important component of contemporary bioassays. This is a far-reaching concept involving established procedures, monitoring, documentation and independent review of all steps and phases in the assay from animal and test agent procurement to final report preparation. The degree and extent of review can vary from that based on statistical sampling considerations to a 100% check and tracking of each data entry item. The depth of monitoring and independent review is based on experience, results of preliminary audits, type of function audited, complexity of the study and the judgment of the reviewers. Too many studies have surfaced with what were later found to be flaws or correctable errors to permit any current significant experiment to proceed without a parallel quality control effort.

J. Pathology Requirements

Recommendations

1. Every animal must be subjected to a thorough post-mortem examination conducted by, or under the direct supervision of, a pathologist as there is only a single opportunity to examine a given animal.
2. The existing heavy burden of microscopic pathology can be reduced through use of either the Inverse Pyramid or Selective Inverse Pyramid method of identifying target tissues in some or all test groups. These methods are detailed in Appendix IV.
3. Statistical inferences from the use of Inverse Pyramid methods should be examined by the NTP staff.

Background and Review of the Literature

This portion of the Subpanel report is based largely on an evaluation of NTP staff papers prepared by Ernest E. McConnell, D.V.M.

TABLE 1

ORGANS/TISSUES EXAMINED HISTOPATHOLOGICALLY

Gross lesions and tissue masses (and regional lymph nodes)

Mandibular and mesenteric lymph node

Salivary gland

Sternebrae, femur; or vertebrae including marrow

Esophagus

Stomach - forestomach and glandular

Small intestine (duodenum, jejunum, ileum)

Cecum

Colon and rectum

Liver - left and right lobes

Gall bladder (when present)

Testes/epididymis

Prostate/seminal vesicles

Uterus

Ovaries

Nasal cavity and nasal turbinates

Trachea

Lungs and mainstem bronchi

Preputial or clitoral glands (paired) - rats only

Heart

Brain (three sections, including frontal cortex and basal ganglia, parietal cortex and thalamus, and cerebellum and pons)

Thymus

Pancreas

Adrenals

Pituitary

Thyroid

Parathyroids

Spleen

Kidneys

Urinary bladder

Spinal cord (in neurologic signs were present

Eyes (if grossly abnormal)

Mammary gland

Skin (skin paint studies to be delineated in specific protocols)

would have been recognized during necropsy and therefore would have been selected for microscopic study. Toxic lesions and the lesions associated with aging were evaluated in a similar manner.

Approximately one-half (56%) of the chemicals were evaluated as animal carcinogens. To be included in this list the chemical had to be associated with an increase in the incidence of neoplasia in at least one organ/tissue of one sex of either species (mouse or rat). The number of tissues/organs is approximately equal (28) to the total number (32-33) of those which are routinely examined. However, of these 28 sites 12 were recognized only as a result of a gross observation during necropsy examination, with subsequent histopathologic verification. In addition, most of the neoplasms in 12 of the remaining 16 tissues/organs were recognized as grossly visible abnormalities during the post mortem examination or during gross trimming of the tissues. Neoplasms of endocrine glands (pituitary, thyroid, adrenal and pancreas) are notable exceptions to the routine ability to presumptively detect a tumor or mass at time of necropsy.

A review of the control animals in the data base of 2-year studies revealed 12 tissues/organs in control F344 rats and 7 tissues/organs in control B6C3F1 mice with an incidence of >1% neoplasia. With the exception of testicle in F344 rats, all tissues/organs with a spontaneous incidence of >1% neoplasia are found in the list associated with chemical exposure.

Definitive information on the incidence of non-neoplastic toxic lesions in two-year studies is unavailable because: 1) non-neoplastic lesions are available only on a study-by-study basis (program wide data is not available); 2) the lesions are coded as morphologic diagnoses without reference to etiology; 3) the primary thrust of chronic studies prior to establishing the NTP was to determine the carcinogenic potential of a given chemical (non-neoplastic lesions

were given less attention); and 4) in chronic studies toxic lesions are often masked by spontaneous aging lesions. However, in the view of NTP the tissues/organs listed in Table 2 are those which most often show direct toxic lesions. Of these 14 tissues/organs only the testicle does not appear in the list of tissue/organs showing chemical related increased incidence of neoplasia.

Recognition of non-neoplastic toxic lesions and spontaneous disease is more complex. While abnormalities in size, shape, texture and color are hallmarks of neoplasia, this is not necessarily true for non-neoplastic lesions. A given tissue/organ may appear absolutely normal at necropsy yet contain significant pathology that can only be detected microscopically. The reason for this is that toxic lesions are manifested by varying degrees of atrophy, degeneration, necrosis, hypertrophy, hyperplasia, and/or distortions in normal architecture. Depending on severity, they may not be visually detected at necropsy. The fact that aging lesions, which may not be macroscopically apparent, occur in a vast array of tissues/organs, also dictates a requirement for histopathological examination of a large number of sites.

While the need to establish the chronic non-neoplastic disease potential of a given chemical is clear, the value of ascertaining this data by using an aged (2+ years) rodent is suspect. Lesions normally associated with aging may obliterate or at least mask a toxic lesion. This is most common in the kidney of rats and mice where chronic nephropathy may be so severe it precludes diagnosis of any lesion other than neoplasia. Spontaneous or induced neoplasms in this and other organs also often obscure toxic lesions and may in fact induce other lesions. For example, lesions in the liver of F344 rats afflicted with mononuclear cell leukemia are so severe that hepatocellular neoplasia is sometimes falsely diagnosed.

TABLE 2

CORE LIST OF ESSENTIAL TISSUES EXAMINED IN THE "SELECTED INVERSE PYRAMID" APPROACH

Adrenal gland-left and right	Pancreas
Brain-3 sections	Pituitary gland
Heart	Prostate/seminal vesicles
Kidney-left and right	Spleen
Liver-left and right anterior lobe	Stomach
Lung	Urinary bladder
Lymph node-submandibular	Testis/epididymis
Ovary/uterus	Thyroid and parathyroid glands

The current histopathology requirements in rodent carcinogenesis studies appear to be in excess of what is needed to answer the question of whether a given chemical causes cancer in the test species. Additionally the present protocol may be flawed in its ability to detect non-neoplastic but chemically related disease. Because of this the NTP is currently including a 15-month evaluation in its protocols to address this latter problem.

Consideration of Options

The ideal modification is one which would retain the test's ability to:

1. Detect an increased incidence of neoplasms.
2. Detect unusual or unique neoplasms.
3. Detect reduced time to the occurrence of neoplasm.
4. Detect direct and indirect toxic lesions.
5. Detect a chemically related increased incidence of/or severity of lesions associated with aging.

These options and details of proposed pathology protocol changes are discussed in Appendix IV.

In deciding whether a modification would be superior to the present protocol, the basic premise needs to be reiterated. Any modification must retain an equivalent power to ascertain an altered rate of neoplasia and define any chemically related non-neoplastic disease. At the same time it should, if possible, add to the scientific data base. The more savings of pathology effort the better. Finally, the modification needs to be straight forward enough to be clearly understood and easily implemented in a wide variety of laboratories. As discussed in Appendix IV either the "Inverse Pyramid" or "Selected Inverse Pyramid" may be suitable and both are superior to the current protocol in terms of savings of pathology workload.

K. Statistical Issues in the Interpretation of Data from NTP Whole Animal Carcinogenesis Bioassays

Recommendations

1. Results of a complete statistical analysis should be made available to the Technical Report Review Committee. Among the various statistical tests available, the test for a linear trend in tumor rates with increasing dose generally provides the single most sensitive and appropriate statistical indication of the presence or absence of a carcinogenic effect. The assumptions that underlie its use should be checked by further analysis of the dose-response curve including pairwise comparisons of the control with each dose group.

These tests should be age adjusted by a method appropriate to the presumed relationship between tumor and death, and additional information should be sought from serial sacrifice experiments regarding this relationship.

2. NTP is encouraged to utilize historical control data, especially recent data from the laboratory that conducted the bioassay, in the interpretation of results of chronic studies. Further analyses should be conducted of the accumulating database to determine the key sources of extraneous variation in tumor rates. Primary reliance should be placed on the comparison of treated animals with concurrent, randomized controls for evaluation of commonly occurring tumors. High priority should be given to ascertaining the extent of extraneous within-assay variation through use of replicate control groups and the analysis of data from recent FDA studies that utilized such a design.

3. Analysis of the NTP database should continue regarding possible negative associations between tumor types and other factors that may contribute to the observation of negative trends.

Background

The statistical procedures currently used to evaluate the results from the NTP carcinogenesis bioassay program are based on established standards of scientific inference. They involve mathematical assumptions whose validity is assured in practice by careful adherence to a well defined experimental protocol. This section starts by outlining some of the basic principles of experimental design that underlie the inferential process. Questions regarding the number and placement of dose levels and the numbers of animals per dose are not considered as they are dealt with in other sections. The role of formal statistical tests is considered, followed by a discussion of issues that arise in the interpretation of bioassay results. The recommendations section suggests modification of the current experimental design and the conduct of further analyses of the historical database so as to resolve some unanswered questions about important sources of variability in the carcinogenesis response. Depending upon the answers to these questions, more formal incorporation of historical data into the statistical evaluation may be desirable in the future.

Issues in Experimental Design

The current NTP bioassay involves the comparison of the carcinogenic response in untreated or vehicle control animals with that in two or more groups of dosed animals. Statistical tests commonly used to carry out this comparison assume that differences between treatment groups are due solely to treatment, or else to factors that are adequately controlled by the test procedure, and that results for individual animals are statistically independent.

Since a large number of genetic and environmental factors are known or suspected to influence tumor occurrence, while variations in pathology may affect the diagnosis of a tumor that has occurred, it is clear that rigorous adherence to the experimental protocol is needed for these assumptions to be met

in practice. For example, the physical act of randomizing individual animals to different dose groups helps to ensure that genetic factors and other pretreatment animal characteristics are equally distributed among treatment groups. Maintenance of uniform conditions within the treatment rooms, or periodic rotation of animals where this is not possible, should minimize or equalize the effects of local environmental factors. Strict rules for animal handling, necropsy and histopathology help to eliminate biases in evaluation. Of course, an experimental protocol that is complicated or rigid may be too costly to implement or may increase the chance for protocol errors. One needs to balance the needs for statistical precision and avoidance of bias against the costs of logistical and administrative problems and the risk of major errors that could invalidate the entire experiment.

Potential Conflicts between Theory and Practice

Various aspects of the NTP bioassay design as currently practiced have the potential to invalidate the statistical assumptions. It is not always clear from the technical reports, for example, whether animals are individually randomized to cages before treatments are assigned. This raises the possibility that genetically similar animals, such as littermates, are more likely than not to receive the same treatment. When animals are caged together, there are serious questions about the assumption of statistical independence of results for individual animals. One possible mechanism for a cage effect on carcinogenesis is the increased likelihood for transmission of biologic agents between animals in the same cage. The potential effect of such "within cage correlations" is to decrease the number of independent sampling units and thereby to increase the variability of response over that assumed by binomial sampling.

Failure to rotate cages opens up the possibility of systematic biases between treatment groups due to their differential exposures to environmental

agents present in the treatment rooms. The retinal pathology experienced by animals in the highest row of cages in some studies, presumably due to their proximity to the light source, illustrates the potential problem. Another example is when an infectious disease outbreak affects only one of the treatment groups. If logistical considerations rule out periodic cage rotation as a feasible part of the experimental protocol, animals receiving the same treatment could be distributed among various possible locations so as to permit covariance analyses that adjust observed differences among treatment groups for the possible effects of location. When confounding of treatments and locations is complete, as for instance when all high dose animals remain in the lowest tiers for the duration of the test, it is impossible to separate the treatment and location effects by any statistical manipulation of the data.

Distinguishing Carcinogenic from Non-Carcinogenic Responses

One of the major difficulties of interpreting data from chronic bioassays is the fact that the administered test compound may have a wide variety of biological effects. These include acute toxicity, early death, induction of bladder calculi or other physical deposits, weight gain or loss, alteration of hormone levels, immunosuppression and induction of cysts or benign growths. Some of these events may represent intermediate or contributory steps in the carcinogenic process, in which case they should presumably be counted in the evaluation process. Others apparently have little or no relationship to cancer formation but interfere with the detection of a carcinogenic response, for example, by shortening of the lifespan. Still others may contribute to the neoplastic process, but by a mechanism that may have little relevance to the human situation. The uncertainty about what the particular biological event means, and about how it should be accommodated in data analysis and interpretation, reflects our lack of knowledge of the mechanisms of tumor formation. Part of the dilemma is due to the necessary focus on carcinogenic as opposed to other

biologic responses for extrapolation to humans. Although these are not statistical issues per se, they are mentioned here because statistical methods have been proposed in order to attempt to deal with them.

One illustration of an interfering effect is to consider how the induction of non-neoplastic lesions can effect the probability that a malignant tumor is detected (Salsburg, 1977). Some pathology protocols call for one or two "random" sections to be taken through each of a number of specified organs, with additional sections taken from suspicious nodules. Since the ultimate diagnosis of cancer is made at the microscopic level, regardless of whether gross lesions are observed, it is clear that increasing the number of sections increases the probability of detecting microscopic lesions that were not evident during gross necropsy. An agent that produced cysts could therefore appear to have a carcinogenic effect simply by increasing the number of slides evaluated. The problem could be avoided, albeit at greatly increased cost, by taking serial step sections through each organ regardless of the presence of any gross lesions. Current NTP practice is to control for the number of sections taken in the design.

Consideration of Options

1. Statistical Tests used to Infer Carcinogenicity

Statistical tests are used to evaluate the possible role of chance mechanisms in producing the differences in tumor rates that are observed between control and treatment groups. The lower the calculated p-value, the less likely it is that the differences can be explained solely on the basis of the random assignment of animals to treatment group. However, the same data can lead to a variety of p-values depending on the particular test selected.

This section discusses the rationale and interpretation of the statistical tests commonly used to compare the observed proportions of tumors of each type in control and treated animals of a given sex and species: 1) the Fisher exact

test (Pearson and Hartley, 1966), used to compare two treatment groups (control vs. low dose or control vs. high dose); and 2) the Cochran-Armitage (Cochran, 1954; Armitage, 1955) test, used to compare three or more dose groups simultaneously for evidence of trend. Under the null hypothesis of no treatment effect, both tests assume that the observed numbers of tumors in the various treatment groups follow binomial sampling distributions with equal probabilities of observing a tumor. This implies that the observations are statistically independent, so that the occurrence of a tumor in one animal provides no information about the likelihood of tumor occurrence in any other animal. Experimental procedures needed to assure the validity of this assumption were considered earlier.

a. The Fisher Exact Test

Statistical tests for carcinogenicity evaluate the null hypothesis that the proportion of animals that develop a specified tumor in the control group is equal to that same proportion in the treatment groups. The usual test for comparing the control and one treatment group is to calculate the "chi-square" statistic for a 2x2 table and to refer it to tables of the theoretical chi-square (squared normal) distribution in order to assess the degree of statistical significance. Unfortunately, even under the null hypothesis, the distribution of the test statistic is not exactly chi-square and in fact depends on an unknown, namely the tumor rate that is assumed common to both groups.

Use of the Fisher exact test alleviates this problem by basing the inference on the conditional distribution of the data given the total number of animals with tumor in both groups combined. The resulting p-value is not dependent for its validity on asymptotic approximations, nor does it vary with the spontaneous rate. However, the price paid for this mathematical security is to restrict the possible p-values to a much smaller set of numbers than might

otherwise be the case (Liddell, 1978). Since a test is declared "positive" only if the observed p-value is smaller than some pre-specified value, typically .05 or .01, use of the exact test induces a degree of conservatism such that the actual probability of declaring a false positive may be much smaller than the nominal 1% or 5%. Table 3, prepared for the typical comparision involving a treated and a control group with 50 animals in each (Haseman, 1983), shows that the conservatism of the exact test is especially pronounced for rare tumors. Use of the trend test discussed below increases the possible set of p-values, and thus reduces the degree of conservatism, and it would be worthwhile to extend Table 3 so as to include results for this test.

The principal cause of this effect is that a certain minimum number of tumors of specified type must be observed in the combined (treated plus control) sample before the exact test can possibly yield a significant result. Under the null hypothesis that the tumor rates are equally low in both groups, it is very unlikely that this minimum number will be seen. The same phenomenon occurs with exact age-adjusted analyses that condition on the total number of tumors observed in each age stratum (Cox, 1970 and Gart, 1971).

While these calculations show that the true false positive rates for rare tumor sites are quite low, the sensitivity of the bioassay for detecting biologically meaningful increases in tumor rates is unfortunately also low. Table 4, prepared by Gart, et al., 1974, shows the false negative rates using the Fisher exact test with a one-sided significance level of $\alpha<0.025$. The comparison of groups of 50 control and 50 treated animals has virtually no chance of detecting a doubling or even a fivefold increase in the rates of rare tumors. If tumor rates are increased from 1% in controls to 10% in treated animals, the probability of obtaining a false negative is still 0.73. However with an increase from 2% to 20% the false negative rate falls to 0.22.

Table 3

True probability of declaring a false positive at the indicated level of statistical significance, using the Fisher exact test

Spontaneous tumor frequency (%)	Nominal Significance Level (one-sided)	
	0.05	0.01
1	0.000009	0.000004
2	0.0012	0.00002
5	0.011	0.0011
10	0.024	0.0038
25	0.029	0.0058

Table 4

True probability of declaring a negative (non-significant) result at the one-sided $\alpha \leq 0.025$ level of significance, using the Fisher exact test

Spontaneous tumor frequency (%)	Factor by which the spontaneous rates are increased in the treated animals		
	Twofold	Fivefold	Tenfold
1	1.00	0.98	0.73
2	0.99	0.81	0.22
5	0.94	0.25	0.0002
10	0.81	0.0068	0.0000
25	0.33	0.0000	0.0000

b. Multiple Comparisions Among Dose Groups

When two or more dose groups are used, it is common practice to make separate comparisons of each one with the controls. Some investigators may also pool together medium and high dose animals for testing against controls, or the controls and low/medium dose animals for testing against the high dose group. The question then arises as to how to interpret the resulting multiplicity of p-values so as to render a single verdict on whether or not a carcinogenic response has been observed. If each separate test has a false-positive rate of 5% (of course they won't if an exact test is used), then the probability that at least one of them is "significant" could be substantially larger than 5%.

One way out of this dilemma is to adjust the p-values for the number of comparisons made. If four different tests are carried out yielding four separate p-values, one reports four times the minimum (most significant) p-value as the adjusted significance level and draws the scientific inference accordingly. (Thus four p-values of 0.067, 0.023, 0.14, and 0.037 would lead to 4 x 0.023 = 0.092 as the composite sigificance level adjusted for the number of comparisons.) The probability that at least one of the tests yields a false positive result is no more than the sum of the probabilities that any one of them does so.

This simple procedure, known popularly as the Bonferroni method, is conservative and may be wasteful since it ignores dependencies between the tests and fails to account for the ordering of the dose groups. An experiment where the control versus low dose comparison was significant at the 1% level whereas the control versus high dose comparison was not significant would be accorded as much weight as one where the control versus high dose comparison was significant at 1% whereas control versus low dose was only significant at 5%. Yet the latter situation would usually give substantially greater evidence for a true carcinogenic effect.

c. Testing for Dose-Response Trends

A better approach to simultaneous testing is to conduct a single test of the joint null hypothesis that the tumor rates in all dose groups are equal, one that is sensitive to alternatives where the tumor rates increase with dose. The Cochran-Armitage trend test assumes that quantitative dose levels are assigned to the control ($x_0 = 0$) and each of K dose groups ($x_1, \ldots, x_K$). It is designed to detect alternatives to the null hypothesis where some transform of the tumor rate, usually the logistic transform, is a linear function of the assigned dose-levels (Tarone and Gart, 1980).

The trend test does not get around the problem of multiplicity entirely since the value of the test statistic still depends on the dose metameter. If the actual doses are equally spaced (e.g., 0, 50, 100, 150 ppm) assignment of the coded dose levels $x_0 = 0$, $x_1 = 1, \ldots, x_3 = 3$ is tantamount to assuming the appropriateness of the arithmetic scale. Multistage models of carcinogenesis and other considerations suggest that tumor incidence at each age is approximately linear in dose, at least for low doses (Crump et al., 1976). This provides some justification for use of the arithmetic scale as recommended by IARC (1980). On the other hand, classical bioassay procedures used in other branches of toxicology are based on the observation that dose-response functions that are linear in log dose give a better description of data in the experimental range. Tukey (Mantel, 1980) has suggested using a dose metameter that is arithmetic at low doses and logarithmic at high ones as a sort of robust approximation to the situation believed to hold in practice. The main point is that the choice of coded dose levels should be based on the best a priori guess as to the nature of the dose-response relationship assuming that the agent being investigated is a true carcinogen.

An exact version of the trend test is available that is conditional on the total number of tumor bearing animals in all dose groups, or on the total of tumor bearing animals in each age stratum in the case of an age-adjusted analysis (Cox, 1970). This has the same advantages and disadvantages as the Fisher exact test used to compare two groups, namely freedom from asymptotic approximations but some restrictions on the range of p-values. The continuity corrected version of the Cochran-Armitage test used in the routine NTP reports provides an excellent approximation to the exact test, especially in situations where a large number of tumor bearing animals make the exact calculation complex and impractical.

d. Time-to-Tumor Analyses

Statistical analysis of carcinogenicity formerly were based on a simple count of the numbers of animals in each treatment group that had a specified neoplasm, divided by the total number of animals examined. More recently it has been common practice to incorporate the animal's age at death into the evaluation, using a variety of statistical approaches. The simplest of these (Gart et al., 1979) is simply to exclude from consideration all animals that died before a certain specified age, or before the age at which the first tumor was found. However, methods based on more complicated parametric (Pike, 1966; Druckery, 1967; Peto and Lee, 1973; and Kodell and Nelson, 1980) or non-parametric (Hall and Walburg, 1972; Peto, 1974; Turnbull and Mitchell, 1978; Mitchell and Turnbull, 1979; Dinse and Lagakos, 1982; and McKnight and Crowley, in press) statistical modeling are increasingly in vogue.

Time to tumor analyses have several goals. One is simply to increase the precision of the statistical comparison by utilizing the additional information available in the ages of tumor bearing animals at death. Most known carcinogens act to increase the rate of tumor appearance at all ages, so that tumors will

start to appear earlier in the treated than in the control group. Statistical tests based on time to tumor are more sensitive and powerful to detect such effects than are tests that ignore age. In theory it might even happen that a test compound accelerates the appearance of tumors, so that those that occur are found earlier in life, but does not increase the total lifetime incidence. Such an effect which is presumed to be evidence of carcinogenicity according to some official sources (WHO Technical Report, 1969; FDA Adv. Comm. on Protocols for Safety Eval., 1971) would be missed unless age was considered. Examples where there is clear evidence of acceleration in the absence of an overall increase in incidence are rare, however, and the issue of whether they are truly indicative of carcinogenicity remains controversial (Mantel, 1980 and Lagakos and Mosteller, 1981).

Another goal of time-to-tumor analyses is to attempt to cope with the effects of differential mortality mentioned earlier. A treated or control animal's age at death has a strong influence on the likelihood that a tumor will be found. This may be due either to an inherent biological process of aging, or to the fact that age is a surrogate for the cumulative effects of nonspecific environmental exposures (Peto et al., 1975). If acute toxicity results in a substantial number of early deaths, the total lifetime tumor count in treated animals could be less than that in controls whereas the percentage of tumor bearing animals among those that survive until terminal sacrifice is higher in the treated group. The issue is complicated by the fact that some tumors are rapidly fatal, so that animals dying with a tumor present may be presumed to have died from it, whereas other tumors are merely incidental findings at necropsy of animals that died of other causes. In the latter case early toxic deaths may lead to the earlier appearance of a tumor even though the agent had no effect on carcinogenesis per se. Since tumors may on some occasions be the

cause of the animal's death, and on others merely an incidental finding, certain pathology protocols call for an assignment of probable cause of death so that these two different situations may be treated differently in the statistical analysis (Peto et al., 1980).

The statistical methods used to control for differential mortality compare the tumor rates of treated and control animals at specific ages. Ideally one would like to compare treated and control groups in terms of age-specific tumor incidence rates, since incidence is the measure of tumor occurrence usually and appropriately applied in the human situation. However, with the exception of skin and mammary tumors, tumor incidence is not generally directly observable in experimental animals. The diagnosis of a tumor of an internal organ is only made at necropsy, which implies either that the tumor had grown large enough to kill the animal or else that the animal died of other causes. If the type of tumor under investigation is uniformly and rapidly fatal (fatal tumor), then age specific tumor mortality rates serve as a good approximation to tumor incidence. In this case the lifetable method of age adjustment is appropriate. One compares mortality rates of treated and control animals, i.e. the number of animals that died with tumor at a given age by the number alive and tumor free just prior to that age. On the other hand, if the type of tumor under investigation generally has nothing to do with the animal's death (non-fatal tumor), but is instead observed only as an incidental finding at necropsy, then treatment groups are better compared in terms of age-specific tumor prevalence rates defined as the number of animals with tumors present at death divided by the number necropsied. Tumors found at terminal sacrifice are treated equally by the two methods of analysis. Composite analyses using both mortality and prevalence rates, with tumor bearing animals being assigned to one category or another on the basis of cause of death or "context of observation" have been

proposed for intermediate situations (Mantel, 1980; Peto, 1974; and Peto et al., 1980). Another approach is to examine the results of both the tumor mortality rate and the tumor prevalence rate analyses and to regard the true situation as being somewhere in between. Generally the results obtained from the two methods will be quite comparable unless there are marked differences in survival between treatment groups, which gives yet another reason to avoid excess toxicity if at all possible.

The categories fatal and non-fatal represent ideal types of tumors that are only imperfectly realized in practice. When the particular tumor in question is not uniformly and instantaneously fatal, but nonetheless may eventually hasten the death of its host, large scale serial sacrifice data are required to disentangle an agent's effects on mortality from its effects on tumor production (McKnight and Crowley, in press). Statistical analysis procedures based on assignment of cause of death or context of observation can only be regarded as providing an imperfect approximation to an experimental protocol where animals are randomly selected for sacrifice at specified ages before the termination of the study (Kodell et al., 1982). Unfortunately, the numbers of animals required for serial sacrifice studies make them prohibitively expensive for use in a routine testing program.

Even if conditions are satisfied that justify the fatal (mortality) or non-fatal (prevalence) approaches to age adjustment, the results of the analysis have the desired interpretation only when the competing risks of tumor and non-tumor deaths operate independently of one another. For example, if animals that were genetically susceptible to tumor development were also selectively killed by acute toxicity, the analyses could be misleading whether or not they were age-adjusted. The effect of such an association could be to spuriously inflate the early tumor prevalence observed among treated animals, since susceptible animals with non-fatal tumors would die sooner than similar animals in the

control group. Alternatively, tumor incidence might be decreased among treated animals by the early elimination of susceptibles from the population at risk. No amount of statistical manipulation could counter the problem (Tsiatis, 1975 and Peterson, 1976). The plausibility of a strong dependence of both cancer and toxicological risk on a common genetic factor is an issue that must be debated on strictly biological grounds. Because of such uncertainties, less confidence has been placed on bioassay results whenever there were large differences in survival between treated and untreated animals, regardless of whether or not age adjustment procedures were used, unless the survival differences were clearly secondary to tumor differences in tumor death rates.

2. Interpretation of Bioassay Results

Because of the complexities of the carcinogenic process, and the uncertainty about the mechanisms involved, it would be unwise to attempt to use the results of the statistical tests in a formal decision framework to arrive at a verdict about any particular test compound. Instead, the evidence from the composition of tumor rates should be combined with whatever additional data are available from the bioassay on associated pathology, or data from *in vitro* tests suggesting possible mechanisms or sites of action, in arriving at an overall judgment as to the strength of the evidence for carcinogenicity. Some of the considerations relevant to this informal reasoning process are discussed below.

a. Coping with a Multiplicity of Tumor Types

Even if attention is focused on the age adjusted trend test, there are still a multiplicity of statistical tests to consider. Separate statistical analyses are carried out routinely for approximately 30 different tumor types or combinations for each sex and species used in the bioassay. The chance of observing a false positive result for at least one tumor type in at least one sex and species could be substantial if each tumor type were tested at a true 5% significance level, or even at a 1% level.

In practice, the overall false positive rate for NTP bioassays appears to be substantially less than naive calculations based on the multiplicity of tests would suggest (Salsburg, 1977; Fears et al., 1977). One reason is evident from Table 3, which as already noted should be extended to include results for the exact trend analysis. The actual false positive rates for rare tumors are much less than the nominal ones. Most of the danger that a chance assignment of animals to dose groups will give a false positive result comes from tumors that occur spontaneously at relatively high rates. An obvious way to protect against the undesirable consequences of making a multiplicity of tests, therefore, is to demand a higher level of statistical significance for common tumors than for rare ones. Nevertheless, some apparently significant findings are bound to arise by chance alone (Salsburg, 1977).

The problem of multiple comparisons is by no means unique to NTP bioassays. Scientists in all fields are faced with the problem of distinguishing true causal relationships from chance associations. Among the criteria typically used to aid in this process are evidence for a dose-response trend, replication of the same result in different settings, biological plausibility and the lack of alternative explanations (Interagency Regulatory Liaison Group, 1979). Rather than using fixed decision rules based solely on the results of formal statistical tests, mature investigators consider each finding in the light of their knowledge and understanding of the processes involved so as to arrive at a scientific judgment regarding its true biological significance. This helps to guard against the "false positive" results possible from use of purely statistical decision procedures. When a positive evaluation of carcinogenesis is based on clear dose-response trends in multiple and/or uncommon sites and is reinforced by ancillary pathology observations and short term tests, the statistical phenomenon of false positives is simply not an issue.

When applied to the bioassay, these criteria suggest that the strongest evidence for a carcinogenic effect accrues when there is a continuous rise in the tumor rate with increasing dose. This re-emphasizes the central role of the trend test. It suggests that stronger evidence could be obtained if the number of dose groups (including control) were increased from three to four or five so that the increasing portion of the dose-response curve were more likely to be included in the experimental range. A limited opportunity to examine the issue of replication within a single bioassay is obtained by comparing results for the same organ in males and females of the same species, or even to compare results between two species. The demonstration of carcinogenicity based on a positive response in the mouse liver is most convincing when a dose-response relationship is evident for both males and females, and formal statistical procedures for combining data from the two sexes are available to quantify the resulting inferences (Gart, 1971). However, it is well known that the target organs for established carcinogens frequently differ from species to species, possibly due to variations in metabolism, so that failure to replicate across species should not be taken as indicative of a negative finding (Tomatis et al., 1973 and Tomatis, 1979).

The statistical evidence may also be examined in the light of additional pathology data and results from other in vitro or in vivo studies of the test compound in order to reach a conclusion regarding its potential carcinogenicity. Even if the tumor type is a common one, it may happen that rare metastatic lesions are concentrated in the treated animals or that the dosed groups have a higher number of tumors per tumor bearing animal. Precursor lesions or other types of pathology may be present in the target organs of treated animals that can aid interpretation of how the disease progresses. Evidence that an agent has mutagenic or teratogenic effects can and should contribute to the overall interpretation of the results of the bioassay.

b. Sources of Variation in Historical Control Rates

The NTP has available a large database of results in control animals (primarily Fischer 344 rats and $B6C3F_1$ mice) from previous bioassays that used protocols similar or identical to the current one. This information is useful for interpreting the results of new assays and for resolving some of the outstanding issues regarding extraneous sources of variation in tumor rates. Possible sources for the variability in tumor rates observed between different control series include: genetic variations in the treated animals; diet; contamination of treatment rooms; methods of animal handling; differential rates of infection or intercurrent mortality; differential body weight gains; the vehicle used to administer the test compound (in case of vehicle control); and differences in pathology techniques and diagnosis. One way of getting some feeling for the magnitude of the extraneous variability from such sources is to study differences in tumor rates between laboratories, between pathologists within laboratories, and over time. Haseman and colleagues (Haseman et al., 1983) have recently carried out a limited analysis of the NTP database and conclude that calendar time and laboratory are both important sources of variation for certain tumors. For F344 rats they noted increases in control rates for mononuclear leukemia, pituitary adenoma and adrenal pheochromocytoma as compared to earlier NCI sponsored studies (Goodman et al., 1979). The rates of liver neoplasms and pituitary adenoma in $B6C3F_1$ mice also have increased (Ward et al., 1979). Notable interlaboratory variations are seen for an even larger number of tumor types including: pancreatic islet cell tumors in male rats; neoplastic nodules of liver, C-cell tumors of thyroid, mammary fibroadenomas and endometrial stromal polyps in female rats; and lymphomas and pituitary adenomas in both male and female mice. Tarone and colleagues (Tarone et al., 1981) conducted a similar analysis of the earlier NCI bioassay program. Tumor types they identified as

having significant temporal or interlaboratory variability include: lymphoma-leukemia, liver tumors, pituitary tumors and adrenal pheochromocytoma in the male rat; pituitary tumors, mammary fibroadenoma and endometrial stromal polyps in the female rat; lung and liver tumors in male mice; and lymphoma-leukemia and liver tumors in female mice.

On the basis of these and other findings, a reasonable recommendation is that historical comparisons be limited to control data for the past 3-4 years from the same laboratory. (While there is some additional variation in tumor rates attributable to pathologist within laboratory, limiting the historical comparison to results for the same pathologist would unduly restrict the size of the historical database.) Even with these restrictions, some variation in control rates within the historical series will be evident. In addition to the variation in control rate between assays, there may also exist within assay variations in excess of that predicted by binomial sampling theory, for example, due to cage, position, litter, pathology or other effects as mentioned earlier. Tests conducted by NTP (Solleveld et al., 1984) and FDA (Lagakos and Mosteller, 1981) have on occasion demonstrated differences between replicate control groups within the same assay that are statistically significant on the basis of the same tests used to compare control and treatment groups. Since such variation obviously invalidates the use of such tests, it is extremely important that NTP ascertain the extent to which it may be present and the particular factors (cage, position) that may be responsible. Further analysis of the recent FDA experiments that utilized replicate control groups is called for, as is the search for within-assay effects in the NTP historical database. Consideration should be given to the use of replicate control groups in future NTP bioassays to monitor the situation within different contracting laboratories as a method of quality control. Changes in the experimental protocol could be instituted to

reduce the effects of any systematic sources of variation that may be found to be problematic. A finding that there was no extra-binomial variation between the replicate control groups would strengthen the credibility of the bioassay and reduce concern about the effects of a high degree of historical variation.

An appreciation of the role of within assay extraneous variation is critical for making a judgment about how the historical database should be used. If the historical variation is due mostly to factors that operate within assays, then we should use the historical database to estimate the degree of extra-binomial variability and adjust p-values accordingly. On the other hand, if the within assay variability is truly binomial, we should use the historical series to get a better estimate of the binomial probability (tumor rate) in the control group for comparison with the observed rates in the treated groups. This latter idea seems to lie behind most of the formal statistical procedures that have been proposed to date (Tarone, 1982 and Dempster et al., 1983).

c. Increasing Power in the Analysis of Rare Tumors

The most important use currently being made of the historical database is to assist reviewers in evaluating the significance of treatment related increases in production of rare tumors. As shown in Table 4, analyses that use concurrent controls exclusively are virtually powerless to detect meaningful increases when the spontaneous rates are low. A finding of 0/50 tumors in controls versus 4/50 in treated animals fails to reach significance using the Fisher exact test (one-sided $p = 0.059$). It takes on a much greater significance when viewed against an historical incidence of 0.1%, say, with no more than 1 or at most 2 tumors of that type being observed in any previous group of 50 control animals. Extraneous sources of variation are evidently low in such cases and one can ascribe the appearance of tumor to the effects of treatment with a reasonable degree of confidence.

d. Interpretation of Results for Common Tumors

The interpretation of results for tumors that have relatively high and variable control rates is much more difficult. Sometimes one has a positive result based on a comparison of control and treated animals within an assay, but finds that the control rate is unusually low in comparison to the historical rates. Reviewers have on occasion attempted to interpret the result by comparing the tumor rates in treated animals to the range of control rates observed in the historical series. Some refuse to accept the finding as evidence for carcinogenicity unless the rate in treated animals exceeds the maximum from all previous control series. Since the value of this maximum increases with the length of the control series, however, this is not a defensible procedure. A more reasonable approach is to compare the rate in treated animals to the 95th percentile, say, of the control rates, although this presumes that a rather large historical series is available.

Such comparisons make the implicit assumption that the historical variability is entirely due to extraneous sources of variation such as cage and litter effects that operate as much within as between different assays. This assumption may be false. For example, the low rate in control animals could be due to a shift in the genetic makeup of the source animals or a change in diet that applied just as well to treated as to control animals. It is precisely because of such uncertainties that one uses a randomized design and places greatest weight on the comparison with concurrent randomized controls.

If the extra-binomial variability were due entirely to between assay factors, an alternate use for the historical data would be to improve the estimate of the (unknown) tumor rate in the concurrent control group but continue to base the control versus treatment comparison on binomial sampling assumptions. Several formal statistical procedures have been proposed recently with this goal

in mind (Tarone, 1982; Dempster et al., 1983; and Hoel, 1983). In effect, they use a weighted average of the control rates observed in the concurrent controls and in the historical series, with more weight going to the latter as the historical variability decreases. This provides a formal justification for use of the historical series to help evaluate results for a rare tumor. If there is any substantial inconsistency between concurrent and historical control rates, however, greatest emphasis clearly needs to be given to the results in concurrent controls (Cox and McCullagh, 1982).

Further study of the historical database is needed to determine the variability associated with inter-assay factors such as cage, position and litter. Only then can one judge which of these two quite different approaches to historical data is most appropriate.

e. Interpretation of Negative Trends

It is not uncommon in NTP bioassays for the chemical agent under study to lower the rates of selected tumors. This is a matter of serious concern since some reviewers have used the occurrence of such decreases to raise doubts about the validity of assays where the agent increased tumor production (Salsburg, 1983 and Young, in press). In many cases there is a simple explanation for the decrease based on the same factors that were noted earlier to complicate the evaluation of positive findings, namely the agent's effects on mortality or weight gain or simply the play of chance. More intriguing is the possibility of genuine and as yet unexplained antagonism between certain tumor types.

In view of the marked dependence of tumor occurrence on age, a lowering of the lifetime tumor rates would be expected if the test compound increased the early mortality of treated animals over that of controls. Age adjusted analyses should then be used. However, a recent review of 25 NTP two-year feeding studies failed to identify differential mortality as an important contributor to

lowered tumor incidence (Haseman, 1983). In fact, more often than not survival was enhanced in the treated animals.

Dietary factors have long been recognized to have an influence on the production of spontaneous tumors (Tannenbaum and Silverstone, 1957; Gilbert et al., 1958; Ross and Bras, 1971; Roe and Tucker, 1978; and Tucker, 1979). Generally speaking, tumor rates are suppressed in undernourished animals. To the extent that treated animals receive less effective nourishment from their feed than control animals, as reflected in a decreased weight gain pattern, one would therefore anticipate that the treated groups would evidence lower rates for some tumors.

Haseman's (Haseman, 1983) recent review of 25 NTP studies confirms this relationship between weight gain and tumor production. After dividing treated groups of animals from all studies into four categories on the basis of their average weight gain relative to controls, there is a statistically significant trend in tumor rates with increasing weight for thyroid and adrenal tumors in male F344 rats and mammary gland fibroadenomas in females. The difference between the first and last categories is particularly striking for the mammary tumors: 187/849 = 22.0% in 16 groups having average weight gains between -2.5% and +6.5% of those of controls, versus 35/399 = 8.8% in 8 groups whose average weight gains were from 18.5% to 34.6% below those of controls. There was a significant increasing trend between leukemia-lymphoma rates and weight gain category in female rats.

Seven of the 25 studies showed a statistically significant ($p<0.05$) decreasing trend of mammary gland fibroadenomas with increasing dose. In all but one the average weight gain in the high dose group was 10% or more below that for controls. On the other hand (see below), the chemicals that significantly decreased leukemia or leukemia-lymphoma rates did not have a consistent effect upon weight gain.

f. Negative Association between Liver Tumors and Leukemia-Lymphoma

Study of the joint occurrence of different tumors in individual CF-1 mice from a multigenerational investigation of the carcinogenic effects of DDT turned up a possible negative association between liver tumors (hepatoma) and lymphomas (Breslow et al., 1974). This result was confirmed in a later study that involved extensive serial sacrifice and thus overcame doubts about whether the original finding may have been an artifact of the rapid lethality of lymphomas (Wahrendorf, 1983). Haseman's (Haseman, 1983) review of 25 NTP studies found a negative association between liver tumor and leukemia-lymphoma rates in groups of F344 rats receiving various treatments. However, there was no naturally occurring inverse association in individual control animals similar to what had been reported for CF-1 mice.

L. Errors and Error Rates

Recommendations

1. Continue monitoring of the quality of bioassays in contractor laboratories by NTP.
2. Systematically analyze the kinds and rates of errors in the conduct of bioassays and develop interventions to reduce or eliminate such errors.
3. Further develop the criteria for carcinogenicity in experimental animals and establish a list of non-carcinogenic substances for reference use by the scientific community.

Background and Options

Reproducibility - Animal bioassays for carcinogenicity are seldom replicated and thus the variability to be expected on repeated testing is not known. In most other scientific investigations, the reproducibility of events is demonstrated from the outset by utilizing multiple samples concurrently and by repeating each experiment under identical conditions (replication). Animal

bioassays for carcinogenicity are so expensive, however, that multiple bioassays with the same test substance are rarely performed. To replicate a carcinogenicity bioassay one would have to have the same batch of test substances, the same types of animals from the same suppliers, and the same experimental conditions. Although it is possible to repeat the bioassays under similar conditions, typically investigators design the next experiment to answer a different question rather than to confirm the results of the first experiment. It would be helpful if those who perform dose-response experiments would: (a) include at least one dose level in the range of expected responses as a positive control, and (b) adjust the size of the treatment groups to retain the power of the test at lower dose levels so that comparisons can be made along the straight portions of the dose curve with the results obtained in previous experiments.

It should be emphasized that the lack of replicate experiments in animal bioassays is partially compensated by the use of more than one dose level, both sexes, and more than one species, all of which represent multiple, concurrent experiments. Positive results in two or more of these groups add heavily to the weight of evidence for carcinogenicity and against the possibility of a false positive result.

Positive and Negative Controls - Another difference between conventional experiments and animal bioassays for carcinogenicity is that the latter seldom employ concurrent positive and negative controls (other than untreated or vehicle-treated controls). Two major factors that contribute to not using a known carcinogen as a positive control are the ever present problems of cost and the question of which carcinogen to use. IARC lists only 147 chemicals or chemical processes (IARC monographs, Suppl. 4) as having sufficient evidence for carcinogenicity in animals. Not all chemical subclasses are included and these "known carcinogens" differ considerably in potency and target sites. It is

sometimes argued that a model carcinogen such as benzo(a)pyrene be used as a positive control in each bioassay to decrease the possibility of false negative bioassays. Experimentalists who use potent carcinogens repeatedly for other purposes, however, find that the qualitative effects are uniformly repeatable and that the possibility that the model compound would not produce cancer is remote even when some previously untried test system is used. In the absence of prior information about possible mechanisms of action, positive controls are not recommended for routine use in bioassays. The use of model compounds as reference standards for potency comparisons, on the other hand, is a commonly accepted practice.

As negative controls, it would be desirable to expose groups of test animals to known non-carcinogens as another check on the variability to be expected in the incidences of background tumors under the test conditions. Unfortunately, few if any compounds have been sufficiently studied to establish them as non-carcinogens. Furthermore, the inclusion of a control group that receives only the vehicle or solvent for the test substance often serves as a suitable negative control.

Errors in the Conduct of the Experiment - The question of error rates in regard to bioassays for carcinogenicity usually brings to mind statistical considerations of the probability that the conclusions reached about carcinogenicity are correct. Long before the statistical evaluation begins, however, there are possible errors in each step of the bioassay. For example, the test substance may be incompletely identified. In the bioassay of N-nitrosodiphenylamine the substance tested in the U.S. (NCI Tech. Report Series No. 164) was dark brown, whereas that tested in Germany was colorless (R. Preussmann, Heidelberg, personal communication). The differences in the outcome of the two bioassays could be attributed to differences in dose levels but the

possibility exists that different substances were tested. Work in histopathology is readily audited at any time because the materials are still available after the experiment is completed, but quality control of the clinical aspects of a bioassay is much more difficult. How does one know that the animals actually received what they were supposed to? Did randomization of animals really take place according to the protocol, was the investigator proficient in deciding when moribund animals should be killed, was a pathologist present at every autopsy, were the samples collected for histopathology without knowledge of the treatment groups to which they belonged? These and many other questions indicate subjects in which further quality control measures are required. Moreover, the quality control procedures provide an opportunity to identify systematic and random errors and to count and record sampling and measurement errors. In the reviews of histopathology where considerable attention has been paid to error rates, in one study (unpublished) pathologists missed 2% of the neoplasms in groups of 600 aged animals. More examples could be given but all of them indicate the need for external monitoring and auditing of bioassays. Moreover, public disclosure of the kinds and rates of errors would assist everyone engaged in such activities in improving the quality (and reproducibility) of the bioassays.

M. Combining Benign and Malignant Neoplasms in Evaluating Carcinogenicity Recommendations

1. Where substantial evidence exists within the specific study that progression occurs from benign to malignant neoplasms in the same organ, then the incidence data may be combined to aid in the evaluation.
2. Neoplasms of the same histomorphogenic type may be combined even if they occur in different anatomic sites.

3. Neoplasms of different morphologic classification may be combined when their histomorphogenesis is comparable.

Background

The major input to the Subpanel on this topic was an invited staff paper prepared by Ernest E. McConnell of NTP. Portions of that document have been excerpted to form this section of the report.

Pathologists examining tumor incidence data from chronic bioassays are often faced with patterns of response which are not readily interpretable. To further aid in the evaluation of the findings once the incidence of benign and malignant tumors has been separately analyzed it may be useful to combine the incidence of tumors in the same organ or tissue or in different organs where the morphology of the tumor is comparable. There are a number of well-reasoned arguments favoring and opposing combination of tumor data and guidelines would be most helpful.

Consideration of Options

Five arguments may be advanced for combining certain benign and malignant tumors in evaluating the strength of evidence of carcinogenicity of a given chemical:

1. The terms benign and malignant are convenient for classification and, at times, clinical prognosis, yet their use may be an artificial division/separation of neoplastic development and progression.
2. For some neoplasms there is substantial evidence for progression from benign to malignant. In such cases evaluating benign and malignant tumors in isolation of each other may mask a true effect.
3. The morphologic criteria for differentiating benign and malignant tumors is often subjective and sometimes arbitrary. In fact, for certain lesions it is difficult to differentiate hyperplasia from neoplasia.

4. The categorization of a neoplasm as benign or malignant may bear little relevance to its adverse biologic potential. Some benign tumors are as life-threatening as their malignant counterparts.
5. In carcinogenesis studies time and resources do not allow for step-sectioning (multiple sections) of a given lesion diagnosed as benign to determine if malignant areas are present. Also the use of interim sacrifices to determine the progression from nonneoplastic lesions through the neoplastic stage is not always feasible.

In view of these difficulties, it is tempting to simply diagnose the lesion as a neoplasm or record only the most severe lesion. This would be a mistake for the following reasons:

1. Differentiating benign and malignant neoplasms provides at least circumstantial evidence for the aggressiveness of a particular type of tumor, thereby providing a better data base for interpretation of carcinogenicity and hazard assessment.
2. In some tissues the difference between benign and malignant neoplasms is clear. In organs where differentiation is less clear and possibly arbitrary it is still useful to differentiate between low and highly aggressive neoplasms. Malignant tumors may also appear to arise de novo.
3. The knowledge that for a given compound the treated animals had a higher incidence of malignant (or benign) neoplasms than the controls is valuable even though the total incidence of neoplasms (benign and malignant) may be comparable. Malignancy implies more extensive damage to the host and by definition irreversibility.
4. Some benign neoplasms have the ability to regress to nonneoplastic lesions.

5. Neoplasia is a complex disease involving many mechanisms. To diagnose a lesion merely as neoplastic or nonneoplastic would ignore this complexity and would not communicate to other scientists the type of information needed for decision making.

Benign and malignant neoplasms may be combined when it appears appropriate to combine them for statistical purposes, but this should be done on a case-by-case basis. In addition, it may be useful to evaluate nonneoplastic proliferative responses in conjunction with neoplastic ones in a given organ for an overall determination of the carcinogenic potential of a given chemical. In practice the NTP evaluates benign and malignant neoplasms separately and combines them to clarify the situation.

In Appendix V of this report are provided two tables with further details of tissue or tumor type combinations. Table 5 is a list of organs or tissues where combining benign and malignant tumors is or is not appropriate in order to better understand the evidence of carcinogenicity. These findings are keyed to the common bioassay species of Fischer 344/N rats and $B6C3F_1$ mice. Scientific judgment should be used in combination with this guide. Table 6 is a list of tissues where combining neoplasms of different sites is or is not appropriate for evaluating carcinogenicity. The same limitations mentioned for Table 5 apply to Table 6 as well.

Exceptions to the rule are possible for a given tissue site. Some of the more notable exceptions are:

Respiratory tract - While neoplasms of the same cell of origin in the larynx, trachea, and major bronchi would usually be combined, neoplasms of the peripheral lung (bronchioalveolar neoplasms) would be analyzed separately. Similarly, neoplasms of the nasal cavity would not normally be combined with tumors found in the rest of the respiratory tract.

Alimentary tract - Squamous cell neoplasms of the tongue, esophagus and forestomach (nonglandular) are often combined for evaluation. Likewise, epithelial derived neoplasms from various portions of the small intestine are combined, as are neoplasms from different areas of the large intestine (cecum and colon). Morphologically similar neoplasms of the small and large intestine are at times combined to evaluate the intestinal tract as a whole. Neoplasms of the glandular stomach are usually evaluated independently.

Smooth muscle - Leiomyomas (sarcomas) are combined for all sites of the body except the gastrointestinal and reproductive tracts where they are evaluated independently.

Genital tract (female) - Neoplasms of glandular epithelium are normally combined regardless of whether these arise in the fallopian tube, uterus, or cervix. However, they are not usually combined with vaginal neoplasms. In contrast, stromal neoplasms would be combined from all areas of the genital tract.

N. Categories of Evidence of Carcinogenicity

Recommendations

1. NTP should continue to develop a system for interpreting the findings of each bioassay (sex and species) with regard to the strength of the experimental evidence supporting the conclusion. These interpretations at present do not take into account potency or mechanism of action.
2. The proposed categories of evidence of carcinogenicity are as follows:
   - Clear evidence of carcinogenicity
   - Some evidence of carcinogenicity
   - Equivocal evidence of carcinogenicity
   - No evidence of carcinogenicity
   - Inadequate study of carcinogenicity

Background

The preparation and review of a bioassay report requires a detailed knowledge of all findings in the study. The more casual reader, however, is less familiar with the details and, in some instances, is not expert in the field of chemical carcinogenesis. The latter individuals are often uncertain of the meaning or significance of the bioassay especially if their reading is confined to the abstract or summary. The investigator's conclusions would not be risk assessments but merely statements regarding the data from that experiment.

Consideration of Options

For several years, the Technical Reports Review staff of NTP has attempted to develop language which would describe the conclusion to be drawn from the carcinogenicity bioassay. The group is aware of the system employed by the International Agency for Research on Cancer in the monograph series, the Evaluation of the Carcinogenic Risk of Chemicals to Humans. However, the IARC system classifies the overall evidence of carcinogenicity in experimental animals. Four categories are employed--sufficient evidence of carcinogenicity, limited evidence of carcinogenicity, inadequate evidence and no evidence. These criteria are given in Appendix VI. IARC holds that data in animals meeting the standard of the first two categories suggests potential carcinogenicity for man.

Table 7 is a general guideline used by NTP for classifying each assay.

Table 7

Guidelines Used by NTP for Classifying Assays

Clear Evidence of Carcinogenicity is demonstrated by studies that are interpreted as showing a chemically related increased incidence of malignant neoplasms, studies that exhibit a substantially increased incidence of benign neoplasms, or studies that exhibit an increased incidence of a combination of malignant and benign neoplasms where each increases with dose.

Some Evidence of Carcinogenicity is demonstrated by studies that are interpreted as showing a chemically related increased incidence of benign neoplasms, studies that exhibit marginal increases in neoplasms of several organs/tissues, or studies that exhibit a slight increase in uncommon malignant or benign neoplasms.

Equivocal Evidence of Carcinogenicity is demonstrated by studies that are interpreted as showing a chemically related marginal increase of neoplasms.

No Evidence of Carcinogenicity is demonstrated by studies that are interpreted as showing no chemically related increases in malignant or benign neoplasms.

Inadequate Study of Carcinogenicity demonstrates that because of major qualitative or quantitative limitations, the studies cannot be interpreted as valid for showing either the presence or absence of a carcinogenic effect.

REFERENCES

Anders, F., Schwab, M., and Scholl, E., "Strategy for Breeding Test Animals of High Susceptibility to Carcinogens," Eds. H.F. Stich and R.H.C. San, In: Short-Term Tests for Chemical Carcinogens, 399-407, Springer-Verlag, New York, Heidelberg, Berlin, 1981.

Arai, M., Aoki, Y., Nakanishi, K., Miyata, Y., Mori, T., and Ito, N., "Long-term Experiment of Maximal Non-Carcinogenic Dose of Dimethylnitrosamine for Carcinogenesis in Rats," Gann, 70:549-558, 1979.

Armitage, P., "Tests for Linear Trends in Proportions and Frequencies," Biometrics, 11:375-86, 1955.

Armitage, P. and Doll, R., "A Two-Stage Theory of Carcinogenesis in Relation to the Age Distribution of Human Cancer," Brit. J. Cancer, 11:161-169, 1961.

Armstrong, B.K., "Cancer After Hormone Therapy," Cancer Surveys, 1:625-652, 198

Baehner, R.L., Boger, L.A. and Ingraham, L.M., "Reduced Oxygen By-Products and White Blood Cells," Free Radicals in Biology V:91-113, 1981.

Baker, N., Mead, J., Jr., and Kannan, R., "Hepatic Contribution to Newly Made Fatty Acids in Adipose Tissue in Rats and Inhibition of Hepatic and Extrahepatic Lipogenesis From Glucose by Dietary Corn Oil," J. Lipids Res., 16:568-576, 1981.

Balmain, A. and Pragnell, I.B., "Mouse Skin Carcinomas Induced In Vivo by Chemical Carcinogens have a Transforming Harvey-ras Oncogene," Nature, 303:72-74, 1983.

Balmain, A., Ramsden, M., Bowden, G.T. and Smith, J., "Activation of the Mouse Cellular Harvey-ras Gene in Chemically Induced Benign Skin Papillomas," Nature, 307:658-660, 1984.

Baron, R.L., "Chronic Studies," In: The Principles and Methods in Modern Toxicology", Eds. C.D. Galli, S.D. Murphy, and R. Paoletti, Elsevier/North-Holland Biomedical Press, New York, p. 191-206, 1980.

Bazin, R., Lavau, M. and Herzog, J., "Pancreatic Lipase and Ketogenic Conditions," Biomedicine, 28:160-165, 1978.

Berenblum, "Carcinogenesis as a Biological Problem," Eds.: Nueberger A. and Tatum, E.L., North-Holland Publishing Company, Amsterdam, Oxford, 1974.

Berenblum and Shubik, "A New Quantitative Approach to the Study of Chemical Carcinogenesis in the Mouse's Skin," Br. J. Cancer, 1:383-391, 1947.

Bieri, J.G., "Second Report of the Ad Hoc Committee on Standards for Nutritional Studies," J. Nutr., 110:1726, 1980.

Bishop, J.M., "Cellular Oncogenes and Retroviruses," Ann. Rev. Biochem., 52:301-354, 1983.

Boden, G., Wilson, R.M., Essa-Koumar, N., and Owen, O.E., "Effects of a Protein Meal, Intraduodenal HCl, and Oleic Acid on Portal and Peripheral Venous Secretion and on Pancreatic Bicarbonate Secretion," Gut, 19:277-283, 1978.

Boehm, T.L.J. and Drahovsky, D., "Alteration of Enzymatic Methylation of DNA Cytosines by Chemical Carcinogens: A Mechanism Involved in the Initiation of Carcinogenesis," Journal of the National Cancer Institute, 71:429-433, 1983.

Bourdel, G., "Effect of Separate Feeding of Proteins and Lipids on Pancreatic Adaptation in the Rat," Am. J. Physiol., 244:125-130, 1983.

Boutwell, R.K., Some Biological Aspects of Skin Carcinogenesis, Progr. Exp. Tumor Res., 4:207-250, 1964.

Breslow, N.E., Day, N.E., Turusov, V.S. Tomatis, L., "Associations Between Tumor Types in a Large-scale Carcinogenesis Study of CF-1 Mice," J. Natl. Cancer Instit., 52:233-9, 1974.

Carlborg, F.W., "Dose-Response Functions in Carcinogenesis and the Weibull Model," Fd. Cosmet. Toxicol., 19:255-263, 1981.

Carr, B.I., Reilly, J.G. and Riggs, A.D., "5-Azacytidine: Promotion Activity for Rat Hepatocellular Carcinoma. In: Börszonyi, M., Day, N.E., Lapis, K., Yamasaki, H. (eds)," Models, Mechanisms and Etiology of Tumour Promotion, IARC Scientific Publication No. 56, International Agency for Research on Cancer, Lyon, in press, 1984.

Cerutti, P.A., Emerit, I. and Amstad, P., "Tumor Promoter Phorbol-myristate-acetate Induces Membrane Mediated Chromosomal Damage. In: Nygaard, O.F. and Simic, M.G. (eds.)," Radioprotectors and Anticarcinogens, New York, Academic Press, Inc., pp. 527-538, 1983.

Chan, P.C., Head, J.F., Cohen, L.A., and Wynder, E.L., "Influence of Dietary Fat on the Induction of Mammary Tumors by N-Nitrosomethylurea: Associated Hormone Changes and Differences Between Sprague-Dawley and F344 Rats," J. Natl Cancer Inst., 59: No. 4, 1977.

Clough, G., "Environmental Effects on Animals Used in Biomedical Research," Biol. Rev. 57:487-523, 1982.

Coates, M., Workshop on Laboratory Animal Nutrition, ICLAS Bulletin, International Council for Laboratory Animal Science, 50:4-11, 1982.

Cochran, W.G., "Some Methods of Strengthening the Common X^2 test," Biometrics, 10:417-51, 1954.

Copeland, E.S. (ed.), "A National Institutes of Health Workshop Report. Free Radicals in Promotion - A Chemical Pathology Study Section Workshop," Cancer Research, 43:5631-5637, 1983.

Cox, D.R., "The Analysis of Binary Data," London: Chapman and Hall, Chapters 4 and 5, 1970.

Cox, D.R. and McCullagh, P., "Some Aspects of Analysis of Covariance," Biometrics, 38:541-61, 1982.

Crump, K.S., Hoel, D.G., Langley, C.H., and Peto, R., "Fundamental Carcinogenic Processes and Their Implications for Low Dose Risk Assessment, Cancer Research 36:2973-79, 1976.

Day, N.E., "Time as Determinant of Risk in Cancer Epidemiology: The Role of Multi-Stage Models," Cancer Surveys, 2, (in press).

Day, N.E. and Brown, C.C., "Multistage Models and the Primary Prevention of Cancer," Journal National Cancer Institute, 64:977-989, 1980.

deBethizy, J.D., Sherrill, J.M., Rickert, D.E., Hamm, T.E., Jr., "Effects of Pectin-containing Diets on the Hepatic Macromolecular Covalent Binding of 2,6-Dinitro-[^{3}H]toluene in Fischer-344 Rats," Tox Appl. Pharm., 69, (in press).

DellaPorta, G. and T., Dragani, A., "Testing for Carcinogenicity," In: The Principles and Methods in Modern Toxicology, Eds. C.D. Galli, S.D. Murphy, and R. Paoletti, Elsevier/North-Holland Biomedical Press, New York, 1980, p. 159-168.

Demol, P. and Sarles, H., "Action of Fatty Acids on the Exocrine Pancreatic Secretion of the Conscious Rat: Further Evidence for a Protein Pancreatic Inhibitory Factor," J. Physiol. 275:27-37, 1978.

Dempster, A.P., Selwyn, M.D., and Weeks, B.J., "Combining Historical and Randomized Controls for Assessing Trends in Proportions," J. Amer Statist Assoc 78:221-7, 1983.

Deschodt-Lanckman, M., Robberecht, P., Camus, J. and Christophe, J., "Short-Term Adaptation of Pancreatic Hydrolases to Nutritional and Physiological Stimuli in Adult Rats," Biochimie, 53:789-796, 1971.

Dinse, G.E. and Lagakos, S.W., "Nonparametric Estimation of Lifetime and Disease Onset Distributions from Incomplete Observations," Biometrics, 1982.

Doerfler, W., "DNA Methylation and Gene Activity," Ann. Rev. Biochem., 52:93-124, 1983.

Doll, R. and Peto, R., "The Causes of Cancer Quantitative Estimates of Avoidable Risks of Cancer in the United States Today," Oxford, New York, Oxford University Press, 1981.

Druckrey, H., "Quantitative Aspects in Chemical Carcinogenesis," In: Truhaut R., ed. Potential Carcinogenic Hazards from Drugs (Evaluation of Risk). New York: Springer Verlag 1967: 60-78.

Eva, A. and Aaronson, S.A., "Frequent Activation of C-Kis as a Transforming Gene in Fibrosarcomas Induced by Methylcholanthrene," Science, 220:955-956, 1983.

Faichney, A., Chey, W.Y., Kim, Y.C., Lee, K.Y., Kim, M.S. and Chang, T.M., "Effect of Sodium Oleate on Plasma Secretin Concentration and Pancreatic Secretion in Dog," Gastroenterology, 81:458-462, 1981.

Farber, E., "Chemical Carcinogenesis. A Biologic Perspective," American Journal of Pathologists, 106:271-296, 1982.

FDA Advisory Committee on Protocols for Safety Evaluation: Panel on Carcinogenesis Report on Cancer Testing in the Safety Evaluation of Food Additives and Pesticides. Toxicol Appl Pharmacol, 20:419-38, 1971.

Fears, T.R., Tarone, R.E. and Chu, K.C., "False-positive and False Negative Rates for Carcinogenicity Screens," Cancer Research, 37:1941-49, 1977.

Feinberg, A.P. and Vogelstein, B., "Hypomethylation of Ras Oncogenes in Primary Human Cancer," Biochemical and Biophy. Res. Commun., 111:47-54, 1983

Fidler, I.J., "Depression of Macrophages in Mice Drinking Hyperchlorinated Water," Nature 270:735-736, 1977.

FDA, Toxicological Principles for the Safety Assessment of Direct Food Additives and Color Additives Used in Food, U.S. Food and Drug Administration: Bureau of Foods, Washington, DC, 1982.

Forman, L.P. and Schneeman, B.O., "Effects of Dietary Pectin and Fat on the Small Intestinal Contents and Exocrine Pancreas of Rats," J. Nutr. 110: 1992-1999, 1980.

Foulds, L., "Neoplastic Development, Vol. 1, 1969, Vol. 2, 1975, Acad. Press, New York.

Frisch, R.E., Hegsted, D.M., Yoshinaga, K., "Body Weight and Food Intake at Early Estrus of Rats on a High-Fat Diet," Proc. Nat. Acad. Sci., 72: 4172-4176, 1975.

Fürstenberger, G., Sorg, B. and Marks, F., "Tumor Promotion by Phorbol Esters in Skin: Evidence for a Memory Effect," Science, 220:89-91, 1983.

Gart, J.J., "The Comparison of Proportions: A Review of Significance Tests, Confidence Intervals and Adjustments for Stratification," Rev Int Statist Instit, 39:148-69, 1971.

Gart, J.J., Chu, K.C., Tarone, R.E., "Statistical Issues in Interpretation of Chronic Bioassay Tests for Carcinogenicity," J. Natl. Cancer Inst., 62:957-74, 1979.

Gateff, E., "Malignant Neoplasms of Genetic Origin in Drosophila Melanogaster," Science, 200:1448-1459, 1978.

Gidez, L.I., "Effect of Dietary Fat on Pancreatic Lipase Levels in the Rat," J. Lipid Res. 14:169-177, 1973.

Gilbert, C. Giehman, J. Loustalot, P., and Lutz, W., "The Modifying Influence of Diet and Physical Environment on Spontaneous Tumor Frequency in Rats," Br. J. Cancer, 12:565-93, 1958.

Goldsworthy, T., Campbell, H.A., and Pitot, H.C., "The Natural History and Dose-Response Characteristics of Enzyme-Altered Foci in Rat Liver Following Phenobarbital and Diethylnitrosamine Administration, Carcinogenesis, (submitted).

Goodman, D.G., Ward, J.M., Squire, R.A., Chu, K.C. and Linhart, M.S., "Neoplastic and Nonneoplastic Lesions in Aging F344 Rats," Toxicol. Appl. Pharmacol, 48:237-48, 1979.

Green, M.H. and Green, J.B., "Effects of Dietary Fat Saturation and Triglyceride and Cholesterol Load on Lipid Absorption in the Rat," Atherosclerosis, 46:181-194, Elsevier Scientific Publishers Ireland, Ltd., 1983.

Guess, H., Crump, K., and Peto, R., "Uncertainty Estimates for Low-Dose-Rate Extrapolations of Animal Carcinogenicity Data," Cancer Res., 37: 3475-3483, 1977.

Hall, J.E., White, W.J. and Lang, L.M., "Acidification of Drinking Water: Its Effects on Selected Biologic Phenomena in Male Mice," Lab. An. Sci., 30:643-651, 1980.

Hamm, T.E., Jr., "Protection of the Animal Environment," in Proceedings of a Workshop on the Optimal Use of Facilities for Carcinogenicity/Toxicity Testing, NTP/NCI Carcinogenesis Testing Program, Boiling Springs, PA, 1980.

Hamm, T.E., Jr., Raynor, T. and Caviston, T., "Unsuitability of the AIN-76A Diet for Male F-344 and CD Rats and Improvement by Substituting Starch for Sucrose," Laboratory Animal Science 32:414-415, 1982.

Haran, N. and Berenblum, I., "The Induction of the Initiating Phase of Skin Carcinogenesis in the Mouse by Oral Administration of Urethane (Ethyl Carbamate)," Brit. J. Cancer, 10:57-60, 1956.

Haseman, J.K., "A Reexamination of False-positive Rates for Carcinogenesis Studies," Fund Appl Toxicol, 3:334-339, 1983.

Haseman, J.K., "Patterns of Tumor Incidence in Two-year Cancer Bioassay Feeding Studies in Fischer 344 Rats," Fund Appl Toxicol, 3:1-9, 1983.

Haseman, J.K., Response to, "Use of Statistics When Examining Lifetime Studies in Rodents to Detect Carcinogenicity," J. Toxicol. Environ. Health, 3:633-36, 1977.

Haseman, J.K., Huff, J. and Boorman, G.A., "Use of Historical Control Data in Carcinogenicity Studies in Rodents." Unpublished manuscript, 1983.

Hecker, E., Fusenig, N.E., Kunz, W., Marks, F. and Thielmann, H.W., (eds.), "Cocarcinogenesis and Biological Effects of Tumor Promoters," Raven Press, New York, 1982.

Herzberg, G.R. and Rogerson, M. "Role for Fatty Acid Binding Protein in the Regulation of Hepatic Lipogenesis by Dietary Linoleic Acid," Nutr. Res., 1:601-607, 1981.

Hicks, R.M., Wakefield, J.S.J., and Chowaniec, J., "Evaluation of a New Model to Detect Bladder Carcinogens or Co-carcinogens: Results Obtained With Saccharin, Cyclamate and Cyclophosphamide," Chem.-Biol. Interactions, 11:225-233, 1975.

Hirono, I. and Shibuya, C., "Induction of Stomach Cancer by a Single Dose of N-Methyl-N'-Nitro-N-Nitrosoguanidine Through a Stomach Tube," In: Topics in Chemical Carcinogenesis, Eds., W. Nakahara, S. Takayama, T. Sugimura and S. Odashima, University of Tokyo Press, Japan, p. 121-132, 1972.

Hoel, D.G., "Conditional Two Sample Tests With Historical Controls. Contributions to Statistics Essays in Honor of Norman L. Johnson," edited by P.K. Sen, North Holland, 229-236, 1983.

Hoel, D.G. and Walburg, H.E., "Statistical Analysis of Survival Experiments," J. Natl. Cancer Instit., 49:361-72, 1972.

Hollstein, M., McCann, J., Angelosanto, F.A. and Nichols, W.W., "Short-Term Tests for Carcinogens and Mutagens," Mutation Res., 65:133-226, 1979.

Homburger, F. and Tregier, A., "Modifiers of Carcinogenesis," Progress in Experimental Tumor Research, 11:86-99, S. Karger, NY, 1969.

Hueper, W.C., "The Skin as an Assay System for Potential Carcinogens," National Cancer Institute Monograph, No. 10, pp. 577-610, presented at the Conference on Biology of Cutaneous Cancer, Philadelphia, Penn., April 6-11, 1962.

IARC, "Long-term and Short-term Screening Assays for Carcinogens: A Critical Appraisal," Supplement 2, International Agency for Research on Cancer, Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Man, Lyon, 1980.

IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans, "Chemicals, Industrial Processes and Industries Associated with Cancer in Humans," V. 1 to 29, Supplement 4, International Agency for Research on Cancer, Lyon, 1982.

IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans, "Some food Additives, Feed Additives and Naturally Occurring Substances," V. 31, International Agency for Research on Cancer, Lyon, 1983.

IARC/IPCS/CEC Working Group Report, "Approaches to Classifying Carcinogens According to Mechanism of Action," IARC Internal Technical Report, 83/1, International Agency for Research on Cancer, 1983.

ILAR, "International Standardized Nomenclature for Outbred Stocks of Laboratory Animals," Institute of Laboratory Animal Resources, National Research Council, Washington, DC, 1972.

ILAR, "Long-Term Holding of Laboratory Rodents," National Research Council, Washington, DC, 1976.

ILAR, "Laboratory Animal Housing," National Academy of Sciences, Washington, D.C., 1978.

Interagency Regulatory Liaison Group, "Scientific Bases for Identification of Potential Carcinogens and Estimation of Risks," J. Natl. Cancer Instit., 63:241-68, 1979.

Ip, C., Yip, P., and Bernardis, L.L., "Role of Prolactin in the Promotion of Dimethylbenz(a)Anthracine-Induced Mammary Tumors by Dietary Fat," Cancer Res., 40:374-378, 1980.

Iritani, N., and Ikeda, Y., "Activation of Catalase and Other Enzymes by Corn Oil Intake," J. Nutr. 112:2235-2239, 1982.

Iritani, N., Fukuda, E., and Kitamura, Y., "Effect of Corn Oil Feeding on Lipid Peroxidation in Rats," J. Nutr., 110:924-930, 1980.

Irwin, M.H., Steenbock, H., Templin, V.M., "A Technique for Determining the Rate of Absorption of Fats," J. Nutr., 12:85-101, 1936.

Jacobs, L.R., "Effect of Short-Term Dietary Fat on Cell Growth in Rat Gastrointestinal Mucosa and Pancreas," Amer. J. Clin. Nutr., 37:361-367, 1983.

King, M.M. and McCay, P.B., "Possibilities for Dietary Fat and Antioxidants as Modulators of Mammary Carcinogenesis," J. Am. Coll. Toxicol., 2:285-306, 1983.

Kinsella, A.R. and Radman, M., "Tumor Promoter Induces Sister Chromatid Exchanges: Relevance to Mechanisms of Carcinogenesis," Proceedings of the National Academy of Sciences, 75:6149-6153, 1978.

Kodell, R.L., Farmer, J.H., Gaylor, D.W. and Cameros, A.M., "Influence of Cause-of-death Assignment on Time-to-tumor Analysis in Animal Carcinogenesis Studies," J. Natl. Cancer Instit., 69:659-64, 1982.

Kodell, R.L. and Nelson, C.J., "An Illness-death Model for the Study of the Carcinogenesis Process using Survival/sacrifice Data," Biometrics, 36:267-77, 1980.

Kraft, P.L., and Bieber, M.A. "The Use of Dietary Fats in Animal Studies," Spring Symposium of the Mid-Atlantic Chapter of the Society of Toxicology, Wilmington, Delaware, May 21, 1983.

Knapka, J.J., Smith, K.P., and Judge, F.J., "Effects of Open and Closed Formula Rations on the Performance of Three Strains of Laboratory Mice," Lab. Animal Sci., 24:480-487, 1974.

Krewski, D., Kovar, J. and Bickis, M., "Optimal Experimental Designs for Low Dose Extrapolation," Topics in Applied Statistics edited by T.W. Dwived: Marcell Dekker, New York, 1983.

Lagakos, S. and Mosteller, F., "A Case Study of Statistics in the Regulatory Process: the FD&C Red No. 40 Experiments," J Natl Cancer Instit, 66:197-212, 1981.

Lambert, M., Gattoni-Celli, S., Kitschmeier, P. and Weinstein, I.B., "Benzo(a)pyrene Induction of Extrachromosomal Viral DNA Synthesis in Rat Cells Transformed by Polyoma Virus," Carcinogenesis, 4:587-594, 1983.

Land, H., Parad, L.F. and Weinberg, R.A., "Tumorigenic Conversion of Primary Embryo Fibroblasts Requires at Least Two Cooperating Oncogenes," Nature, 304:596-602, 1983.

Lavau, M., Bazin, R., and Herzog, J., "Comparative Effects of Oral and Parenteral Feeding on Pancreatic Enzymes in the Rat," J. Nutr., 104:1432-1437, 1974.

Lavi, S., "Carcinogen-Mediated Amplification of Specific DNA Sequences," Journal of Cellular Biochemistry, 18:149-156, 1982.

Lee, S.Y., and Rogers, A.E., "Dimethylbenzanthracene Mammary Tumorigenesis in Sprague-Dawley Rats Fed Diets Differing in Content of Beef Tallow or Rapeseed Oil," Nutr. Res., 3:361-371, 1983.

Liddell, D., "Practical Tests of the 2x2 Contingency Tables," Statistician, 23:295-304, 1978.

McCay, P.B., King, M., Rikans, L. and Pitha, J., "Interactions Between Dietary Fats and Antioxidants on DMBA-Induced Mammary Carcinomas and on AAF-Induced Hyperplastic Nodules and Hepatomas," J. Environ. Pathol. Toxicol. 3:451, 1980.

McKinney, J.D., Maurer, R.R., Hass, J.R. and Thomas, R.O., "Possible Factors in the Drinking Water of Laboratory Animals Causing Reproductive Failure," Ed. K., Lawrence, Identification and Analysis of Organic Pollutants in Water, Ann Arbor Science, Michigan, USA, pp. 417-432.

McKnight, B. and Crowley, J.J., "Methods for Testing for Differences in Tumor Incidence Based on Animal Carcinogenesis Experiments," J. Amer Statist Assoc (in press), 1984.

McCormick, D.L., Ronan, S.S., Becci, P.J., and Moon, R.C., "Influence of Total Dose and Dose Schedule on Induction of Urinary Bladder Cancer in the Mouse by N-butyl-N-(4-hydroxy-butyl)nitrosamine," Carcinogenesis, 2: 251-254, 1981.

Mainz, D.L., Black, O., and Webster, P.D., "Hormonal Control of Pancreatic Growth," J. Clin. Invest., 52:2300-2304, 1973.

Mantel, N., "Assessing Laboratory Evidence for Neoplastic Activity," Biometrics, 36:389-99, 1980.

Marshall, E., "EPA's High Risk Carcinogen Policy," Science, 218:975-978, 1982.

Medinsky, M.A., Popp, J.A., Hamm, T.E. and Dent, J.G., "Development of Hepatic Lesions in Male Fischer-344 Rats Fed AIN-76A Purified Diet," Tox. Appl. Pharm., 62:111-120, 1982.

Meyer, J.H. and Jones, R.S., "Canine Pancreatic Responses to Intestinally Perfused Fat and Products of Fat Digestion," Amer. J. Physiol., 226: 178-1187, 1974.

Miller, T.A., Konturek, S.J., Llanos, O.L., Rayford, P.L., and Thompson, J.C., "The Effect of Fat on Secretin Release," Ann. Surg. 187:303-308, 1978.

Mitchell, T.J. and Turnbull, B.W., "Log-linear Models in the Analysis of Disease Prevalence Data from Survival/sacrifice Experiments," Biometrics, 35:221-35, 1979.

Montesano, R. and Slaga, T.J., "Initiation and Promotion in Carcinogenesis: An Appraisal," Cancer Surveys, 2, in press, 1984.

Munro, I.C. and Krewski, D., "Risk Assessment and Regulatory Decision Making," Food and Cosmetics Toxicology, 19:549-560, 1981.

National Cancer Institute, Guidelines for Carcinogen Bioassay in Small Rodents, NCI-CG-TR-1, 1976.

Newberne, P.M., Bieri, J.G., Briggs, G.M. and Nesheim, M.C., "Control of Diets in Laboratory Animal Experimentation," A Report of the Committee on Laboratory Animal Diets, Institute of Laboratory Animal Resources, Assembly of Life Sciences, National Research Council, National Academy of Sciences, Washington, DC, 1978.

Newberne, P.M., Weigert, J. and Kula, N., "Effects of Dietary Fat on Hepatic Mixed-Function Oxidases and Hepatocellular Carcinoma Induced by Aflatoxin B_1 in Rats," Cancer Res., 39:3986-91, 1979.

Newbold, R.F, and Overell, R.W., "Fibroblast Immortality is A Prerequisite for Transformation by EJ c-Ha-ras Oncogene," Nature, 304:468-651, 1983.

Nguyen, H.T. and Woodard, T.C., "Intranephronic Calcinosis in Rats," Am. J. Path., 100:39-55, 1980.

Notario, V., Sukumar, S., Santos, E. and Barbacid, M., "A Common Mechanism for the Malignant Activation of Ras Oncogenes in Human Neoplasia and in Chemically-Induced Animal Tumors. In: Vande Woode, G.F. et al. (eds): Cancer Cells 2/ Oncogenes and Viral Genes, 1984.

NTP/TR-210, "Carcinogenesis Bioassay of 1,2-dibromoethane in F344 Rats and $B6C3F_1$ Mice (inhalation study)," National Toxicology Program, Technical Report No. 210, 1982.

OECD, Guidelines for Testing Chemicals," Publ. #ISBN92-64-12221-4. Organization for Economic Cooperation and Development, Paris, France, 1981.

Organization for Economic Cooperation and Development, OECD Guidelines for Testing of Chemicals, Guidelines 451, 452 and 453, 1981.

Pearson, E.S. and Hartley, H.O., "Biometrika Tables for Statisticians," Third Edition Cambridge University Press, 71-7, 1966.

Pershagen, G., Nordberg, G. and Bjorklung, N.-E., "Carcinomas of the Respiratory Tract in Hamsters Given Arsenic Trioxide and/or Benzo(a)pyrene by the Pulmonary Rote," Environmental Research, (in press).

Peterson, A., "Bounds for a Joint Distribution Function with Fixed Sub-distribution Functions: Applications to Completing Risks," Proc. Natl. Acad Sci. USA, 73:11-13, 1976.

Peto, R., "Epidemiology, Multistage Models and Short-term Mutagenicity Tests". In: Hiatt, H.H., Watson, J.D., and Winsten, J.A. (Eds.): Origins of Human Cancer. Cold Spring Harbor Laboratory, 1403-1428, 1977.

Peto, R., "Guidelines on the Analysis of Tumor Rates and Death Rates in Experimental Animals (editoral)," Br. J. Cancer, 29:101-3, 1974.

Peto, R. and Lee, P, "Weibull Distributions for Continuous Carcinogenesis Experiments," Biometrics, 29:457-70, 1973.

Peto, R., Gray, R., Brantom, P., and Grasso, P., "Effects on Two Tones of Inbred Rats of Chronic Ingestion of Diethyl- or Dimethyl-nitrosamine: An Unusually Detailed Dose-Response Study," Presented at Banbury Conference - Possible Role of Nitrosamines," In: Human Cancer, April 4-7, 1982.

Peto, R., Pike, M.C., Day, N.E., et al., "Guidelines for Simple, Sensitive, Significance Tests for Carcinogenesis Effects in Long-term Animal Experiments. Long-term and Short-term Screening Assays for Carcinogenesis: A Critical Appraisal," IARC Monogr, 2(suppl):311-426, 1980.

Peto, R., Roe, F.C., Lee P., Levy, L. and Clark, J., "Cancer and Aging in Mice and Men," Br. J. Cancer, 32:411-26, 1975.

Peto, J., Seidman, H., and Selikoff, I.J., "Mesothelioma Mortality in Asbestos Workers: Implications for Models Carcinogenesis and Risk Assessments," Brit. J. Cancer, 45:124, 1982.

Pike, M.C., "A Method of Analysis of a Certain Class of Experiments in Carcinogenesis," Biometrics, 22:142-61, 1966.

Pitot, H.C., "Contribution of Our Understanding of the Natural History of Neoplastic Development in Lower Animals to the Cause and Control of Human Cancer," Cancer Survey, 2: (in press).

Portier, C. and Hoel, D., "Low-dose-rate Extrapolation Using the Multistage Model," Biometrics, 39:897-906, 1983a.

Portier, C. and Hoel, D., "Design of Animal Carcinogenicity Studies for Goodness-of-fit of Multistage Models," Fundamental and Applied Toxicology, (in press).

Portier, C., and Hoel, D., "Optimal Design of the Chronic Animal Bioassay," J. Tox. Env. Health, 12:1-19, 1983b.

Purchase, I.F.H., "Inter-Species Comparisions of Carcinogenicity," Br. J. Cancer, 41:454-468, 1980.

Pye, V.I., and Patrick, R., "Ground Water Contamination in the United States," Science, 21:713-718, 1983, 1:183-197, 1977.

Ramaley, J.A., "Puberty Onset in Males and Females Fed a High Fat Diet (41062)," Proc. Soc. Exp. Biol. and Med., 166:294-296, 1981.

Raynor, T.H., White, E.L., Cheplen, J.M., Sherill, J.M., and Hamm, T.E., Jr., "An Evaluation of a Water Purification System for Use in Animal Facilities," Lab. Animals, (in press).

Raynor, T.H., Steinhagen, W.H., and Hamm, T.E., Jr., "Differences in the Microenvironment of a Polycarbonate Caging System: Bedding vs. Raised Wire Floors," Lab. Animals, 17:85-89, 1983.

Reddy, B.S., Druckrey, H. , Cohen, L.A., McCoy, G.D., Hill, P., Weisburger, J.H., and Wynder, E.L., "Nutrition and its Relationship to Cancer," Adv. Cancer Res., 32:237-345, 1980.

Riggs, A.D. and Jones, P.A., "5-Methylcytosine, Gene Regulation and Cancer," Advances in Cancer Research, 40:1-30, 1983.

Roe, F.J.C., "Testing in vivo for General Chronic Toxicity and Carcinogenicity," In: Testing for Toxicity, Ed. J.W. Gorrod, Taylor and Francis, Ltd., London, p. 29-43, 1981.

Roe, F.J.C. and Tucker, M.J., "Recent Developments in the Design of Carcinogenicity Tests on Laboratory Animals," Proc Europ Soc Studies Drug Toxicity, 13:171-7, 1978.

Roebuck, B.D., Yager, J.D., Jr., and Longnecker, D.S., "Dietary Modulation of Azaserine-Induced Pancreatic Carcinogenesis in the Rat," Cancer Res., 41:888-893, 1981.

Roebuck, B.D., Yager, J.D., Jr., and Longnecker, D.S., and Wilpone, S.A., "Promotion by Unsaturated Fat of Azaserine-Induced Pancreatic Carcinogenesis in the Rat," Cancer Res., 41:3961-3966, 1981.

Rogers, A.E., and Newberne, P.W., "Lipotrope Deficiency in Experimental Carcinogenesis," Nutr. and Cancer, 2:104, 1980.

Ross, M.H., Lustbader, E.D., and Bras, G., "Dietary Practices of Early Life and Spontaneous Tumors of the Rat," Nutr. and Cancer, 3:150-167, 1982.

Ross, M.H. and Bras, G., "Lasting Influence of Early Caloric Restriction on Prevalence of Neoplasms in the Rat," J. Natl Cancer Instit, 47:1095-1113, 1971.

Rudnay, P. and Borzsonyi, M., "The Tumorigenic Effect of Treatment with Arsenic Trioxide," Magyar Onkologia, 25:73-77, 1981.

Ruley, H.E., "Adenovirus Early Region 1A Enables Viral and Cellular Transforming Genes to Transform Primary Cells in Culture," Nature, 304:602-606, 1983.

Salsburg, D.S., "The Lifetime Feeding Study in Mice and Rats - An Examination of its Validity as a Bioassay for Human Carcinogenesis," Fund Appl. Toxicol., 3:63-7, 1983.

Salsburg, D.S., "Use of Statistics when Examining Lifetime Studies in Rodents to Detect Carcinogenecity," J. Toxicol. Environ. Health," 3:611-28, 1977.

Saraux, B., Girard-Globa, A., Ouagued, M., and Vacher, D., "Response of the Exocrine Pancreas to Quantitative and Qualitative Variations in Dietary Lipids," Am. J. Physiol., 243:10-15, 1982.

Schartl, M., Barnekow, A., Bauer, H., and Anders, F., "Correlations of Inheritance and Expression Between a Tumor Gene and the Cellular Homolog of the Rous Sarcoma Virus-Transforming Gene in Xiphophorus," Cancer Res., 42:4222-4227, 1982.

Scherer, E., and Emmelot, P., "Kinetics of Induction and Growth of Precancerous Liver-Cell Foci, and Liver Tumour Formation by Diethylnitrosamine in the Rat," Europ. J. Cancer, 11:689-696, 1975.

Schwetz, B.A., "Prechronic Toxicological Tests - Scope and Use," National Institute of Environmental Health Sciences, National Toxicology Program.

Shubik, P., and Clayson, D.B., "Application of the Results of Carcinogen Bioassays to Man," Inserm., 52:241-252, 1976.

Solleveld, H.S., Haseman, J. and McConnell, E.E., "The Natural History of Body Weight Gain, Survival and Neoplasia in the Fischer F344 Rat," J Natl Cancer Instit, 72:929-940, 1984.

Sontag, J.M., "Aspects in Carcinogen Bioassay," Origins of Human Cancer, Eds. H. Hiatt, J. Watson, and J. Winsten, pp. 1327-1338, 1977.

Sontag, J.M., Page, N.P., and Saffiotti, U., Guidelines for Carcinogen Bioassay in Small Rodents, National Institutes of Health, DHEW Publ. No. (NIH) 76-801, Washington, DC, 1976.

Squire, R.A., "Ranking Animal Carcinogens: A Proposed Regulatory Approach," Science, 214:877-880, 1981.

Sukumar, S., Notario, V., Martin-Zanca, D., and Barbacid, M., "Induction of Mammary Carcinomas in Rats by Nitroso-Methyl-Urea Involves the Malignant Activation of the H-ras 1 Locus by Single Point Mutations," Nature, 306:658-661, 1983.

Sukumar, S., Pulciani, S., Doniger, J., DiPaolo, J.A., Evans, C.H., Zbar, B., and Barbacid, M., "Transforming ras Gene in Tumorigenic Guinea Pig Cell Lines Initiated by Diverse Chemical Carcinogens," Science, (in press).

Symons, J.M., Bellar, J.A., Carswell, J.K. Kropp, K.L., Roebuck, G.G., Seeger, D.R., Slocum, C.J., Smith, B.L. and Stevens, A.A., National Organics Reconnaisance Survey for Halogenated Organics in Drinking Water, J. Am. Water Works Assoc., 67:634-637, 1975.

Tannenbaum, A. and Silverstone, H., "Nutrition and the Genesis of Cancer," In: Raven RW, Ed. Cancer, London, 1957, Butterworth: I:306-34.

Tarone, R.E., "The Use of Historical Control Information in Testing for a Trend in a Proportion," Biometrics, 38:215-20, 1982.

Tarone, R.E., Chu, K.C. and Ward, J.M., "Variability in the Rates of Some Common Naturally Occurring Tumors is Fischer 344 Rats and (C57B1/6N x C3H/H3N) F_1 ($B6C3f_1$) mice," J. Natl. Cancer Instit., 66:1175-81, 1981.

Tarone, R.E. and Gart, J.J., "On the Robustness of Combined Tests for Trends in Proportions," J. Amer. Statist. Assoc., 75:110-16, 1980.

Tomatis, L., "The Predictive Value of Rodent Carcinogenicity Tests in the Evaluation of Human Risk," Ann. Rev. Pharmacol. Toxicol., 19:511-30, 1979.

Tomatis, L., Partensky, C., and Montesano, R., "The Predictive Value of Mouse Liver Tumour Induction in Carcinogenicity Testing - A Literature Survey," Int. J. Cancer, 12:1-20, 1973.

TOSCA, "Proposed Health Effect Tests Standards for Toxic Substances Control Act Test Rules," Fed. Register, 44(91):27334-27374, 1979.

Tsiatis, A., "A Nonidentifiability Aspect of the Problem of Competing Risks," Proc. Natl. Acad. Sci., USA, 72:20-22, 1975.

Tucker, M.J., "The Effect of Long Term Food Restriction on Tumors in Rodents," Int. J. Cancer 23:803-7, 1979.

Turnbull, B.W. and Mitchell, T.J., "Explanatory Analysis of Disease Prevalence Data from Survival/sacrifice Experiments," Biometrics, 34:555-70, 1978.

USDHEW, Guide for the Care and Use of Laboratory Animals, U.S. Department of Health, Education and Welfare Publication No. (NIH) 78-23, ILAR, National Research Council, Washington, DC, 1978.

U.S. Food and Drug Administration, Bureau of Foods, "Toxicologic Principles for the Safety Assessment of Direct Food Additives and Color Additives Used in Food," pp. 34, 72, 1982.

US EPA, Pesticide Assessment Guidelines, Subdivision F Hazard Evaluation: Human and Domestic Animals, Report PB83-153916, p. 109, Nov. 1982.

US EPA, Health Effects Test Guidelines PB82-232984, Section HG-Chronic, HG-Chronic-Onco, HG-Chronic-Combines, August 1982.

US EPA, Health Effects Test Guidelines, EPA 560/6-82-001. U.S. Environmental Protection Agency, Washington, DC, 1982.

Varshavsky, A., "Phorbol Ester Dramatically Increases Incidence of Methotrexate-Resistant Mouse Cells: Possible Mechanisms and Relevance to Tumor Promotion," Cell, 25:561-572, 1981.

Vesselinovitch, et al., 1979 (to be supplied).

Wagner, D.A., Naylor, P.H., Kim, U., Shea, W., Ip, C., and Ip, M.M., "Interaction of Dietary Fat and the Thymus in the Induction of Mammary Tumors By 7, 12-Dimethylbenz(a)Anthracene," Cancer Res., 42:1266-1273, 1982.

Wahrendorf, J., "Simultaneous Analysis of Different Tumor Types in a Long-term Carcinogenicity Study with Scheduled Sacrifices," J. Natl. Cancer Instit., 70:915-21, 1983.

Ward, J.M., Goodman, D.G., Squire, R.A., Chu, K.C. and Linhart, M.S., "Neoplastic and Nonneoplastic Lesions in Aging (C57BL/6N x C3H/HeN)F_1 ($B6C3F_1$) Mice," J. Natl. Cancer Instit., 63:849-54, 1979.

Watanabe, P.G., and Gehring, P.J., "Dose-Dependent Fate of Vinyl Chloride and its Possible Relationship to Oncogenicity in Rats," Environmental Health Perspectives, 17:145-152, 1976.

Weinberg, R.A., "Fewer and Fewer Oncogenes," Cell, 30:3-4, 1982.

Weinberg, R.A., "A Molecular Basis of Cancer," Scientific American, 249:126-142, 1983.

Weinstein, I.B., "Current Concepts and Controversies in Chemical Carcinogenesis," J. of Supramol. Structure and Cellular Biochem., 17:99-120, 1981.

Weinstein, I.B., "Protein Kinase, Phospholipid and Control of Growth," Nature, 302:750, 1983.

Weinstein, I.B., "Letter to The Editor," Science, 219:794-796, 1983.

Weinstein, I.B., Gattoni-Celli, S., Kirschmeier, P., Lambert, M., Hsiao, W., Backer, J., and Jeffrey, A., "Molecular Mechanisms in Multistage Chemical Carcinogenesis." In: Greim, H., Jung, R., Kramer, M., Marquardt, H., and Oesch, F. (eds.). Biochemical Basis of Chemical Carcinogenesis, Raven Press, New York, 1984.

Weisburger, J.H. and Williams, G.M., "The Decision-Point Approach for Systematic Carcinogen Testing," Fd. Cosmet. Toxicol., 19:561-566, 1981.

Weisburger, J.H. and Williams, G.M., "Carcinogen Bioassay in Chemical Carcinogens," Ed. C. Searle, American Chem. Soc., 1983, (in press).

Weisburger, J.H., and Weisburger, E.K., "Tests for Chemical Carcinogens" Methods in Cancer Res., 1:307-399, 1967.

Wetsel, W.C., Rogers, A.E., and Newberne, P.M., "Dietary Fat and DMBA Mammary Carcinogenesis in Rats," Cancer Det. and Prev. 4:535-543, 1981.

WHO Scientific Group, "Report of the WHO Scientific Group on Principles for the Testing and Evaluation of Drugs for Carcinogenicity," WHO Tech. Rep. Ser., 426:1-26, 1969.

Wilbourn, J., Haroun, L., Heseltine, E., Kaldor, J., Montesano, R., Vainio, H., "How Do Animals Respond to Human Carcinogens," (in preparation).

Williams, G.M., "Classification of Genotoxic and Epigenetic Hepatocarcinogens Using Liver Culture Assays," Annals of the New York Academy of Sciences, 349:273-282, 1980.

Williams, G.M., "Epigenetic Effect of Liver Tumor Promoters and Implications for Health Effect," Environ. Health Perspectives, 50:177-183, 1983.

Williams, G.M., Furukawa, K., Maeura, Y. and Croci, T., "Induction of Hepatic Enzymes by Corn Oil Gavage," Science, 1983, (in press).

Wilson, V.L. and Jones, P.A., "Inhibition of DNA Methylation by Chemical Carcinogens In Vitro," Cell, 32:239-246, 1983.

Wise, A., "Interaction of Diet and Toxicity - The Future Role of Purified Diet in Toxicological Research," Arch. Toxicol., 50:287-299, 1982.

Yamasaki, H. and Weinstein, I.B., "Cellular and Molecular Mechanisms of Tumor Promotion and Their Implications with Respect to Risk Assessment," SGOMSEC Workshop on "Quantitative Estimation of Risk to Human Health from Chemicals," 1984.

Young, S.S., "Analysis and Interpretation of Long Term Rodent Studies with Attendant Problems of Risk Assessment." In: Symposium on Safety Evaluation of Drugs and Chemicals. Hemisphere Press (in press).

Appendix I

In the dietary history of the laboratory rat, cereal-based diets and purified diets were devised that provided nutritional adequacy. Several oils and fats have been used to provide the essential fatty acids (EFA). Corn oil was selected as a model fat likely because it was available in pure form and inexpensive. From the chemical and nutritional standpoint, fats and oils vary with respect to saturation, (mono-unsaturation, polyunsaturation) "essential" fatty acid composition, cis-trans configurations, potential rancidification, vitamin (especially tocopherol) content and the presence of added anti-oxidants.

It is well established that dietary vegetable oils used by humans are easily digested by rats; the coefficient of digestibility is from 94 to 99 percent. In 1936, Irwin et al. reported that, while various oils had significantly different rates of absorption when given by gavage to 48-hour fasted rats, absorption in all cases was virtually complete within 8 to 12 hours. More than half of a 1.5 ml dose of corn oil was absorbed within 2 to 4 hours. This fact has toxicological significance because it means that test substances dissolved in the corn oil may reach peak blood levels within this time period, although other factors may inhibit or enhance absorption. This effect could modify the pharmacokinetics of the test substance significantly. This specific aspect of oil gavage of test chemicals has received very little study.

The essential fatty acid requirements of rats and other rodents are in the range of 1% by weight of the diet, or about 2-4% of calories (NAS-NRC, 1982). These data are available primarily from rat studies; it is assumed that mice respond in a similar way. In the administration of test substances by gavage, rats have been given as much as 5-10 milliliters of corn oil per kilogram of body weight. Administration of 5 ml/kg raises the lipid intake, expressed as percent by weight of diet, from about 4-5% (the content in most natural ingredient diets, such as NIH-07 or Purina Laboratory Chow^{R}) to about 15%. Rats

given that amount of fat ingest about 30% of calories as fat compared to ingestion of about 10% of calories as fat in the natural ingredient diets (Kraft & Bieber, 1983).

A diet containing 15% fat by weight, or 30% of the total calories, is a high fat diet for rodents. The nutritional effects of the high fat intake may be expected to include: 1) decreased nutrient intake to compensate for calories derived from the administered lipid; 2) altered absorption of fat-soluble vitamins and other fat-soluble compounds because of their altered solubility in the bile acid micelles in the gut lumen; 3) increased selenium, Vitamin E, and lipotrope requirements for metabolism of the load of fat, especially the polyunsaturated fatty acids (PUFA).

The biological significance of the nutritional effects induced by a change in fat intake of this magnitude is not clear and has not been adequately studied. Diets fed to laboratory animals generally are designed to contain all nutrients in amounts greater than the minimal requirements. Nutrient content varies considerably and may be only borderline or even deficient, depending upon the source and composition of ingredients and conditions of mixing and storing. Prolonged ingestion of a diet borderline in certian nutrients and high in fat may cause significant alterations in the animals' nutritional and metabolic status. In particular, the effect of the PUFA present in vegetable oils should be considered because they can differ significantly from more saturated fats (Green & Green, 1983) with regard to digestibility and absorption.

Participants in the 1981 laboratory workshop on animal nutrition at the XIIth International Congress of Nutrition concluded that animal diet formulation has not been adequately considered in regard to the unique nutrient needs of animals used in long-term carcinogenicity and toxicity testing. Specially formulated diets are needed to minimize the number of naturally occuring tumors (Ross et al., 1982).

Body weights of rats receiving corn oil gavage as vehicle controls in NTP bioassays were higher than untreated controls by about 20 grams, a difference of 5% or less. This indicates that rats given the additional caloric load of fat partially compensated for the calories by decreasing their non-fat ad libitum food intake.

The physiological and biochemical effects of increased polyunsaturated lipid intake may be expected to include: 1) stimulation of pancreatic enzyme and bicarbonate secretion; followed by 2) increased pancreatic enzyme synthesis; and 3) increased pH in the upper small bowel; 4) alteration of the quality and quantity of bile acid secretion; 5) alteration of gastrointestinal motility; 6) increased prostaglandin synthesis in the gut mucosa; 7) increased peroxidation of tissue lipids and tissue superoxide and peroxide content (Iritani et al., 1980 a,b); 8) altered fatty acid composition of tissues and consequent changes in membrane fluidity; 9) altered activity of hepatic and other microsomal oxidases; 10) decreased hepatic and adipose tissue lipogenesis (Baker et al., 1981; Herzberg & Rogerson, 1981).

Pancreatic and biliary secretions are influenced by intralumenal lipid and the resulting effects on digestion and absorption have been studied. In rats, pancreatic secretions are influenced by trypsin and possibly other enzymes in the duodenal contents and by the presence of fat, carbohydrate and protein in the small bowel (Boden et al., 1978; Deschodt-Lanckman et al., 1971; Faichney et al., 1981; Gidez, 1973). Vegetable oils are highly effective inducers of pancreatic and intestinal lipase, and suppress intestinal and pancreatic amylase and trypsin (Jacobs, 1983; Deschodt-Lanckman et al., 1971). Pancreatic lipase activity can be stimulated up to 200 times basal activity when a high fat diet is fed to a rat (Bazin et al., 1978; Deschodt-Lanckman et al., 1971; Jacobs, 1983; Lavau et al., 1974). Stimulation of pancreatic secretion by corn oil in dogs requires digestion of the added oil (Meyer & Jones, 1974).

In contrast to the effects of the polyunsaturated oils, oleic acid and olive oil decreased total pancreatic protein (enzyme) output while increasing pancreatic secretion volume in rats (Demol & Sarles, 1978). This effect of olive oil is dependent on the presence of pancreatic juice, presumably on its content of lipase. In a study of effects of lard (which contains about 40% oleic acid) on pancreatic enzyme content, increasing lipase and colipase, and decreasing amylase, concentrations were found as the amount of lard in the diet increased from 12 to 45%. Colipase, but not the other two enzymes, correlated also with protein intake (Saraux et al., 1982). Protein concentration in the diet and duodenum appears to be a significant factor in determining how lipids affect pancreatic secretion (Bourdel, 1983).

One might expect hypertrophy or hyperplasia of the acinar cells of the pancreas in response to an increased demand for enzyme output. Although histologic studies were not performed along with enzymatic measurements, neither pancreatic weight (Forman & Schneeman, 1980) nor DNA synthesis (Jacobs, 1983) was increased after several weeks of intake of diets high in corn oil. However, other studies using oleic acid administered as sodium oleate showed enhanced secretion of three gastrointestinal hormones: gastrin, secretin and cholecystokinin (Boden et al., 1978; Faichney et al., 1981; Miller et al., 1978). These gastrointestinal hormones can also regulate pancreatic cell growth. Mainz et al. (1973) showed that cholecystokinin and pancreozymin administration to rats resulted in pancreatic hypertrophy and hyperplasia. Another variable which can dramatically affect pancreatic secretion is the level of trypsin inhibitor. Although no data are available, it is reasonable to assume that the NIH-07 or Purina Chow diets contain residual trypsin inhibitors which can trigger release of cholecystokinin (ISEO Report 28 Nov 83).

The question of the effect of physiological stimuli and the hormone levels induced by fat intake on cell division in the pancreas or other organs is still

unanswered. It is an important question because of the observation of pancreatic hyperplasia and adenomas in the vehicle control rats in certain NTP studies.

Hormonal effects of high fat intake have been reported in studies of sexual maturation in rodents (Frisch et al., 1975). However, recent studies in large numbers of rats in two laboratories have reported no consistent effect of diets high in either polyunsaturated vegetable oils or saturated fat on sexual maturation of female rats (Ramaley, 1981; Wetsel et al., 1981; Lee & Rogers, 1983).

Diets high in corn oil have been reported to increase serum prolactin in female rats bearing 7,12-dimethylbenz(a)anthracene (DMBA)-induced tumors, bled under anesthesia and examined at about 30 weeks of age (Ip et al., 1980). However later study from the same laboratory reported that the high corn oil diet did not increase serum prolactin significantly (Wagner et al., 1982). Alteration of serum estrogen has also been reported in rats fed a high lard diet (Chan et al., 1977). Other studies of serum prolactin, estrogen and progesterone in cannulated unanesthetized, undisturbed rats reported no dietary effects on the three hormones at any point in the estrous cycle (Rogers, 1983; Wetsel et al., 1983). The earlier studies cited may have showed an effect of diet on prolactin because of the stress produced by bleeding under anesthesia. There was no effect on basal levels in undisturbed, cannulated rats.

The control diets in the studies of Ip et al. (1980), contained only 0.5% corn oil which is borderline in EFA. Their failure to find a signficant difference in prolactin between rats fed 5% and 20% corn oil in the later study suggests that the difference in the early study was due to use of an EFA-deficient diet (Wagner et al., 1982).

Effects on Chemical Bio-availability - The effects of various vehicles on the uptake of subcutaneously injected 3,4,9,10 dibenzpyrene was studied by Homburger

and Tregier (1969). For the same dose of carcinogen, the latent period for development of sarcomas was 16 weeks after administration in peanut oil, 37 weeks in a lipoprotein vehicle, and 62 weeks in Ringers solution. No studies on the influence of different edible oils on the availability or absorption of orally or intragastrically administered carcinogens are available.

Stomach emptying is affected by the presence of fat. For example, Hirono and Shibuya (1972) found that 4.0 ml of saline were emptied from the stomach in two hours, whereas 20-30% of the same volume of olive oil still remained in the stomach at that time. Consequently, when N-methyl-N'-nitro-N-nitrosoguanidine was administered in these two vehicles, saline suspension produced only a 36% incidence of gastric cancer in contrast to 70% with olive oil.

Effects on Chemical Biostransformation - The effects of intragastrically administered corn oil on the biotransformation of xenobiotics have been studied only to a limited extent. Newberne, et al. (1979) have shown that high doses of corn oil in the rat increase liver mixed function oxidases which, in turn, correlated with increased carcinogenicity of aflatoxin B_1. Williams et al. (1983) have shown that corn oil gavage increases rat liver glutathione S-transferase and UDP-glucuronyl transferase. These changes, in turn, would be expected to modify the disposition of xenobiotics with which they form derivates. Iritani and Ikeda (1982) observed that 10% dietary corn oil, compared to lower concentrations, enhanced hepatic catalase, gamma glutamyl-trans peptidase, glutathione peroxidase, and superoxide dismutase.

Effect on Neoplasms - The modulating effect of specific antioxidants (Rogers and Newberne, 1980) on tumor induction indicates another possible mechanism in tumor promotion by PUFA, that of lipid peroxidation. This involves lipid radicals and chain reactions of peroxidation that challenge the integrity of the cellular membrane (Baehner et al., 1981) and may contribute to tumor promotion.

The observed direct in vitro effects of fatty acids (detergent action) on membranes suggest a potential tumor promotion mechanism. However, the significance of these membrane changes on neoplasms is not known. Different antioxidants show varying results in modulating breast tumor growth of animals fed a high fat diet (King and McCay, 1983). Further possible direct effects of polyunsaturated fatty acids may include a non-toxic stimulation of cell proliferation by alteration of the properties or number of hormone receptors.

Tumor Promotion - There is now some evidence that corn oil and other vegetable oils and fats enhance tumor promotion in several organs and tissues (Reddy et al., 1980) compared to high saturated fat, low fat or low EFA diets. These studies have been carried out in mice, rats and hamsters using several carcinogens as initiators; amongst these were azaserine (Roebuck et al., 1981), 2-acetylaminofluorene (AAF) (McCay et al., 1980), N-nitroso bis(2-oxypropyl) amine, DMBA, and aflatoxin-B_1 (Newberne et al., 1979). The tumor sites include breast, colon, liver and pancreas. In the carcinogenesis bioassay program, the occurrence of pancreatic tumors in association with corn oil gavage has not been consistently observed and the mechanism involved is not known (Eustis and Boorman, 1983). The precise mechanism of action of these lipids in the process of tumor promotion is not known although it may be relevant that the phorbal ester tumor promoters induce a number of changes in the structure and function of cellular membranes (Weinstein, 1981).

Although the need for studies which provide information on the susceptibility of test animals during crucial stages of development has been discussed for cancer testing since 1970 (FDA, 1970), there has been no general acceptance of this approach.

In order to evaluate the usefulness of these studies, it is important to review the available toxicity data relating to the problem. Most of the studies involving in utero exposure have involved the use of known carcinogens. The IARC monograph, "Transplacental Carcinogenesis" provides information on much of the earlier research on these substances. In general, the reports indicated that transplacental carcinogenesis could occur and the carcinogenic effects in the offspring may be seen at sites different from those observed in the parent. For example, in transplacental experiments with rats ethylnitrosourea shows a striking neurooncoselectivity that is not observed in the parent. In addition, the fetus may also be more susceptible to tumor development than adult rats, e.g., in the case of ethylnitrosourea adult rats are as susceptible to tumor development as young rats and young rats are not as susceptible as the fetus. This is not always the case since for some carcinogens, e.g., methylnitrosourea, adults are more sensitive than the offspring.

Although transplacental carcinogenesis has been the major interest of in utero studies, there is an increasing amount of information that indicates that non-carcinogenic substances may be the cause of a variety of biochemical and other toxic effects to the developing fetus. Some of this information has developed following a study of the toxic effect of an environmental contaminant such as methylmercury. In the case of methylmercury poisoning, episodes in both Japan and Iraq have demonstrated that the developing fetus could receive toxic doses of methylmercury, although the mother was virtually asymptomatic. The

fetus per se did not appear to be more sensitive than the mother. The distribution of methylmercury in the developing fetus, particularly the brain, showed quite a different pattern than that observed in adults, i.e., in the brain the distribution of methylmercury in transplacentally exposed adults showed selective localization. Similarly the studies of Allen have quite clearly demonstrated the effects of transplacental exposure to PCB. Monkeys with body burdens of PCB derived from previous exposure produced offspring which showed behavioral and learning deficiencies. It was estimated that approximately 40% of the body burden of the offspring was derived from placental transfer.

Transplacental studies using rats have also demonstrated that there is a possibility that the course of enzyme development in the fetus may be markedly altered. This so-called programming may result either in the early or late development of specific enzymes, or in changes in the patterns of development of sex dependent enzymes, i.e., male offsping will develop enzyme profiles more characteristic of female offspring. The extent of such changes will obviously require much research but clearly point to a major area of concern.

The use of progeny of exposed parents in chronic studies will serve as a fitness test for subliminal effects of the type described above.

Some studies with food additives have involved in utero exposure. In the case of saccharin it was clearly demonstrated that in utero exposure resulted in a greater sensitivity of the carcinogenic effects of saccharin. A study with safrole in utero produced renal epithelial tumors in 7% of the female offspring; none of the other experimental or control animals had these tumors. Male offspring (34%) nursed by mothers treated with safrole during lactation had hepatocellular tumors, but not female offspring. However, when safrole was administered post-weaning, a significant increase in hepatocellular tumors was observed in females (48%) but not in males (8%). Of the tumors observed in the

females, 86% were hepatocellular carcinomas of which 42% had pulmonary metastases (Vesselinovitch et al., 1979).

In mammals the fetal membranes and yolk sac show the greatest variation, e.g., in rat the yolk sac may play an important role in the secretion of chemicals in the uterine lining, whereas in humans, the yolk sac appears to be vestigal and its function is not known. Histologically, the epitheliochorial placenta in rodents has 1, 2 or 3 distinct tissue membranes, with the inner two layers being less well differentiated than those in man. The structural differences may result in vastly different rates of transfer of chemicals across the placental membrane.

Regarding the distribution and concentration of the test chemical in the developing fetus, the available (although limited) information suggests that in the case of the rat the test compound is equally distributed throughout the fetuses (e.g., saccharin and acrylamide). It is possible that the fetuses will accumulate more of the test compound than the mother, e.g., in the case of methylmercury, the mercury blood level of newborn infants was higher than that of their mothers. This may reflect the inability of the developing fetus to excrete the test compound. The net result of such effects may be embryotoxicity, which would clearly be reflected in reproduction studies. In order to overcome this problem it would be necessary to reduce the dose of test compound to produce progeny capable of survival in the chronic test that follows.

However, because it may be necessary to use low levels of the test compound to ensure that reproductive or teratogenic effects do not occur, the dose selection may be much lower than those selected for conventional carcinogenicity tests, and the validity of the test may be questioned. Although it may be possible to increase the dose level during the course of the study, such changes would cause problems in determining dose responses.

A number of methods have been described for carrying out such studies. One method involves dosing a parent generation prior to impregnation, during subsequent pregnancy and lactation, and throughout the remainder of their lifetimes, the offspring being dosed with the chemical from weaning and throughout their lifetimes. Such a study provides information on both generations, and would permit the identification of effects specific to in utero and lactation exposure. Other methods described involve utilizing a parent generation for a 3-generation reproduction study, after the F_1 offspring are weaned (FDA) or alternately just maintaining the parents until the offspring are weaned. In either instance the F_1 offspring would be fed throughout their lifetimes.

The FDA proposed a 4 week treatment period of both males and females. Concern has been raised that this may cause genetic damage in both males and females. However, dosing of the females through pregnancy only may not permit significant build up of the test compound in storage tissues to permit exposure of the developing fetus. Adequate pharmacokinetic studies will be required to determine the period required to develop the equilibrium state.

This must take into account the needs of the dams, neonate, juvenile and F_0 and F_1 adults. Detailed information will be required from reproduction, teratogenic and pharmacokinetic studies. Reproduction and teratogenic studies will indicate what dose levels will not cause reproductive or teratogenic effects in the F_1 generation. Pharmacokinetic studies will provide information on whether or not metabolic processes in the dam or off-spring are saturated. Information will also be obtained on rate of uptake and distribution pattern and rate of elimination of the test compound and its metabolites, which will enable suitable dose levels to be selected for the developing fetus and nursing pup, as well as the dose level and period of exposure for the dam, prior to mating.

At present, in utero exposure may be most useful in studying the effects of known carcinogens or compounds strongly suspected of being carcinogenic, since

the sensitivity of the developing fetus, and the suckling animal to the test compound or its metabolite, may result in a substantially reduced latent period. Although final evaluation of the usefulness of this technique as a source of additional toxicologic information must await the testing of a wide range of chemicals, this does not mitigate against the validity of the technique as a more exact means of reflecting the human experience.

Appendix III

It is generally recognized that there are only a few organizations that have an animal care program adequate to maintain the health of laboratory rodents for a two-year experiment. For example, many multipurpose, centralized laboratory facilities maintain other animals and research projects that carry infectious organisms. These can serve as a source of infection for bioassay animals. Any practice which mixes animal species suppliers and other agents without adequate attention to facilities or programs, is almost assured of a spread of diseases.

All rodent pathogens must be considered undesirable, but some more so than others. There is a temporal relationship that must be considered in preparing a more definitive list. At the time of procurement from the vendor, animals must be free from mycoplasma, ectromelia, LCM virus, polyoma virus, hemobartanella, epyrythrozoon, tape worms, and all ectoparasites. Procurement of animals with any other pathogen should be evaluated on a case by case basis.

Specific diseases which will result in termination of an individual study will vary from study to study. In the meantime documentation of endemic diseases is essential. Monitoring should be a continuous process from production through the course of the bioassay study. Monitoring should include viral serology, parasitology, bacteriology (both aerobic and anaerobic), and complete necropsy of the animals with gross pathologic and histopathologic examinations. At a minimum, it should encompass all of those agents considered unacceptable. The number of animals to be monitored must be determined by statistical analysis based on the incidence of disease that is deemed necessary to be documented. Monitoring on arrival at a subcontracted laboratory is recommended. Sufficient sentinel animals should be provided to allow monitoring at the end of 1 month, 3 months, 9 months, and at termination of each study. Interim deaths that occur

during the study must be examined to determine the presence or absence of concurrent disease. It is recommended that the pathology protocol for interim deaths be revised to include the steps necessary to accomplish this objective. Steps to be considered in addition to histopathology and gross pathology include storing of tissues and sera for future examination as needed. Sentinel animals should be randomized by cage throughout racks of cages of the same species. Laboratories should monitor the environment as part of their own management process.

Among the key animal environmental factors to be considered are the following:

1. Filters

   Participants in the Boiling Springs meeting uniformly agreed that cage top filters have no specific disease control function in normal laboratory operations. Because they prevent passive diffusion of air, they tend to have an adverse physiologic impact on the animals. The Committee recommended that filters only be used to protect against overt or gross contamination by particulate matter. Their use is of particular importance where there is an active compound in the feed or excreted by the animal.

2. Bedding

   The Committee recommended that the use of stainless steel wire bottoms inside polycarbonate solid bottom cages instead of bedding be investigated. If wire bottoms are functionally acceptable, their use would be recommended in preference to bedding. If bedding is used, it should be changed sufficiently often so that the ammonia levels remain acceptable.

3. Temperature and Humidity

   The cage environment or the micro-environment of the animals is the important issue in looking at temperature and humidity. The Committee

recommended that the standards for temperature and humidity be the thermal-neutral temperature for the animal on test ±2° and that the humidity be 50% ±10%, both measured in the cage. Determiniation of the room temperature that achieves this temperature in the cage environment would be the responsibility of the investigator.

4. Air Quality

It is the consensus of the committee that air quality should be standardized according to some parameter that would permit or give the performing laboratory the flexibility to make adjustments. There was general agreement that ammonia levels would be the most effective parameter. As a starting point, the recommendation was that the maximum allowable room level of ammonia would be 25 parts per million, measured at any one point in time. This is a tentative recommendation, pending a review of the available data on the relationship between room and cage concentrations of ammonia. Another comment regarding air quality is that the air should be recirculated only within a single room rather than within a section of the laboratory. Recirculation of air in a zone of the laboratory would be the second choice, and acceptable only if the experiments being performed within that zone were controlled and of equivalent type.

5. Light

The Committee recommends that the light period be reduced from 12 hours of light and 12 hours of darkness to 10 hours of light and 14 hours of darkness. The rationale for this recommendation is that it would constitute an energy savings. The intensity of light should be minimum lumens consistent with good husbandry and good study practices. A two-stage light system whereby the light can be increased above minimum levels while the technician is working in the room was encouraged. Red lights

synchronized with the room lights should be used in service halls if animal room doors have windows or are to be opened during the dark cycle.

6. Noise

Noise level should be kept at an absolute minimum. The use of piped-in music generally was approved because it improves technician morale and performance and tends to cover up the noise technicians make while they are working in the animal room. It is recommended that indiscriminate introduction of radios into the animal room by technicians should be discouraged.

7. Sanitizing Agents

All members of the committee agreed that the use of sanitizing agents containing essential oils should be prohibited in the animal rooms. It was generally agreed that not enough information exists to prohibit or to endorse the use of other sanitizing agents. The proposition that water alone be used to clean animal rooms was discussed. The problem with water is that it is not an effective cleaning agent for either waste or for chemicals used in the room that are not water soluble.

8. Vaccines

It was generally agreed that more study of vaccines and their use in long-term studies is needed. Clearly, vaccines cannot be used if it is known that they will influence the results of the study. Cost-benefit and risk-benefit analyses are required. The majority of the committee favored the position that if effective vaccines were available and the disease agents could not be excluded by the use of facilities and management procedures alone, then the use of vaccines is an alternative that must be considered.

9. Feed

Diet is an important variable in toxicology studies (Wise, 1982) and the need for a well-defined diet has been emphazied (Newberne et al., 1978). A committee of the American Institute of Nutrition was formed in 1973 to develop recommendations on nutritional methodology to serve as guidelines to scientists in such areas as toxicology, carcinogenesis, and aging for devising diets which could be used for rats and mice in long-term studies. Their report (American Institute of Nutrition, 1977) defined the types of experimental diets and described recommended diets. The first type of diet defined by the committee was the cereal based diet also known as the unrefined diet or non-purified diet. These terms were applied to diets composed predominantly of unrefined plant and animal materials which may contain added vitamins & minerals. There are two types of cereal based diets, the "open formula" and the "closed formula." An open formula diet is diet whose precise composition is published. These diets must always be assembled exactly as specified. NIH-07 and NIH-31 are the most commonly used open formula rodent diets. The "closed formula" diet is a diet for which the manufacturers do not disclose the exact composition. Purina Lab Chows and Wayne Laboratory animal diets are two commonly used closed formula diets. The AIN committee recommended that the NIH-07 open formula diet be used. This recommendation has been made by other committees (National Cancer Institute, 1976; Newberne et al., 1978; IARC, 1980; Hamm, 1980; Coates, 1982). The National Institutes of Health has used this diet since 1972, the National Cancer Institute - National Toxicology Program Bioassay Program has been using NIH-07 for all bioassys since 1979 and CIIT has been using it for all bioassays since 1981. The formula for this diet has been published

(Knapka, 1974). If an autoclavable open formula diet is required, NIH-31 is a frequent choice. Precautions should be taken to avoid adversely affecting the nutrient composition or physical properties of the diet during autoclaving (ILAR, 1976).

The other type of diet discussed by the AIN committee was the purified diet which are diets composed primarily of commercially refined proteins, carbohydrates, and fat with added mineral and vitamin mixtures. A new purified diet, AIN-76^{TM} purified diet, was proposed. Problems with this diet prompted the publication of changes in the formulation (Bieri, 1980). This new formulation, AIN-76a still caused problems such as hepatic lipidosis (Medinsky, 1982; Hamm et al., 1982), kidney mineralization (Nguyen and Woodard, 1980) and effects on metabolism by gut flora (deBethizy et al., 1983). Therefore purified diets as currently formulated are unacceptable for carcinogen bioassays. See ILAR (1976) for a discussion of the shipment handling and the storage of feeds. Feed should be used within 90 days of manufacture. Feed should be analyzed for nutrient value (i.e. protein, fat, fiber, ash, phosphorus, thiamine, vitamin A, carotene) and contaminants (i.e., aflatoxins, nitrosamines, arsenic, cadmium, calcium, lead, mercury, selenium, nitrate, nitrite, bacteria, coliforms, E. coli, BHA BHT, alph BHC, beta BHC, gamma BHC-lindane, delta BHC, heptachlor, aldrin, heptachlor, epoxide, DDE, DDD, DDT, HCB, mirex, methoxychlor, dieldrin, endrin, telodrin, chlordane, toxaphene, estimated PCB's ronnel, ethion, trithion, diazinon, methyl parathion, ethyl parathion, malathion, endosulfan I, endosulfan II, endosulfan sulfate). Various organizations have published lists of permissible concentrations of contaminants in animal diets (TOSCA, 1979;

ILAR, 1976). Results of feed analysis should be included in the publication of the bioassay results.

10. Water

Based on the lack of published information, the potential effect of chemical contamination of the water apparently has been given little attention. The Environmental Protection Agency has reported extensive chemical contamination of both raw and treated water in various areas of the U.S. (Symons et al., 1975) and a recent review of the problem has been published (Pye and Patrick, 1983). Acidification of drinking water has been shown to affect selected biologic phenomena in male mice (Hall et al., 1980), hyperchlorination of drinking water has been shown to depress the activity of macrophages in mice (Fidler, 1977), and variation in water quality has been reported to cause reproductive failure in female mice (McKinney, et al., 1976). A system for producing purified water for laboratory animals has been tested and is one way to minimize water contaminants (Raynor, et al., 1983). Water should be monitored for contaminants (Raynor, et al., 1983) and the results included in the publication of the bioassay data.

11. Caging

Animals exposed by inhalation should be housed in stainless steel wire mesh cages. Chemical distribution studies should be done to guarantee that the cageing does not affect exposure. Animals exposed by other routes should be housed in solid cages that contain the chemical. Plastic cages with filter tops are commonly used. Animals exposed to highly toxic or carcinogenic chemicals should be housed in systems that completely contain the chemical. Disposable cages that are incinerated after use are expensive but simplify handling dangerous chemicals.

Regardless of cage type, animals should be given adequate space that meets or exceeds the recommendations in the guide for care and use of laboratory animals (USDHEW, 1978). The decision must be made to house the animals individually or in groups. Individual housing has the advantages that identification of the animals is faster since a large copy of the ear tag number can be placed on the cage, and fighting and cannabalism are eliminated. In dosed feed studies individual housing is required if the dose each animal is getting needs to be determined. If group housing is chosen the groups should not be changed as animals die and the original cage should be large enough to hold all the animals at their maximum weight. Cages should be rotated in the rack to assure that all animals stay in each location on the rack for an equal amount of time. Cages should be changed at intervals which prevent the ammonia level in the cage from reaching unacceptable levels. Only heat treated hardwood chip bedding manufactured for animal use should be used. To minimize inhalation or ingestion of the bedding, wire floorwalks that suspend the animals above the bedding can be used (Raynor et al., 1983). Autoclaving the bedding minimizes the potential for disease transmission from organisms in the bedding and reduces ammonia generation from urease positive bacteria in the bedding.

The ideal modification of the pathology protocol would retain the basic ability to detect increased numbers, unusual types or reduced latency of neoplasms as well as chemically induced toxic lesions yet not require every tissue from every animal be subjected to the pathologist's review. The statistical inferences from such modifications should also be considered. If the modified protocol cannot address these objectives as well or better than the current protocol, the value of the resulting scientific data would be compromised and incorporation of the alternative approach may not be justified. However, it would also be a mistake not to recognize the fact that chronic carcinogenesis studies involve the use of a complex test system (animals), that these studies may have certain inherent flaws in terms of detailed reproducibility, sensitivity and the ability to interpret observations is sometimes difficult. In other words, it is a mistake to ask the test method to answer more questions than it is capable of.

Various alternatives to the current pathology protocol portion of the NTP carcinogen bioassay were investigated. The alternatives were compared with the current protocol using the NTP historical data base. Several assumptions must be made prior to evaluation of the alternatives. They are as follows:

1. Number of animals per sex/dose remains essentially the same as the current protocol.
2. Number of dose groups may vary but are a minimum of 2.
3. All animals from all dose groups will receive a complete post mortem examination as defined in Table 1.
4. A "complete" set of tissues will be preserved in a suitable fixative for possible histopathologic examination.
5. A "complete" histopathologic examination is defined as microscopic examination of the 31-33 tissues/organs.

6. A typical study will have 6-8 tissues/organs/animal showing grossly visible lesions (or a target organ effect).
7. Males and females are examined in the same manner.
8. Each sex/species within a given study is evaluated independently.
9. The same pathologist will examine all of the histologic sections of a given animal species.
10. A 15-month evaluation for the purpose of defining chronic toxicity is part of the protocol.

The alternatives chosen for evaluation in this presentation are as follows:

Option 1 - Inverse Pyramid

1. "Complete" (31-33 tissues/organs) histopathologic examination of all animals from the highest dose and control groups.
2. Histopathologic examination of tissues/organs at lower doses where chemically related neoplastic or non-neoplastic effects were identified in the high dose animals or in which there is a grossly visible lesion.
3. If survival in the high dose is reduced because of toxicity (unrelated to neoplasia), "complete" histopathology is conducted on all of the animals in the next high dose in addition to the high dose. Target organs and gross lesions are examined in lower dose group(s) if they are part of the study.

Option 2 - Selective Inverse Pyramid

1. Histopathologic examination from all high dose and control animals of a subset of "core" tissues (16-18) which have previously been associated with neoplasia (Table 2) or which are suspect target organs based on subchronic studies, previous studies with this chemical or related chemicals and/or which are suspect based on clinicopathological observations.

2. Histopathologic examination of tissues/organs at lower doses where chemically related neoplastic or non-neoplastic effects (target organs) were identified in the high dose group.
3. Histopathologic examination of all gross abnormalities in all animals.
4. If survival in the high dose is reduced because of toxicity (unrelated to neoplasia), histopathology will be conducted on the "core" tissues in the next high dose group as well as the high dose group animals. Target organs will be examined in lower dose(s) groups.

Each option was evaluated as to its impact on: pathology workload; ability to answer the essential questions of neoplastic and non-neoplastic chemical effects; and the level of management required to implement.

It is apparent after reviewing the alternative described, as well as the current protocol, that there is one essential prerequisite that is required for any of the approaches to succeed; a thorough post mortem examination. A thorough necropsy entails much more than the collection of tissues for subsequent histopathologic examination. If a given lesion is not recognized at necropsy and therefore not saved (fixed), it is lost forever. A pathologist has only one chance to perform the necropsy of a given animal and to recognize abnormalities; in contrast to stained sections which can be re-evaluated innumerable times. It is for this reason that the NTP insists upon a professional pathologist to conduct or directly supervise the post mortem examinations. In addition, it is preferable that the pathologist evaluating the histopathology be the same person who conducted the necropsy. A better correlation of gross and microscopic pathology is attained in this way.

It soon became apparent that the background incidence of a given neoplasm or other lesion was necessary information for interpreting potential chemical effects. Therefore it became essential to examine as many tissues/organs as

possible from control and treated animals to make legitimate conclusions. This view or approach has been the state-of-the-art since the decade of the 1960's. Because of this past approach, we now have an extensive data base upon which to consider more efficient methodologies. In this regard, the NTP data base comprising over 250 carcinogenesis bioassays represents a vast amount of information upon which to draw for evaluating alternative approaches.

With the above background in mind the various modification were compared:

Option 1 (Inverse Pyramid) is straightforward, requires a minimum amount of scientific judgment to implement, should be easy to manage and should yield an equivalent amount of scientific information as the current protocol. It also has an added benefit that it has been used previously in NTP subchronic studies and in other government and industrial laboratories where it has withstood the test of time and rigorous peer review. This modification would provide control animal data for all tissues/organs and allow for continual updating of the historical data base. While concurrent control data is the most valuable for comparative purposes, inconsistent results can often be identified by comparing the data to historical data. This modification's only drawback is that the savings in pathology effort, while substantial (21%), are relatively small compared to the other modification evaluated. The savings would be greater in studies involving more than two dose levels.

A continual updating of the historical data base for the core tissues would be possible with the selective inverse pyramid approach. It would allow for comparison of pathology data to that developed in the study laboratory as well as program wide data. However, over a period of time the data base for noncore tissues would be diminished.

APPENDIX V

Detailed Guidelines for Combining Tumors by Type and Anatomic Site

TABLE 5. Guidelines for Combining Benign and Malignant Neoplasms in the Fischer 344 Rat and $B6C3F_1$ Mouse

Tissue	Tumors	Combine
Liver	Neoplastic nodule (rat) or Hepatocellular adenoma (mouse) and Hepatocellular carcinoma	Yes
	Bile duct adenoma and Bile duct carcinoma	Yes
Mammary Gland	Fibroma and Fibroadenoma	Yes
	Carcinoma and Adenocarcinoma	Yes
	Fibroma/Fibroadenoma and Carcinoma/Adenocarcinoma	No
Thyroid	*Follicular cell adenoma and Follicular cell carcinoma	Yes
	*C-cell adenoma and C-cell carcinoma	Yes
	Follicular cell neoplasms and C-cell neoplasms	No
Pituitary	*Adenoma and Carcinoma	Yes
Lung	Bronchioalveolar adenoma and Bronchioalveolar carcinoma	Yes

*Neoplasms where the incidence of hyperplasia is taken into consideration in the evaluation of a carcinogenic response.

Table 5. (Cont.)

Hematopoietic System	Rat	
	Leukemia Mononuclear cell (Fischer rat) Lymphocytic Undifferentiated	Yes
	Myelogenous leukemia and Leukemias-other types	No
	Malignant lymphoma (lymphosarcoma) Lymphocytic Lymphoblastic Histiocytic Reticulum cell Mixed cell	Yes
	Leukemias and Lymphomas	No
	Mouse	
	Lymphocytic leukemia and Undifferentiated leukemia	Yes
	Myelogenous leukemia and Leukemia-other types	No
	Malignant lymphoma (lymphosarcoma) Lymphocytic Lymphoblastic Histiocytic Reticulum cell	Yes
	Leukemias-all except myelogenous Lymphomas-all types	Yes
Pancreas	*Islet cell adenoma and Islet cell carcinoma	Yes
	*Acinar cell adenoma and Acinar cell carcinoma	Yes
	Islet cell neoplasms and Acinar cell neoplasms	No

*Neoplasms where the incidence of hyperplasia is taken into consideration in the evaluation of a carcinogenic response.

Table 5 (Cont.)

Gastrointestinal Tract	*Forestomach-Papillomas and Squamouse cell carcinomas	Yes
	*Glandular region and intestinal Adenomas/adenomatous polyps Adenocarcinomas	Yes
	Leiomyomas and Leiomyosarcomas	Yes
	Fibromas and Fibrosarcomas	Yes
	Squamous cell neoplasms Glandular neoplasms Mesenchymal neoplasms	No
	Leiomyomas/leiomyosarcomas and Fibromas/fibrosarcomas	No
Kidney	Tubular cell adenomas and Tubular cell carcinomas	Yes
	Transitional cell papillomas and Transitional cell carcinomas	Yes
	Lipomas and Liposarcomas	Yes
	Transitional cell neoplasms and Tubular cell neoplasms	No
	Lipomatous neoplasms and Other types of renal neoplasms	No
Urinary Bladder	Transitional cell papillomas and Transitional cell carcinomas	Yes
Skeletal System	Osteoma and Osteosarcoma	Yes
	Chrondroma and Chrondrosarcoma	Yes
	Osteoma/osteosarcoma and Chrondroma/chrondrosarcoma	No

*Neoplasms where the incidence of hyperplasia is taken into consideration in the evaluation of a carcinogenic response.

Table 5 (Cont.)

Adrenal Gland	*Cortical adenomas and Cortical carcinomas	Yes
	*Pheochromocytoma and Malignant pheochromocytoma	Yes
	Cortical neoplasms and Medullary neoplasms	No
Brain	All gliomas, i.e. Oligodendroglioma Astrocytoma	Yes
	Granular cell neoplasms and Gliomas	No
	Nerve cell neoplasms and Gliomas	No
	Meningiomas-all types Other CNS neoplasms	No
Ovary/Testicle	Germ cell neoplasms-all types	Yes
	*Stromal neoplasms-all types	Yes
	Germ cell neoplasms and Stromal neoplasms	No
Uterus	*Stromal polyps and Stromal sarcomas	Yes
	Glandular adenomas and Adenocarcinomas	Yes
	Stromal neoplasms and Glandular neoplasms	No
Integument	*Basal cell tumors all types and Pilomatrixoma	Yes
	Sebaceous gland tumors-all types	Yes

*Neoplasms where the incidence of hyperplasia is taken into consideration in the evaluation of a carcinogenic response.

Table 5 (Cont.)

Squamous cell	Squamous cell papilloma Squamous cell carcinoma	Yes
	Squamous cell neoplasms and Adnexal neoplasms	No
	Basal cell neoplasms Squamous cell neoplasms	No
	Keratoacanthoma Squamous cell carcinoma	No
Subcutis	Fibromas and Fibrosarcomas	Yes
	*Hemangiomas and Hemangiosarcomas	Yes
	Leiomyomas and Leiomyosarcomas	Yes
	Fibromas/fibrosarcoma and Leiomyomas/leiomyosarcomas	No
	Connective tissue neoplasms and Endothelial neoplasms	No
Preputial/Clitoral Gland	*Adenoma and Carcinoma	Yes
Zymbal Gland	*Adenoma and Carcinoma	Yes
Nasal Cavity	*Adenoma and Adenocarcinoma	Yes
	Squamous cell neoplasms and Glandular neoplasms	No
	Esthestioneuralepithelial neoplasms and Other neoplasms	No

*Neoplasms where the incidence of hyperplasia is taken into consideration in the evaluation of a carcinogenic response.

TABLE 6. Guidelines for Combining Neoplasms from Different Anatomic Sites in the Fischer 344/N Rat and B6C3F$_1$ Mouse

Tissue	Combine
Adnexa	Yes
Alimentary tract	Sometimes
Bone	Yes
Cartilage	Yes
Central nervous system	
glial	Yes
nerve cell	Yes
Endothelium	Yes
Fibrous connective tissue	Yes
Genital tract	Sometimes*
Hematopoietic	Yes
Lymphoreticular	Yes
Peripheral nervous system	
nerve sheath	Yes
Respiratory tract	No*
Skeletal muscle	Yes
Skin	Yes
Smooth muscle	Yes
Urothelium	Yes

*Tissue sites with significant exceptions.

Appendix VI

The detailed criteria used by IARC to classify the overall evidence of carcinogenicity of a chemical agent to experimental animals are as follows:

1. Sufficient evidence of carcinogenicity, which indicates that there is an increased incidence of malignant tumors: a) in multiple species or strains; or b) in multiple experiments (preferably with different routes of administration or using different dose levels); or c) to an unusual degree with regard to incidence, site or type of tumor, or age at onset. Additional evidence may be provided by data on dose-response effects, as well as information from short-term tests or on chemical structure.
2. Limited evidence of carcinogenicity, which means that the data suggest a carcinogenic effect but are limited because: a) the studies involve a single species, strain or experiment; or b) the experiments are restricted by inadequate dosage levels, inadequate duration of exposure to the agent, inadequate period of follow-up, poor survival, too few animals, or inadequate reporting; or c) the neoplasms produced often occur spontaneously and, in the past, have been difficutlt to classify as malignant by histological criteria alone (e.g., lung adenomas and adenocarcinomas and liver tumors in certain strains of mice).
3. Inadequate evidence, which indicates that because of major qualitative or quantitative limitations, the studies cannot be interpreted as showing either the presence or absence or a carcinogenic effect.
4. No evidence applies when several adequate studies show, within the limits of the tests used, that the chemical is not carcinogenic, the number of negative studies is small, since, in general, studies that show no effect are less likely to be published than those suggesting carcinogenicity.

IARC holds that data in animals meeting the standard of categories 1 or 2 (sufficient evidence, limited evidence) suggests potential carcinogenicity for man. The Preamble of recent monographs develops these ideas in a broader context.

☆ U.S. GOVERNMENT PRINTING OFFICE: 1984-421-132:4726