Available online at www.sciencedirect.com
ScienceDirect
Food and Chemical Toxicology 45 (2007) 759–796
www.elsevier.com/locate/foodchemtox

Framework for use of toxicity screening tools in context-based decision-making

John Doull [a], Joseph F. Borzelleca [b], Richard Becker [c], George Daston [d], John DeSesso [e], Anna Fan [f], Penelope Fenner-Crisp [g], Michael Holsapple [h], Joseph Holson [i], G. Craig Llewellyn [j,1], James MacGregor [k,2], Jennifer Seed [l], Isabel Walls [g,3], Yin-tak Woo [l], Stephen Olin [g,*]

[a] *University of Kansas, United States*
[b] *Virginia Commonwealth University, United States*
[c] *American Chemistry Council, United States*
[d] *The Procter & Gamble Company, United States*
[e] *Mitretek Systems, United States*
[f] *California Environmental Protection Agency, United States*
[g] *ILSI Research Foundation, Risk Science Institute, One Thomas Circle, NW, Suite 900, Washington, DC 20005, United States*
[h] *ILSI Health and Environmental Sciences Institute, United States*
[i] *WIL Research Laboratories, United States*
[j] *Kraft Foods, United States*
[k] *US Food and Drug Administration, United States*
[l] *US Environmental Protection Agency, United States*

Received 14 September 2006; accepted 27 October 2006

Abstract

One of the principal applications of toxicology data is to inform risk assessments and support risk management decisions that are protective of human health. Ideally, a risk assessor would have available all of the relevant information on (a) the toxicity profile of the agent of interest; (b) its interactions with living systems; and (c) the known or projected exposure scenarios: to whom, how much, by which route(s), and how often. In practice, however, complete information is seldom available. Nonetheless, decisions still must be made. Screening-level assays and tools can provide support for many aspects of the risk assessment process, as long as the limitations of the tools are understood and to the extent that the added uncertainty the tools introduce into the process can be characterized and managed. Use of these tools for decision-making may be an end in itself for risk assessment and decision-making or a preliminary step to more extensive data collection and evaluation before assessments are undertaken or completed and risk management decisions made. This paper describes a framework for the application of screening tools for human health decision-making, although with some modest modification, it could be made applicable to environmental settings as well. The framework consists of problem formulation, development of a screening strategy based on an assessment of critical data needs, and a data analysis phase that employs weight-of-evidence criteria and uncertainty analyses, and leads to context-based decisions. Criteria for determining the appropriate screening tool(s) have been identified. The choice and use of the tool(s) will depend on the question and the level of uncertainty that may be appropriate for the context in which the decision is being made. The framework is iterative, in that users may refine the question(s) as they proceed. Several case studies illustrate how the framework may be used effectively to address specific questions for any endpoint of toxicity.

* Corresponding author. Tel.: +1 202 659 3306.
E-mail address: solin@ilsi.org (S. Olin).
[1] Present address: Wm. Wrigley Jr. Company.
[2] Present address: Toxicology Consulting Services.
[3] Present address: USDA Food Safety and Inspection Service.

0278-6915/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.fct.2006.10.025

Keywords: Toxicity testing; Decision tree; Structure–activity relationships; Endocrine disruptor; Carcinogenicity; Immunotoxicity

1. Introduction

Definitive information about the potential human health impacts of an agent in human exposure scenarios is often not available, so risk management decisions may more typically be made on the basis of more limited information. The amount of information required to make decisions is dependent upon the context. In some situations, decisions may be made based on the results of short term assays, while in other situations extensive information may be available. Sometimes gaps in the toxicology database can be filled with screening assay results.

Screening tools are usually relatively simple and rapid studies or models that are designed to assist in decision-making and in determining the need for additional research and/or testing. In medicine, screening tools are used for diagnostic purposes to detect disease and abnormal function, and to guide and monitor therapy. In toxicology, screening tools are used to identify those agents that exhibit a particular set of attributes that either exclude them from further research and/or testing or cause them to undergo more rigorous evaluation. Screens also are used to conserve resources (animals, test agents or time), as range-finding tests, as surrogates for more complex toxicity studies or to provide supplemental information about specific adverse effects. Screens are particularly useful in those situations where decisions have to be made in a short time frame. The outputs from screens often identify a specific trait (or limited set of characteristics) associated with a known outcome.

It is important to define the overall goal for screening, because the selection of the appropriate screening tool(s) for use in a specific situation depends primarily on the question(s) the screen is intended to answer. When used to assist decision-making regarding potential human health risks, screening tools need to be selected to provide appropriate information about both hazard and dose–response to help determine whether the expected or estimated human exposure(s) will be of concern. The degree of certainty that is needed from the exposure and hazard datasets depends on the nature of the question being asked and the time available to make the decision. For example, a scenario in which exposure is likely to be low and involves a limited number of people may be amenable to the use of less-than-definitive hazard data, particularly if the available hazard data indicate that the compound in question has a low order of toxicity. Conversely, an exposure scenario that includes a large population for which the extent and magnitude of exposure is unknown, and for which there is an indication of potential toxicity of concern, may require a much larger and more definitive dataset than can be generated solely with screening-level assays.

Prior to employing a screening tool, the decision-maker should indicate the level of risk and the level of uncertainty in the assessment deemed to be acceptable. While this is ultimately a management decision, it should be informed by the available science. When choosing the screening tool(s), it is important to understand the strengths and limitations of the various screening approaches in order to select the appropriate screen(s) for a particular purpose. This should follow the development (and, hopefully, codification) of criteria for determining when screening data are sufficient to support a hazard characterization. A number of factors should be considered when determining the adequacy of a screening system, including, for example, (1) whether the appropriate life stages are represented; (2) whether appropriate animal species is/are used; (3) whether the dose–response relationship can (or, need) be established; (4) whether endpoints that are relevant for assessing the toxicity of concern have been included; and (5) for assays that are to be used in the risk assessment process, whether the route of exposure is relevant (or can be extrapolated to the relevant route(s)) using credible pharmacokinetic procedures.

Screening approaches often are used in medical practice to identify patients with particular health-related characteristics. Biomarker-based screening approaches are commonly used, for example, presence of occult fecal blood is used as a screen for colon cancer and alpha-fetoprotein in maternal blood is used as an indicator of abnormalities in the fetus (Mandel et al., 1993; Hardcastle et al., 1996; Wald et al., 1977; Milunsky and Alpert, 1984; Macri and Weiss, 1982). Both of these screens use a marker that has a known link to a particular disease state to identify individuals with elevated risk for the disease, so that a definitive clinical follow-up can be undertaken and a therapeutic decision made. In the case of fecal occult blood, the mechanistic link is the fact that the progression of colon polyps generally leads to intestinal bleeding. The issues of sensitivity and specificity of this screening method have been well studied, and illustrate the need to define both false negative and false positive rates when employing the assay (Levin et al., 2003). Alpha-fetoprotein is a component of fetal plasma, and its presence at measurable levels in the maternal system has been correlated with the presence of an open neural tube, and possibly other defects. Data on the rates of identified disease states that do not show up in these screening assays (false negatives), and the frequency of elevated levels of the marker due to causes other than the disease (false positives) determine the utility of the screening approach. Mechanistic knowledge lends substantial weight to the reliability of the findings and is very useful in identifying sources of potential error, such as why the screening marker might be elevated in the

absence of the disease, or not elevated in the presence of the disease.

The first step in predicting the potential risks associated with exposure to chemical or physical agents is to evaluate the available information on the agent, the target population and the conditions of exposure. For the agent, this includes its identity and characterization of physical and chemical properties such as vapor pressure, solubility, stability, purity, molecular weight, melting point, boiling point, isomerism, and reactivity. Data on individual and population exposure situations and on sensitive segments of the population are examined and used to estimate plausible "worst case" scenarios. This information then is used to determine the number and type of toxicity and epidemiologic studies that will be needed to identify exposure situations that provide "reasonable certainty of no harm" or equivalent protection. In those situations where there is no exposure or the exposure is expected to be below the minimal level established by a regulatory agency to have no significant risk to health (e.g., the threshold of toxicological concern), no further studies are indicated. The US Food and Drug Administration and certain other national food regulatory agencies as well as the Joint FAO/WHO Expert Committee on Food Additives (JECFA) have used this concept to distinguish indirect food additives (contact materials) (USFDA, 1995) and flavoring agents (JECFA, 2005) that pose a significant level of concern from those that pose negligible concern. As would be expected, the regulatory requirements for agents in the former category would be expected to be more rigorous than for those in the latter category.

Adverse effects (toxicity) are the result of the interaction of agents (chemical or physical) with biological systems (the host). This interaction is defined by the exposure scenario. The two major factors in exposure are dose and time. Dose or dose rate is a simple variable (e.g., mg/kg/day) but time is a more complex variable because it involves not only the duration and frequency but also the persistence of the agent in the host. The hazard profile of an agent is defined by the specific adverse effects that it is inherently capable of causing following exposure which may be of an acute, intermediate or chronic duration. The host or biological system is defined by its susceptibility to the adverse effects of the agent. Agents also can be characterized by their mode or mechanism of action and by their potency. This paradigm can be used to design more efficient protocols for toxicity studies by first determining whether the rate-limiting events responsible for adverse effects are occurring in the kinetic or in the dynamic pathways leading to the effect (Fig. 1). Key rate-limiting events that occur in the dynamic pathway (e.g., enzyme induction or inactivation, formation of adducts or other types of irreversible or long-lasting effects) should direct further research towards characterization of the dynamic mechanisms of injury and recovery. Agents with longer half-lives require a focus toward studies on the kinetic pathways of absorption and elimination.

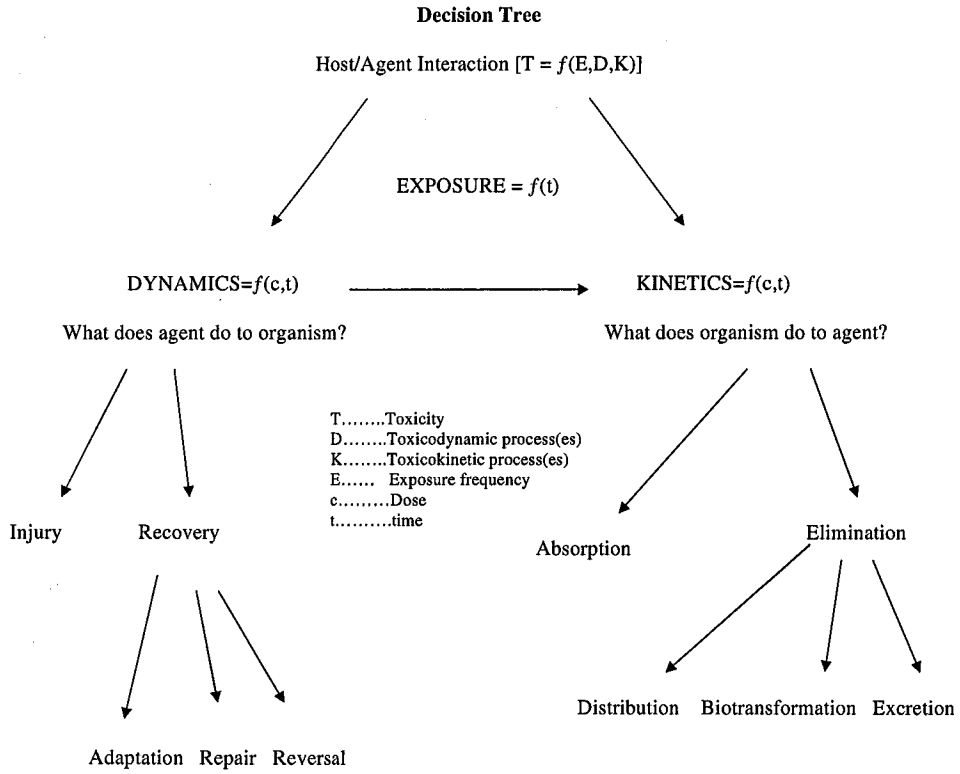


Fig. 1. Conceptual outline of the decision-tree approach showing dose and time as variables of toxicity (from Rozman and Doull, 2000).

The ILSI Risk Science Institute established its Expert Working Group to develop a framework for evaluating the utility and application of screening tools for decision-making. The framework is illustrated by a number of case studies covering selected endpoints that give examples of specific questions and demonstrate how screening tools can be used to address the questions at hand.

2. Screening tools framework

A general framework for decision-making using screening tools is presented below (Fig. 2). The framework shares many characteristics with the framework that has been developed by USEPA for its human health and ecological risk assessment processes (e.g., USEPA, 1998a). The initial and most crucial step to assure attainment of a credible decision is problem formulation. Problem formulation involves the construction of the question(s) to be addressed by the screening tool. Specific questions should be identified at the beginning of the exercise that consider the health issue, host factors and the exposure scenario. It is critical to (1) clearly define the toxicological endpoint(s) or health status to be investigated; (2) delineate the putative causative agent or situation to be tested; (3) describe the environmental conditions for the study (these include the exposure parameters of dose, route and rate of administration, and such host factors as the species, sex, and life stage of the test subjects); (4) determine the length of time that is

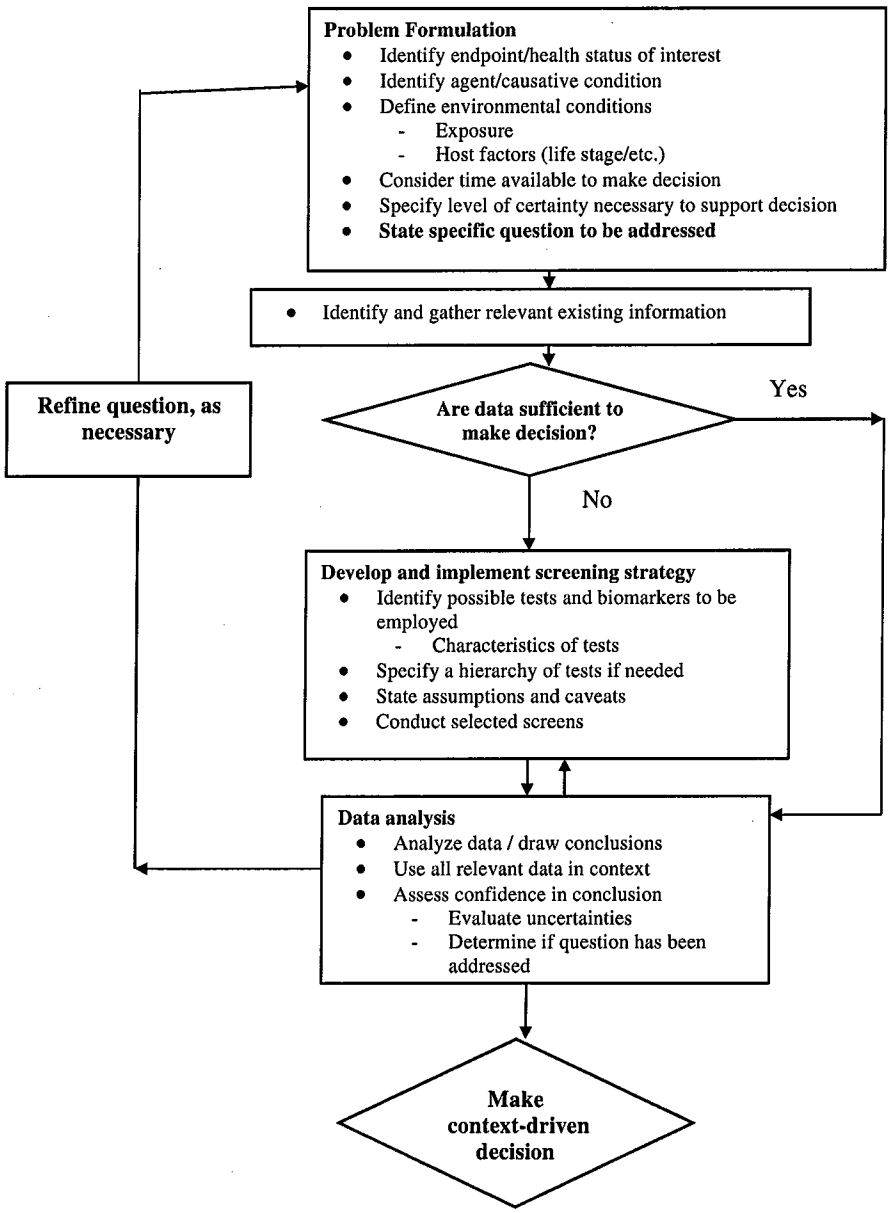


Fig. 2. Scheme for use of screening tools in contextual decision making.

available for one to generate data, analyze results and draw conclusions, and (5) identify the level of confidence needed to support the decision. Once these issues are considered, specific questions can be formulated and should be stated clearly and concisely. An example of a specific risk management question might be "Is it safe for the population to drink water containing Agent X, which is present in the town's water supply at a specified level, or should we recommend that the population use an alternative source?"

At this point, the existing data should be evaluated critically, using a weight of evidence approach, to determine the adequacy and relevance for making a decision to address the specified question. In a specific situation, a decision may be needed within a very short timeframe, even though some of the information one would like to have is not available. If the data are not appropriate for making a decision, a plan should be developed to generate the information needed to provide an answer within the confines of the situation described in problem formulation. Depending on the nature and extent of data that are required, the characteristics of additional tests or series of tests will be specified. This may involve a tiered scheme whereby simpler tests are conducted first and more complex studies follow. Depending upon the constraints placed on the screening methods (e.g., limited time, money or resources), assumptions may be required to complete the evaluation. Assumptions should be stated clearly, and any resultant caveats should be acknowledged.

Characteristics that distinguish successful screening tools include the following:

- Relevance: the extent to which the method is related to the endpoint(s) of interest, and the method's usefulness for its specified purpose.
- Biological validity: the ability of the method to minimize false positive and false negative findings.
- Sensitivity: the capacity for the method to identify the outcome of interest.
- Specificity: the degree to which the method responds to factors not associated with the outcome of interest.
- Accuracy: a measure of the degree to which the method generates the correct answer.
- Precision: a measure of the degree to which the method is reproducible within or across laboratories.
- Robustness: degree to which similar results are obtained in different laboratories or with minor modifications to the protocol or reagents.
- Practicality: the extent to which conduct of the method is economically and methodologically prudent.
- Formalization: a standardized protocol is available that permits some flexibility.

In order for the results of a screening method to be useful, it is important that the results are well-corroborated. This is often referred to as "validation." An assessment of the validity of a test methodology can be made by applying the following criteria, as summarized from the ICCVAM report on validation (NIEHS, 1997):

- The proposed use is clearly stated.
- A defined biological basis or relationship links the endpoint(s) of the screening test to the effect(s) of interest.
- The methods are set forth in a formal, detailed protocol.
- The reliability of the screening test method has been assessed.
- The relevance of the screening test method has been assessed.
- The limitations of the screening method are known and have been described.
- Previously generated data are available for independent scientific review.
- Data quality is adequate (preferably of GLP quality).
- The screening test methodology has been subjected to, and has passed, independent peer-review.

Screening tests generally employ biomarkers—objective measurable endpoints that are believed to reflect a functional effect or structural alteration of interest (Biomarkers Definitions Working Group, 2001). For example, serum transaminases and bilirubin are biomarkers generally employed to identify liver toxicity in animal or human screening tests. As with any given test system, the characteristics of biomarkers employed must be defined in the context of the screening objective. It is desirable that a well-understood mechanistic relationship exists between the biomarker and the functional or structural event that is to be identified, and it is useful to define the conditions that may yield "false-positive" responses (a biomarker response not related to the event of interest) or "false-negative" responses (failure of the biomarker to respond when the functional or structural event of interest occurs).

Once a screening strategy has been agreed upon, screening can proceed. In some cases, data may already exist. In other cases, data will have to be generated *de novo*. The data generated according to the plan should be analyzed in the context of all other existing relevant data. In particular, it is recommended that a weight-of-evidence approach be applied to address the following questions:

- Is there an indication of toxicity?
- What is the nature of the dose–response relationship?
- Is the response species or sex-specific?
- Is the response relevant for humans?
- Were the data obtained using the appropriate route of exposure? (which should simulate human exposure conditions)
- Is the response reversible?
- Is the response consistent with other published data?
- Is the response biologically plausible?
- If there is a signal of toxicity, would additional data improve the ability to predict risk (or to answer the particular question being asked)?

- If additional data are needed, do the screening level results suggest a more tailored approach for the definitive assessment?
- If human exposure data are available, have they been considered in determining the adequacy of the screening results? (e.g., low potential for exposure combined with low potential for toxicity may lead one to conclude that no further testing is necessary).

The results of this context-based decision-making process must be evaluated with respect to the experimental uncertainties, to determine one's confidence in the decision. The use of screening tools should be considered to be an iterative process. If one's confidence in the decision is sufficiently high, the process is complete. If, however, one's confidence in the decision is low, the question should be refined and the process repeated, as time and resources allow. In addition, the situation should be monitored so that periodically the decision can be re-evaluated.

3. Case studies to illustrate framework

3.1. Structure activity relationships (SAR/QSAR) as a screening tool

3.1.1. Introduction

Structure Activity Relationships/Quantitative Structure Activity Relationships (SAR/QSAR) provide a basis for the study of qualitative and quantitative relationships between chemical structure and biological activity of chemical compounds. It is often considered the first tool for strategic screening in product development, hazard identification and risk assessment. Given *sufficient* knowledge on structurally or functionally related compounds, SAR/QSAR can be used for screening virtually any reasonably well-defined biological, toxicological or pharmacological endpoint of interest including kinetic characteristics (i.e., ADME: absorption, distribution, metabolism, excretion). Traditionally, SAR/QSAR studies have been conducted by, with or for experts in the field (domain experts) who have extensive knowledge on the endpoint of interest. However, the explosive growth in knowledge of the molecular basis of many biological endpoints has increased the need to handle the vast amount of data generated to make complex predictions and discoveries. The technological advances in predictive methodologies have resulted in SAR/QSAR analysis becoming a multi-disciplinary study with increasing reliance on computerization for data mining, machine learning and prediction. The proliferation of commercially available predictive software has expanded the end-users from domain experts to risk managers, regulators, health protection professionals and even casual users who may have little or no technical expertise and, thus, need guidance to interpret the end results correctly. The purpose of this section is to provide some general background information on SAR/QSAR, their known and potential uses, criteria for assessing their scientific soundness and ways to avoid pitfalls when using these tools.

3.1.2. Brief overview of SAR, QSAR, artificial intelligence (AI) and data mining

The field of SAR/QSAR analysis has undergone tremendous evolution as our knowledge of the molecular basis of toxicological/pharmacological/biological activity of chemicals rapidly increases and more powerful predictive technology becomes available. A variety of SAR/QSAR methods have been developed for application, depending on the nature of the endpoint/characteristic to be evaluated and the complexity of the question(s) being asked. These include: (a) qualitative SAR, which involves expert judgment of structural features (e.g., "structural alert"), assessment of factors that affect absorption/distribution/metabolism/excretion (ADME) and consideration of other supportive information as a basis for prediction; (b) classical QSAR methods which involve development of mathematical models that relate the biological activity of molecules to their chemical structures and corresponding physicochemical properties and other molecular descriptors (e.g., topological, quantum mechanical), using various statistical methods, such as regression analysis, principal component and factor analysis, discriminant and pattern recognition analysis and similarity analysis; (c) computer-assisted methods, which include knowledge rule-based expert systems as well as machine learning, neuronal networking, artificial intelligence (AI) or data mining systems to identify molecular fragments of interest, discover SAR features, classify active and inactive compounds, induce knowledge rules, and/or develop decision logic; (d) biologically-based models such as 2D or 3D receptor modeling, binding, and ligand SAR; and (e) integrative models that incorporate or combine both chemical and biological information.

SAR/QSAR methods may be classified as statistically-based (relying on statistical, deterministic or probabilistic association) or mechanistically-based (e.g., receptor modeling, electrophilicity-based) or they may be a combination of both. Ideally, SAR/QSAR studies should strive to achieve statistical association but have a mechanistic foundation. The importance of mechanistic understanding in SAR/QSAR will be discussed further in the following section.

When selecting chemicals for model development, SAR/QSAR studies may be undertaken using structurally closely related sets of chemicals (termed homogeneous or congeneric) or structurally diverse sets of chemicals (termed heterogeneous, noncongeneric or global). In general, predictive models using congeneric data require fewer chemicals for model development and tend to perform better (presumably because they are more likely to act by a particular mechanism) but are more limited in scope. Predictive models using heterogeneous data tend to be more adept in discovering new insights but may be more likely to yield false negative results if the predicted chemical is not well-represented in the training database. Beyond structural homogeneity, SAR/QSAR studies also have

been conducted on classes of chemicals with similar biological activity/function (e.g., peroxisome proliferators or physicochemical characteristics such as lipophilicity, amphiphilicity, acidity, reactivity) to identify common factors beyond chemical structure.

3.1.3. Uses of SAR, QSAR, AI and data mining

SAR/QSAR is a powerful, efficient and economical screening tool that can be used for many different purposes. These may be loosely classified under the following categories:

(a) *Product development:* SAR/QSAR is a crucial tool in the development of new products in the pharmaceutical and chemical industries. In addition to contributing to the discoveries and development of products with desired dynamic and kinetic properties and early exclusion from development of products with undesirable characteristics, SAR/QSAR also has been used to design environmentally safer chemicals as well as to search for safer substitutes to existing chemicals.

(b) *Hazard identification and risk assessment:* For untested chemicals, such as those found in the ambient environment or under development for introduction into commerce, SAR/QSAR has been used to identify potential hazards, to design experimental studies, and to select surrogate/analog chemicals that may be used for interim risk assessments. For tested chemicals, SAR/QSAR may be used to provide input into weight of evidence assessments, to elucidate potential mechanisms of action, and to contribute to selection of the quantitative risk assessment methodology to be applied.

(c) *Health and environmental protection:* SAR/QSAR has been used widely to assess the environmental fate, ecotoxicity and human health hazard of chemicals. It also has been used as a tool for prioritization of environmentally-occurring chemicals (e.g., air pollutants or disinfection by-products in drinking water) for testing and monitoring so that limited resources can be used effectively.

(d) *Regulation:* SAR/QSAR has been used extensively for regulating new chemicals for which test data are not available or are inadequate (e.g., Wagner et al., 1995). There also is increasing worldwide interest (e.g., Cronin et al., 2003) in expanding the use for other regulatory purposes although the legal ramifications surrounding the uncertainties associated with SAR/QSAR studies remain to be resolved.

3.1.4. Criteria for assessing the scientific soundness of SAR/QSAR

SAR/QSAR studies represent a specialized type of screening tool. Beyond the usual criteria for assessing the predictive capabilities of screening tools (e.g., predictive accuracy, sensitivity, specificity), there are other special considerations. Since SAR/QSAR studies can be conducted by virtually limitless means and approaches, it is essential to evaluate the scientific soundness of the specific method/model selected by the user for the specific objective of interest. The following are such criteria as well as issues that the users should be aware of for effective use of SAR/QSAR:

(a) *Selection of endpoints or characteristics:* The selection of endpoints can have a significant impact on the outcome of any SAR/QSAR studies. In general, well defined endpoints with a relatively simple mechanism of action (e.g., mutagenicity in Ames test) are easier to model than those that may involve multiple components or mechanisms of action (e.g., acute lethality, carcinogenicity). Endpoints that are a composite of multiple contributing sub-endpoints should not be directly modeled unless there is some common basis among the targeted chemicals.
With regard to screening for kinetic characteristics, similar care must be taken to incorporate those parameters which are known to contribute to impacting the characteristic under evaluation (e.g., the octanol–water coefficient and skin absorption or reactions catalyzed by the enzymes of metabolism, such as the cytochrome P450 superfamily).

(b) *Knowledge-based vs. statistically-based models:* For knowledge-based SAR models, it is important to scrutinize the expertise and track record of the domain experts involved. For statistically-based models, factors for consideration include: data quality and variability (ideally from a single reputable source or using the same protocol), coverage and representativeness (e.g., covering sufficient congeneric range for interpolation; for noncongeneric, structurally or mechanistically diverse or with sufficient "informational content"), and active/inactive distribution (ideally close to 1:1). The lack of data coverage and representativeness has been considered to be one of the main reasons for the overall poor performance in a recent international predictive challenge (PTC-2000) using the NTP cancer bioassay data to train various data mining models to predict FDA data (Benigni and Giuliani, 2003; Helma and Kramer, 2003).

(c) *Methodology, selection and handling of molecular descriptors:* To the extent possible, users should be aware of the fundamental basis (e.g., molecular fragment, receptor modeling) of the SAR/QSAR models and suitability to the endpoint of their interest. The basis for selection of molecular descriptors, including procedures to focus on the key descriptors and to avoid interdependent or "overfitting" descriptors should also be scrutinized. Several recent reviews discuss issues on methodology and modeling techniques (e.g., Eriksson et al., 2003; Franke and Gruska, 2003).

(d) *Validation:* Depending on the type of SAR/QSAR method, a number of different validation methods have been used. These include: (i) peer review, (ii) dividing modeling data into a "learning set" for

model training and a "test set" for validation, (iii) 10- or 20-fold cross-validation for large databases, (iv) "leave-one-out" technique for smaller databases, (v) retrospective validation using data from different sources (e.g., the NTP vs. the FDA databases) for model development and validation, and (vi) prospective validation of *real life* predictions of the outcome of future tests. By far the most stringent test is the prospective validation. In two NTP-sponsored predictive exercises in which various SAR/QSAR modelers were asked to make predictions on the outcome of an NTP bioassay underway, a number of internally cross-validated systems did not perform as well as expected (Benigni and Zito, 2004).

(e) *Transparency and rationale:* "Transparency" should be an important component of any SAR/QSAR prediction. The scientific rationale, the rules used and/or the analogs used to support the prediction should be provided so that the user may be able to judge the reasonableness of the prediction. "Black box" approaches and predictions based on vague, abstract or overly sophisticated information or methods should be avoided.

(f) *Confidence and uncertainty:* The degree of confidence or uncertainty also should be a component of prediction. This is particularly important for regulatory uses. Probability, predictive domain and reasoning are some of the approaches used to express confidence/uncertainty. In general, predictions by interpolation from a congeneric study of compounds with reasonably well understood mechanism(s) of action tend to inspire higher confidence, particularly if the user can locate additional supportive information. On the other hand, negative predictions based on lack of information should always be closely scrutinized to prevent false negatives. An in-depth review of methods for assessing reliability and uncertainty of QSARs has been published (Eriksson et al., 2003).

(g) *Strengths, weaknesses and limitations:* As may be expected from the diversity of approaches and data used in model development, different SAR/QSAR models may perform differently and have their own unique strengths, weaknesses and limitations. Many reviews have addressed these issues (e.g., Hulzebos et al., 2001; Greene, 2002). Users should be aware of these issues relative to their specific endpoint of interest or objective. For any specific endpoint/characteristic, it is advisable to use more than one model and compare the predictions.

3.1.5. Importance of mechanistic considerations: chemical carcinogenicity as an example

Mechanistic considerations often are overlooked by most statistically-based QSAR studies. This may lead to errors, particularly for endpoints that are known to involve multiple mechanisms. Mechanistic considerations may help almost any QSAR study by helping to (a) select the most appropriate molecular descriptors (e.g., electrophilicity vs. receptor-based); (b) serve as a criterion to assess whether the training database is suitable for making predictions on the chemicals of interest; (c) stratify the training database into smaller but mechanistically more homogenous subsets to improve predictive capability; (d) interpret outliers; (e) guide hypothesis testing to fill data gaps; and (f) assess the human significance of predictions based on animal data.

The importance of mechanistic understanding in SAR/QSAR studies can be illustrated by using chemical carcinogenesis as an endpoint. Chemical carcinogenesis is a multistage, multi-factorial process, which conceptually consists of three operational stages – initiation, promotion, and progression. Initiation involves a mutational event that may include gene mutation, chromosome aberration, translocation and instability. Promotion involves clonal expansion of initiated cells to reach a critical mass by a variety of means such as cell proliferation, inhibition of programmed cell death, inhibition of terminal differentiation, and loss of growth control. Progression may involve a second mutational event, the loss of tumor suppressor gene, impairment of immune surveillance and acquisition of ability to metastasize. From the mechanistic point of view, carcinogens may be classified as genotoxic and epigenetic/nongenotoxic. Genotoxic carcinogens, also known as DNA-reactive carcinogens, generally are chemicals that directly interact with DNA either as parent chemicals or as reactive metabolites to form DNA adducts or lesions which, if unrepaired, may initiate carcinogenesis. Epigenetic carcinogens are agents that act through secondary mechanisms that do not involve direct DNA damage.

There are numerous examples of chemical carcinogens that act predominantly by genotoxic mechanisms. The classical major structural classes of genotoxic carcinogens include direct-acting carcinogens (e.g., epoxides), aromatic amines, *N*-nitroso and hydrazo compounds, halogenated hydrocarbons, aflatoxin-type furocoumarins, homocyclic and heterocyclic polycyclic aromatic hydrocarbons. The key common features for most of these potent genotoxic carcinogens are: (a) propensity to be or to generate electrophilic intermediates; (b) availability of a stabilizing mechanism to allow transport of reactive intermediates from the site of activation to the site of interaction for DNA covalent binding; (c) characteristics of persistent DNA adducts or chromosomal lesions; and (d) ability to act on various stages of carcinogenesis.

The scientific literature on epigenetic mechanisms of chemical carcinogens has been growing at an explosive pace in the past several years because of the importance of mechanistic understanding in elucidating the molecular basis of carcinogenesis, considering human relevance of animal data, and modeling quantitative risk assessment. Epigenetic carcinogens include agents that (a) act via receptor-mediated mechanisms; (b) generate reactive oxygen species to cause secondary DNA damage; (c) cause aberrant gene expression; (d) disturb the homeostatic status of cells; and

(e) elicit cytotoxicity with subsequent compensatory regenerative hyperplasia. For agents that involve reversible binding to receptors, the two key common elements that are favorable are: (a) molecular size/shape; and (b) long biological half life to allow sustained binding/activation of the receptor. The principal known mechanisms include: (a) peroxisome proliferation; (b) arylhydrocarbon (Ah) receptor-mediated effects and other enzyme induction; (c) inhibition of gap junctional intercellular communication; (d) oxidative stress; (e) perturbation of DNA methylation and gene expression; (f) hormonal imbalance; (i) cytotoxicity-induced regenerative cell proliferation; (j) inhibition of microtubulin polymerization; and (k) impairment of immune surveillance. The SAR features of the epigenetic carcinogens are not as well studied as those of genotoxic carcinogens; most of the available information has been summarized by Woo and Lai (2003).

As may be expected from the fundamental difference between genotoxic and epigenetic carcinogens, the SAR features contributing to genotoxic mechanisms often differ from those contributing to epigenetic/nongenotoxic mechanisms. Using a molecular fragment approach, it has been shown that the toxicophores associated with (a) inhibition of gap junctional intercellular communication (an epigenetic/promoting mechanism) and (b) mutagenicity in Ames *Salmonella* tests (a genotoxic/initiating indicator) have very little structural overlap suggesting that (a) and (b) lack commonality and may contribute to carcinogenicity independently (Rosenkranz, 2003). Thus, it would be important to consider both the genotoxic and epigenetic SAR features for integrative assessment of the carcinogenic potential of chemicals.

3.1.6. Future perspectives

SAR/QSARs are indispensable screening tools in product development, hazard identification and risk assessment. However, much work is still needed before they can be used with confidence by average users and regulatory agencies. Several future directions should be explored:

(a) *Critical evaluation of current SAR/QSAR methods:* The number of commercial SAR/QSAR predictive software programs has been rapidly increasing in recent years. Although most are internally cross-validated, their predictive capabilities in real life situations remain to be tested. More critical evaluations such as the NTP predictive exercises should be conducted to gain valuable feedback to accentuate strengths and reduce weaknesses. Expressions of confidence in the results and the attendant uncertainty, avoidance of false negatives due to lack of information and false positive due to overconservatism, and defining of predictive domains are important issues to consider. More domain experts, of whom many may be skeptics, should be involved in these evaluations to provide insight and first hand predictive experience on specific endpoints.

(b) *Expansion of database/knowledge base:* The predictive capability of SAR/QSAR methods is limited by the input of training information. For some structural or functional classes/types of chemicals, valuable training knowledge and data are often not accessible or available in the open literature. More efforts should be made to build centralized, publicly available databases (e.g., Richard and Williams, 2003) and to find means to make proprietary or confidential data available for model development and evaluation. Such databases will be highly useful for more effective data mining as well as developing strategic, goal-oriented research to fill key knowledge/data gaps. The National Center for Toxicogenomics has initiated a decade-long project to develop a Chemical Effects in Biological Systems (CEBS) knowledge base with the ultimate goal to assimilate toxicological, biological and chemical information from multiple public domain databases to meet the information needs of "systems toxicology" (Waters et al., 2003).

(c) *Expansion of integrative approaches:* For endpoints with multiple mechanisms and complex system biology, more integrative approaches should be considered to cover multiple mechanisms and/or to model individual steps of the overall process to improve predictive capability. The US Environmental Protection Agency (USEPA, 2003a) has developed a framework for a Computational Toxicology Research Program, which aims to utilize novel technologies derived from computational chemistry, molecular biology and systems biology to develop integrative models to improve hazard screening, enhance quantitative risk assessment and reduce uncertainties in modeling the linkages in the source-to-outcome continuum of environmental release → environmental concentration → exposure concentration → target organ dose → early biological effects → adverse outcome.

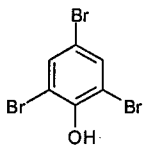
(d) *Utilization of input from emerging technologies/computational biology/bio-informatics:* Functional criteria or activity–activity relationships analysis using short-term tests or biomarkers to predict long-term endpoints has long been used as an effective tool. The recent rapid growth in "omics" technology (genomics, proteomics, metabonomics, systemomics, etc.), epigenetic biology, and high through-put screening assays should be capitalized on along with computational biology and bioinformatics tools to provide support for hypothesis-driven SAR/QSAR for the most effective predictions. For example, toxicogenomics (e.g., Fielden and Zacharewski, 2001; MacGregor, 2003) offers great promise of integrating with traditional SAR/QSAR to reveal, assess and elucidate the mechanisms of action/toxicity pathways of known and suspected toxicants, to serve as a tool for generating and supporting hypotheses, and to improve predictive capability.

3.1.7. Case examples

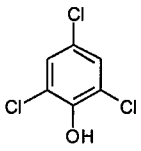
3.1.7.1. 2,4,6-Tribromophenol (TBP). Problem formulation: TBP, a suspect carcinogen, was detected at US Superfund sites (uncontrolled or abandoned places where hazardous waste is located, possibly affecting local ecosystems or people (http://www.epa.gov/superfund/sites)). A cancer bioassay was planned by NTP but has yet to be conducted. The US EPA, under the mandate of the Resource Conservation and Recovery Act (RCRA), had to answer the question "Is there a sufficient basis to add TBP to the list of hazardous wastes and the list of hazardous constituents for regulation?"

Review of existing data: Review of existing data showed the availability of some mutagenicity data on TBP but no cancer bioassay data.

Development and implementation of strategic plans: Qualitative SAR was selected as the initial screening tool. A closely-related structural analog, 2,4,6-trichlorophenol (TCP), had been determined to be a rodent carcinogen in the NTP bioassay (NTP, 1979) (positive in both sexes of mice and in male rats; negative in female rats). The structural analogy was considered to be very high because (a) with the exception of the halogen substitutions, TBP and TCP are structurally identical; (b) there are a number of clear examples of brominated hydrocarbons suspected to be carcinogenic by structural analogy to their chlorinated analogs and subsequently tested and found to be carcinogenic (e.g., PBBs and PCBs). The structural analogy was reinforced further by comparative evaluation of available mutagenicity data. The mutagenicity profile of TBP in several short-term *in vitro* tests matched that of TCP. The results for TCP were predominantly negative (positive only in an *E. coli* assay and *S. typhimurium* TA1535 and *Bacillus subtilus* without S9 activation (ATSDR, 1999)). TBP has been tested in the same bacterial systems and found to be negative in all instances, with and without activation (WHO, 2005). The overall weight of evidence provided a reasonable basis to support the consideration of TBP as a potential carcinogen. The lack of mechanistic information on TCP, however, may pose limits on its use in quantitative risk assessment. It appears not to be a promoter as a 20% solution of 2,4,6-trichlorophenol in benzene, applied dermally 2 times/week for 15 weeks, did not promote skin tumors in four strains of mice (albino Suher, Holtzman, C3H and CAF) initiated with 9,10-dimethyl-1,2-benzanthracene (Boutwell and Bosch, 1959).



2,4,6-tribromophenol



2,4,6-trichlorophenol

Outcome: Qualitative SAR was used as a basis for proposing the addition of TBP to the list of hazardous wastes and the list of hazardous constituents (USEPA, 1998b); regulation was, however, hampered by legal challenges.

3.1.7.2. Prioritization of disinfection byproducts (DBPs). Problem formulation: DBPs are formed when disinfectants such as chlorine and chloramine react with organic and inorganic matter in water (Weinberg et al., 2002). The discovery of carcinogenic DBPs in drinking water raised a public health concern. Several hundred DBPs have been detected to date. Under the mandate of the Disinfectants/Disinfection Byproducts Rule of the Safe Drinking Water Act, EPA is required to answer the question "What are the priorities for research and monitoring of DBPs, so we can concentrate resources on the more hazardous DBPs?"

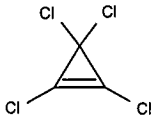
Review of existing data: A review of existing data showed the availability of cancer bioassay data on a small fraction of DBPs. In addition, a number of DBPs currently are being tested or selected for testing by NTP. These DBPs have been excluded from consideration. About half of all DBPs also have been excluded from consideration for a variety of other reasons (e.g., instability, lack of bioavailability, possible detection of artifacts, or on the GRAS list). The remaining DBPs have been organized into 18 structural classes for literature searches and data review. Limited mutagenicity data are available on some of the DBPs.

Development and implementation of strategic plans: Mechanism-based SAR analysis has been selected as the initial screening tool. Codified mechanism-based SAR knowledge rules captured in EPA's OncoLogic Cancer Expert System as well as expert judgment were used to conduct a semiquantitative analysis of relative carcinogenic concern of each DBP (Woo et al., 2002). Table 1 illustrates the SAR analysis of the haloalkane/haloalkene class. The rules used included consideration of (a) the nature, number, and position(s) of halogen(s); (b) the molecular size of the compound; (c) potential direct acting activities; (d) metabolic activation to reactive intermediates; and (e) resonance stabilization of reactive intermediates. The initial SAR consideration was supplemented by an analysis of the available mutagenicity data. The final concern level together with the scientific rationale used have been delineated (see Table 1), making the entire process transparent to the public. The relative concern level has been suggested as a basis of prioritization to supplement occurrence analysis.

Outcome: DBPs with moderate or higher concern levels have been selected by EPA for initiation of testing programs by EPA research laboratories and NTP. In addition, a national monitoring study has been initiated to determine the occurrence of these DBPs in drinking water. This prioritization effort reduced the number of candidate DBPs from several hundred to a more manageable size of less than twenty.

3.1.7.3. QSAR analysis of carcinogenic aromatic amines. Problem formulation: Company X detected low levels of a

Table 1
SAR analysis of carcinogenic potential of haloalkanes and haloalkenes detected as DBPs (M = moderate; LM = low-moderate; mar = marginal; L = low; Woo et al., 2002)

Structure	Concern level	Rationale
CH_2Br_2	M	Structural analogy to dichloromethane which is a rat carcinogen. The brominated compound is expected to be more hazardous than the chlorinated compound due to more favorable leaving tendency and GSH-mediated activation. Positive genotoxicity (Ames, ara forward mutation, *E. coli*) data
CH_2BrCl	M	Structural analogy to dichloromethane. The brominated compound is expected to be more hazardous due to more favorable leaving tendency and GSH-mediated activation
CHClBrI	M	Structural analogy to bromodichloromethane which is a rodent carcinogen. The iodo group is expected to be a better leaving group than the chloro/bromo group
$CHCl_2I$	M	Structural analogy to bromodichloromethane and chloroform which are both carcinogenic. The iodo group is expected to be an even better leaving group than the chloro/bromo group
$Br_3C–CHBrCl$	M	Structural analogy to pentachloroethane, which is a mouse carcinogen and 1,1,2,2-tetrabromoethane, which is hepatotoxic
CH_3Cl	LM	Structural analogy to iodomethane and chloroethane, which are both carcinogenic. Positive Ames assay
$CH_3CH_2CH(Br)CH_3$	LM	Positive lung adenoma assay and positive Ames assay. The internal location of bromine may limit its genotoxic potential
$CH_3CH(Cl)CH(Cl)CH_3$	LM	Vicinal dichloro substituted may lead to GSH-mediated activation but internal location of chlorine may limit its genotoxic potential
$Cl_3CC(CH_3)=CH_2$	LM	Structural analogy to 1,3-dichloropropene which is a mouse carcinogen but not as favorable due to steric hindrance by methyl and marginally active trichloro group
$ClCH_2C(Cl)(CH_3)CH_2CH_3$	LM	Structural analogy to 2-chloroisobutane which is positive in the lung adenoma assay. Vicinal dichloro substitution may lead to GSH-mediated activation but methyl substitution may lead to steric hindrance
	LM	Limited structural analogy to hexachloropentadiene, which has negative bioassay data. However, may have some genotoxic potential. One of the chlorines at the bridged carbon may leave and generate a carbonium ion which can be stabilized by the ring by resonance stabilization
$Cl_2CH(CH_2)_3CHCl_2$	mar	Potential alkylating agent but its genotoxic potential may be reduced because the potentially reactive terminal carbons are both dichlorinated making them not as favorable as mono chlorine as leaving groups
$ClCH_2(CH_2)_6CH_3$	mar	Despite the presence of a terminal chlorine, this compound is expected to be a weak alkylating agent due to its high molecular weight and its saturated chain
$CH_3CH(Cl)(CH_2)_9CH_3$	L	Expected to be a very weak alkylating agent due to its high molecular weight and its saturated chain

monocyclic aromatic amine Y in its consumer product Z. Toxicity studies of Y revealed mutagenic potential and suggestive evidence of carcinogenicity. A 2-year cancer bioassay was started. Before the results became available, Company X needed interim information to answer the question "What is the estimated potential carcinogenic potency of Y?" to conduct a quantitative risk assessment to ensure the safety of Z.

Review of existing data: Aromatic amines represent one of the most important structural classes of carcinogens and extensive mechanistic studies exist. For most genotoxic aromatic amines, covalent binding of the metabolically activated, electrophilic nitrenium ion metabolites to DNA is considered the first critical step toward tumor initiation. Factors that may affect this key step include favorable physicochemical properties for reaching target organs, ease of metabolic activation of the amino group, or resonance stabilization of reactive intermediates to allow traveling from the site of activation to site of interaction with DNA. A robust QSAR study of a congeneric series of 37 monocyclic carcinogenic aromatic amines is available in the literature (Benigni and Passerini, 2002). It has been shown that the carcinogenic potency in mice (BRM) can be effectively modeled by the following QSAR equation:

$$\begin{aligned}\mathrm{BRM} = {} & 0.88 \times \log P \times I(\mathrm{monoNH_2}) + 0.29 \times \log P \\ & \times I(\mathrm{diNH_2}) + 1.38 \times E_{\mathrm{HOMO}} - 1.28 \times E_{\mathrm{LUMO}} \\ & - 1.06 \times \Sigma \mathrm{MR}_{2,6} - 1.1 \times \mathrm{MR}_3 - 0.2 \times E_s(R) \\ & + 0.75 \times I(\mathrm{diNH_2}) + 11.16 \\ & [\text{standard deviations not shown}] \\ & n = 37,\ r = 0.907,\ r^2 = 0.823,\ s = 0.381, \\ & F = 16.3,\ P < 0.001\end{aligned}$$

where BRM is $\log(\mathrm{MW}/\mathrm{TD}_{50})$ (molecular weight divided by daily dose required to induce 50% tumor incidence at the end of standard lifespan); $\log P$ is log of octanol/water partition coefficient (a measure of hydrophobicity), I ($\mathrm{monoNH_2}$) and I ($\mathrm{diNH_2}$) are indicator variables, which are equal to one if the condition within the parentheses is true; E_{HOMO} and E_{LUMO} is energy of highest occupied molecular orbital (useful as a measure of potential for oxidative metabolism) and energy of lowest unoccupied molecular orbital (a measure of reactivity), respectively; $\mathrm{MR}_{2,6}$ and MR_3 are molar refractivity of substituents at *ortho* and *meta* positions of aniline (a measure of bulk in relation with molar volume; useful as indicator of steric hindrance surrounding the amino group); $E_s(R)$ is

Charton's substituent constants for substituents at the functional amino group.

This QSAR equation has high statistical significance and captures most of the mechanistic highlights, making it an excellent case of statistical association with mechanistic backing.

Development and implementation of strategic plans: The utility of QSAR in predicting the carcinogenic potency of untested chemicals is a subject of debate. In general, QSAR models developed from chemicals with diverse chemical structures (global models) do not perform as well as those developed using congeneric series of chemicals (congeneric/local models). Even for a congeneric model, beyond assessing the robustness of the model, the suitability of the model for the chemical of interest also must be evaluated. In the current specific case example, the available model as described above is robust both from the statistical and mechanistic viewpoints. Comparison of the chemical structure of compound Y with the 37 chemicals used to develop the model showed that compound Y can be considered a member of the congeneric series. The various ring and amino group substituents in compound Y are adequately covered within the set of 37 chemicals to allow interpolative evaluation. Overall, the model is considered suitable for compound Y and may be used to provide an estimate of mouse TD_{50} as the first step of interim safety assessment. Additional factors such as supportive data, species extrapolation, and mixture effect of product Z also must be considered for total assessment of safety.

Outcome: This is a hypothetical case with emphasis on the need for considering various issues before any QSAR models can be used for predicting toxic potency.

3.1.7.4. Towards achieving sustained therapeutic steady-state blood plasma levels. Problem formulation: Inhibitors of angiotensin-converting enzyme (ACEIs) are effective and widely-used drugs for the treatment of hypertension and congestive heart failure. At least a dozen members of this class are approved for use in the US. Most of these are prescribed to be taken in single oral daily doses. A few, however, must be taken two or three times a day to assure efficacy because of their rapid uptake and short elimination half-life. Of interest would be the designing of ACEI analogues which would exhibit enhanced percutaneous penetration resulting in sustained steady-state blood plasma levels of therapeutic benefit.

Review of existing data: Captopril was the first oral ACEI to be marketed (Duchin et al., 1988; Jackson, 2001). The unchanged molecule is the pharmacologically-active moiety. It is poorly absorbed through the skin, but rapidly absorbed orally with a bioavailability of ~75%. Peak concentrations in plasma occur within an hour. The clearance half-life is approximately two hours. The primary route of elimination is the kidney. The presence of food in the GI tract reduces its bioavailability by 25–30%, so the drug should be taken at least one hour before a meal.

Development and implementation of strategic plans: Moss et al. (2006) conducted a study the aim of which was to synthesize a series of prodrugs of captopril and to determine if a quantitative structure-permeability relationship (QSPR) model could be used to design therapeutically viable prodrugs. Molecules with the highest predicted $k(p)$ values were synthesized and characterized, and $J(m)$ measured in Franz diffusion cells from saturated aqueous donor across porcine skin (fresh and frozen). *In vitro* metabolism also was measured. Captopril and the prodrugs crossed the skin relatively freely, with $J(m)$ being highest for ethyl to butyl esters. Substantial first-order metabolism of the prodrugs was observed, suggesting that their enhanced percutaneous absorption was complemented by their metabolic performance.

Outcome: The application of the QSPR model predicted that ester and thiol prodrug derivatives of captopril would have lower maximal transdermal flux ($J(m)$) than the parent drug, since the increases in permeability coefficient ($k(p)$) of prodrugs would be outweighed by the reductions in aqueous solubility. The results suggested that QSPR models provided excellent enhancements in drug delivery. This was not seen at higher lipophilicities, suggesting that issues of solubility need to be considered in conjunction with any such use of a QSPR model.

3.2. Predicting and characterizing carcinogenic potential

Testing chemicals for carcinogenic potential in laboratory animals provides the most common data source for evaluation of potential carcinogenic risk to humans. Because long-term bioassays in experimental animals require extensive time and resources to perform, attempts have been made to use short- and medium-term assays to provide data to predict potential carcinogenicity. These include short-term genotoxicity tests which assess the potential of a chemical to interact directly or indirectly with deoxyribonucleic acid (DNA), and medium-term assays such as the rat liver altered foci assay, which is based on the two-stage or multi-stage theory of carcinogenesis. More recently, several genetically-modified mouse models have been developed and evaluated for their ability to appropriately identify potential carcinogens. Some are being explored as potential adjuncts to, or replacements for, one or both of the species used in the traditional four-cell rodent bioassay.

The determination or prediction of carcinogenicity of a chemical is an integral component of the risk assessment process and a guide to regulatory decision-making. Risk assessment is a tool used to determine the potential health effects (in this case, carcinogenicity) that may result from human exposure, as individuals or populations, to hazardous materials (such as environmental contaminants or a chemical product). The inherent uncertainties in this process must be acknowledged and documented. Risk assessment is used in the product decision-making process (e.g., decision on further testing, product development or use)

and the regulatory decision-making process (e.g., food or consumer product recall, provision of an alternative drinking water supply, cancellation of registration or restricted use of a pesticide).

The determination of the carcinogenic potential of a chemical is dependent upon the availability of data and other relevant information. Such data and information may include: (1) epidemiologic evidence linking human exposure and cancer; (2) evidence in experimental animal studies; (3) results from short-term *in vitro* and *in vivo* genotoxicity tests in non-mammalian and mammalian systems; (4) knowledge of mode of action, and (5) information that the mode of action is or is not anticipated to occur in humans and progress to tumors. In certain situations, in spite of the absence of a robust level of knowledge of these aspects, or when the pertinent data are inadequate for a full evaluation, a risk management decision may still have to be made. Human data on the evaluation of carcinogenicity of chemicals are limited. Most of the data available for evaluation are based on animal testing in rats and mice.

3.2.1. Genotoxicity assays

Genotoxicity tests in non-mammalian and mammalian systems are often among the first studies conducted as the toxicological profile of an agent of interest is being developed. These tests are short-term in duration and less resource-intensive than the more complex, longer-term whole animal studies. Companies involved in product development will use the results of genotoxicity studies to determine if there is value in continuing to develop and market potential candidates.

Regulatory agencies, both domestic and international, which have authority to require testing of chemicals in the context of pre-market approval or continuation of existing uses, often specify which kinds of genotoxicity assays must be conducted to fulfill their testing requirements. For instance, the regulatory bodies of the European Union, Japan and the United States, through their International Conference on Harmonization (ICH), have articulated a standard test battery for the testing of pharmaceuticals (ICH, 1997a). It consists of a test for gene mutation in bacteria, an *in vitro* test with cytogenetic evaluation of chromosomal damage with mammalian cells or an *in vitro* mouse lymphoma tk assay and an *in vivo* test for chromosomal damage using rodent hematopoietic cells. Similarly, pesticide regulatory programs in the US and their Organization for Economic Cooperation and Development (OECD) counterparts also require a three-component battery, which varies somewhat from the ICH battery in that they allow some flexibility in the choice of the specific tests to evaluate the potential to cause gene mutations or structural chromosomal aberrations. The third assay is selected "as appropriate for the test substance" and may be characterized by the ability to determine numerical chromosome aberrations, direct DNA damage and repair, mammalian cell transformation or target organ/cell analysis. Eighteen OPPTS harmonized genotoxicity test guidelines are available for use in generating the relevant data (USEPA, 2006). These are compatible with the test guidelines available for use by OECD members.

The (US) National Toxicology Program also includes genetic toxicity testing in its chemical testing program (NTP, 2006). The current NTP genetic toxicity testing program evolved from a broader testing initiative originally developed as a potential predictor of rodent carcinogenicity based on structure–activity relationships and chemical-induced mechanisms of DNA damage. Selection of studies comes from a menu of three *in vitro* assays (*Salmonella*, mouse lymphoma and CHO cell cytogenetics) and two *in vivo* assays (*Drosophila* and rodent micronucleus).

For about 30 years, the US government has conducted a cancer bioassay program, for the first few years managed by the National Cancer Institute and for the past 25 years by the National Toxicology Program housed at the National Institute for Environmental Health Sciences. Well over 500 chemicals, chemical mixtures or physical agents have been tested to date. Most of these have also been examined in one to several genotoxicity screening assays. Periodically, over the past 20 years, assessments of the performance of the more commonly used *in vitro* and *in vivo* assays have been conducted to establish the value and limitations each, individually or in a battery, possesses to predict carcinogenic potential. These evaluations compare the results in the genotoxicity tests with the results in the two-species, two-sex ("four cell") long-term rodent bioassay. Most studies have made the comparisons with agents in the NTP database, although a few studies have used the agents classified for human carcinogenic potential in the IARC monograph series or the Carcinogen Potency Data Base (CPDB) housed at the University of California-Berkeley (Bartsch and Malaveille, 1990; Waters et al., 1999a,b; Ashby and Paton, 1993). (It should be noted that agents assessed by IARC or stored in the CPDB include those tested in the NTP program as well as elsewhere.) All of these studies show the positive value of genotoxicity testing as a tool in the prediction of rodent, and, by extension with appropriate qualification, human carcinogenicity. These evaluations also have served to inform the selection of assay types to be used in the test battery mandated by the various regulatory authorities.

Tennant et al. (1987a,b) were the first to conduct a systematic assessment. Seventy-three chemicals (44 carcinogens, 20 non-carcinogens, 9 equivocal) for which adequate four-cell NTP bioassay results were available were tested in four *in vitro* assay systems: (1) the Salmonella/microsome mutagenicity assay (four or five bacterial strains employed); (2) the mouse L5178Y lymphoma $TK^{+/-}$ forward mutation assay; (3) the chromosome aberrations assay in Chinese hamster ovary (CHO) cells; and (4) the assay for sister chromatid exchange (SCE) in CHO cells. The results indicated that, while overall concordance of the mutagenic response in any *Salmonella* strain with the rodent carcinogenesis response was 62%, the agreement between mutagenicity in *Salmonella* and

tumorigenesis in any one of the four cells of the bioassay was 83%. The investigators concluded that mutagenicity in *Salmonella* indicates a high probability, though not a certainty, of carcinogenicity in rodents. The authors concluded that none of the other three assays, alone or in a battery, could serve in a complementary role to the *Salmonella* assay. None improved performance over that possessed by the *Salmonella* assay alone.

These findings prompted the follow-up evaluation of the performance of several additional assays (Tennant et al., 1987b). Some preliminary results were presented on *in vivo* assays in *Drosophila* and mouse bone marrow, both *in vivo* and *in vitro* unscheduled DNA synthesis assays and mammalian cell transformation *in vitro*. Aside from suggesting that the *Drosophila* test may not be appropriate as a predictor of carcinogenicity, given that it is an *in vivo* germ cell mutagenicity assay where effects are observed only in the offspring, no definitive conclusions about the value of the other assays could be offered. However, they did offer the opinion that results obtained in *in vivo* studies could be highly informative.

Data on an additional 41 chemicals were evaluated by Zeiger et al. (1990), using the same analytical template employed by Tennant et al. (1987a,b). The predictive values of, and interrelationships among, the test results for these 41 chemicals were similar to those previously reported for the 73 chemicals addressed in Tennant et al. and confirm those earlier results.

Additional analysis of 301 chemicals in the NTP database showed that positive results for a chemical in the *Salmonella* gene mutation assay when coupled with having structural alerts to DNA reactivity were highly correlated with carcinogenicity in multiple species/sexes of rodents and at multiple tissue sites (Ashby and Tennant, 1991).

Over time, some genotoxicity assays have fallen out of favor (e.g., the sister chromatid exchange assay) and others have been newly-developed or updated. For instance, a consensus protocol for an *in vitro* micronucleus assay, which can use human or experimental animal cell lines or lymphocytes, has recently been described (Kirsch-Volders et al., 2003). Studies of the correlation between mutagenicity test data and rodent carcinogenicity showed a strong association between clearly positive results in long-term mouse peripheral blood micronucleus tests and rodent carcinogenicity (Witt et al., 2000). Most organic chemicals (other than hormones) identified by the International Agency for Research on Cancer as human carcinogens are genotoxic, and the vast majority of these are detected by both the *Salmonella* assay and rodent micronucleus tests (Shelby, 1988; Shelby and Zeiger, 1990).

A recent evaluation of the performance of a battery of three of the currently most commonly used *in vitro* genotoxicity assays has been published by Kirkland et al. (2005). Their evaluation focused on the ability of the *Salmonella* assay, the mouse lymphoma assay and the *in vitro* micronucleus (MN) or chromosomal aberration (CA) test, alone or in combination, to discriminate rodent carcinogens and non-carcinogens from a database of over 700 chemicals assembled from the NTP Bioassay Program, the CPDB, IARC monographs and other publications. Of the 553 carcinogens for which there were valid genotoxicity data, 93% of the carcinogens gave positive results in at least one of the three tests. Combinations of two and three assays had greater sensitivity than individual assays resulting in sensitivities of around 90% or more, depending on test combination. Only 19 carcinogens (out of 206 tested in all three tests, considering CA and MN as alternatives) gave consistently negative results in a full three-assay battery. Most were either carcinogenic via a non-genotoxic mechanism (liver enzyme inducers, peroxisome proliferators, hormones) or were extremely weak (presumed) genotoxic carcinogens (e.g. *N*-nitrosodiphenylamine). The authors identified 183 chemicals that were non-carcinogenic after testing in both male and female rats and mice. There were genotoxicity data on 177 of these. The specificity of the Ames test was reasonable (73.9%), but all mammalian cell tests had very low specificity (i.e. below 45%), and this declined to extremely low levels in combinations of two and three assay systems. When all three tests were performed, 75–95% of the non-carcinogens gave positive (i.e. false positive) results in at least one test in the battery. Deficiencies remain in our ability to predict the *in vivo* situation from such *in vitro* results. The authors suggest that there may be a need for either a reassessment of the conditions and criteria for positive results or the development and use of a completely new set of *in vitro* tests. The authors concluded that positive results in all three tests indicate the chemical is three times more likely to be a rodent carcinogen than a non-carcinogen. Likewise, negative results in all three tests indicate the chemical is two times more likely to be a rodent non-carcinogen than a carcinogen.

3.2.2. Medium-term whole animal assays

Initiation–Promotion Models: Over the past 20 years, Ito and his colleagues at Nagoya City University in Japan have developed several types of medium-term bioassays utilizing young adult male F344 rats to bridge the information gap between the short-term genotoxicity assay and the long-term chronic/carcinogenicity in rats (Ito et al., 1988a,b,c, 1992, 1994, 1996, 1998a,b, 2000, 2003). These assays are based upon the initiation–promotion model of multi-stage carcinogenesis. Their medium-term assays are considered by the International Conference on Harmonization to be an acceptable second *in vivo* test for carcinogenic potential of a pharmaceutical when coupled, in the basic scheme, with a long-term rodent (usually, rat) carcinogenicity study (ICH, 1997b).

The liver model takes 8 weeks to conduct. It begins with a single 200 mg intraperitoneal injection of *N*-nitrosodiethylamine, followed two weeks later by six weeks' treatment with the test agent(s), presented in the feed or drinking water. In the third week, the rats are subjected to a two-thirds partial hepatectomy. At termination, livers

are examined for the presence of immunohistochemically identified glutathione-*S*-transferase placental form-positive liver cell foci. By 2003, 313 chemicals had been tested for carcinogenic potential in the liver (Ito et al., 2003). Of the chemicals tested, 30/31 (97%) of the mutagenic hepatocarcinogens and 29/33 (88%) of the non-mutagenic hepatocarcinogens tested positive. Ten out of 43 (23%) agents known to be carcinogenic in organs other than the liver were also positive. It is particularly important that only one of 48 non-carcinogens gave a very weak positive result, so that the system has a very low false-positive rate.

The Nagoya group also has developed a medium-term multi-organ assay, dosing adult male F344 rats with five different initiating carcinogens over the first three weeks of the experiment for the detection not only of liver carcinogens, but those that target other organs (e.g., Ito et al., 1998a,b). The animals are sacrificed at the end of week 36. Major organs are examined histologically for indications of carcinogenic potential relevant to the target tissue. Tamano et al. (2001) reported that 17/17 (100%) known liver carcinogens exhibited positive results. In addition, 19/22 (86%) of substances known to be carcinogenic at sites other than the liver were positive. Five non-carcinogens were negative, and for those substances of unknown carcinogenic potential, 9/19 (47%) showed positive results.

Transgenic and neonatal models: As noted above, the ICH supports the principle that short- or medium-term *in vivo* rodent systems can serve as acceptable additional *in vivo* tests for carcinogenic potential of a pharmaceutical when coupled, in the basic scheme, with a long-term rodent (usually, rat) carcinogenicity study (ICH, 1997b). Models using transgenic or neonatal rodents are viewed as suitable candidates. In light of that pronouncement, interest in gaining experience with these models and understanding their strengths and weakness burgeoned. To accelerate this process, the Health and Environmental Sciences Institute of the International Life Sciences Institute ("ILSI HESI") coordinated a multi-year, multi-laboratory collaborative research program to examine several models that could prove useful as a complement to the rat two-year study, or as was hoped by some, an alternative (Robinson and MacDonald, 2001). To maximize reproducibility and comparability of data, standardized protocols were developed for each of the models evaluated. Specific criteria were developed for chemical and dose selection, histopathology review, quality control and evaluation of study outcome. Six *in vivo* (and one *in vitro*) models were studied. The study duration for *in vivo* models generally is 6 months to 1 year. The models included were:

1. The rasH2 transgenic mouse

The rasH2 mouse is a hemizygous transgenic mouse carrying the c-Ha-ras oncogene and that gene's promotor/enhancer within the genetic background of a BALB/cByJ ×C57BL/6J F1 mouse (Tamaoki, 2001). Spontaneous tumors in hemizygous transgenic mice are rare until 6 months of age. The observed rasH2 tumor spectrum, including lung adenoma/adenocarcinoma, forestomach and skin papillomas, Harderian gland adenoma, liver proliferative lesions, splenic hemangioma/sarcoma, and lymphoma is consistent with the BALB/c and C57BL/6 background. In the rasH2 mouse, point mutations of the transgene induced by genotoxins are reported frequently but not in all tumors. Elevated levels of transgene expression were detected in all genotoxin-induced tumors in the rasH2. Increased transgene expression was independent of the mutation rate in transgenic and endogenous ras genes. These observations suggest that the overexpression of transgenic c-Ha-ras is responsible for accelerated tumor development.

Usui et al. (2001) reported their data for compounds tested in this model for the ILSI HESI project as well as some data from other studies. Three of the four known genotoxic human carcinogens tested were positive in the rasH2 mouse (*N*-Methyl-*N*-nitrosurea (MNU), cyclophosamide and phenacetin); melphalan was equivocally positive. Results for the immunosuppressive human carcinogen, Cyclosporin A, were equivocal. Two human hormonal tumorigens, diethylstilbestrol and ^{17}beta-estradiol, gave positive and negative results, respectively. Of the twelve additional compounds tested that are classified as non-genotoxic rodent carcinogens and putative human non-carcinogens, only the two peroxisome proliferators (clofibrate and di(2-ethylhexyl) phthalate (DEHP)) produced a positive response (liver effects). The three non-genotoxic non-carcinogens that were tested (phenobarbital, methapyrilene and reserpine) also gave negative responses. Earlier validation studies with 29 compounds had shown that both genotoxic and some non-genotoxic carcinogens yielded positive responses, but all genotoxic or non-genotoxic substances testing negative as carcinogens in the standard long-term bioassay were negative in this transgenic model (Yamamoto et al., 1998a,b). This result provides confidence that the model is likely to have a low false-positive rate.

2. The Tg.AC transgenic mouse

The Tg.AC (v-Ha-ras) transgenic mouse model provides a reporter phenotype of skin papillomas in response to either genotoxic or nongenotoxic carcinogens (Tennant et al., 2001). The Tg.AC responds to known human carcinogens and does not respond to noncarcinogens. It also does not respond to most chemicals that are positive in conventional bioassays principally at sites of high spontaneous tumor incidence. The mechanism of response of the Tg.AC model is related to the structure and genomic position of the transgene and the induction of transgene expression through specific mediated interactions between the chemicals and target cells in the skin.

Test chemicals are most often applied topically, but oral exposures have been employed as well (Eastin et al., 2001). Skin papillomas are the response of interest. Fourteen pharmaceutical agents administered by the topical and oral routes were selected for study in the ILSI HESI project. The Tg.AC mice seem to respond to topical

application of either mutagenic or nonmutagenic carcinogens with papilloma formation at the site of application. Cyclosporin A, ethinyl estradiol and diethylstilbestrol and clofibrate, were considered clearly positive in the topical studies. In the oral studies, ethinyl estradiol and diethylstilbestrol were negative, cyclosporin was considered equivocal, and results were not available for the clofibrate study. Of the 3 genotoxic human carcinogens (phenacetin, melphalan, and cyclophosphamide), phenacetin was negative by both the topical and oral routes. Melphalan and cyclophosphamide are, respectively, direct and indirect DNA alkylating agents and topical administration of both caused equivocal responses. With the exception of clofibrate, Tg.AC mice did not exhibit tumor responses to the rodent carcinogens that were putative human noncarcinogens, (di(2-ethylhexyl) phthalate, methapyraline HCl, phenobarbital Na, reserpine, sulfamethoxazole or WY 14,643, or the nongenotoxic, noncarcinogenic, sulfisoxazole) regardless of route of administration. It was concluded that the Tg.AC model was not overly sensitive and possesses utility as an adjunct to the battery of toxicity studies used to establish human carcinogenic risk.

3. The $XPA^{-/-}$ single and $XPA^{-/-}/p53^{+/-}$double knockout mouse models

The $XPA^{-/-}$ homozygous knockout mouse with a complete defect in nucleotide excision repair was developed to study the role of this defect in chemical-induced tumorigenesis (van Steeg et al., 2001). These mice develop skin tumors at high frequency when exposed to UV light and also appear to be susceptible to genotoxic carcinogens given orally. In an attempt to further increase both the sensitivity and specificity of the XPA model in carcinogenicity testing, the authors crossed XPA mice with mice having a heterozygous defect in the tumor suppressor gene p53. $XPA/p53^{+/-}$ double knockout mice develop tumors earlier and with higher incidences upon exposure to carcinogens as compared to their single knockout counterparts.

In the ILSI HESI project, 13 pharmaceuticals were tested in the XPA model, and 10 in the XPA/p53 model (van Kreijl et al., 2001). With one exception, all studies were 39 weeks long. The observed spontaneous tumor incidence for the XPA model after 9 months was comparable to that of wild-type mice (total 6%). For the XPA/p53 model, this was somewhat higher (9%/13% for males/females). The 3 positive controls gave positive and consistent tumor responses in both the XPA and XPA/p53 model, but no or lower responses in wild-type mice. From the 13 ILSI compounds tested, the single genotoxic carcinogen (phenacetin) was negative in both the XPA and XPA/p53 model. Positive tumor responses were observed for four compounds (Cyclosporin A, DES, estradiol, and WY 14,643). Negative results were obtained with five other nongenotoxic rodent carcinogens (phenobarbital, clofibrate, reserpine, DEHP and sulfamethazole), and two noncarcinogens (ampicillin and D-mannitol) tested. Combined with previous results, 6 out of 7 (86%) of the genotoxic human and/or rodent carcinogens tested are positive in the XPA model. The positive results obtained with the four nongenotoxic ILSI compounds may point to other carcinogenic mechanisms involved, or may raise some doubts about their true nongenotoxic nature. In general, the XPA/p53 model appears to be more sensitive to carcinogens than the XPA model.

4. The $p53^{+/-}$ knockout mouse model

The heterozygous Trp53 null allele C57BL/6 (N5) mouse is susceptible to the rapid development of neoplasia by mutagenic carcinogens relative to control strains (French et al., 2001). This model demonstrates (1) rapid dose-related induction of tumors (26 wks), (2) multiple sites of carcinogen-specific tissue susceptibility, and (3) carcinogen-induced loss of heterozygosity involving the Trp53 wild-type allele or a p53 haploinsufficiency permitting mutation of other critical protooncogenes and/or inactivation of tumor suppressor genes driving tumorigenesis. Demonstration of mutation or loss of heterozygosity involving the Trp53 locus is consistent with a common finding in human cancers and supports extrapolation between rodents and humans. Almost all mutagenic rodent carcinogens (including all mutagens that are carcinogenic to humans), but not nonmutagenic rodent carcinogens, induce tumors within 26 weeks of continuous exposure. These characteristics and results indicate that the mouse heterozygous for the Trp53 null allele may be of significant use for the prospective identification of mutagenic carcinogens of potential risk to human health.

For the ILSI HESI project, this model was evaluated in 31 short-term carcinogenicity studies with 21 compounds, together with data from other studies which used comparable protocols (Storer et al., 2001). Seventy-five percent (12/16) of the genotoxic human and/or rodent carcinogens tested were positive and the positive control, *p*-cresidine, gave reproducible responses across laboratories (18/19 studies positive in bladder). Cyclosporin A was positive for lymphomas but produced a similar response in wild type mice. Diethylstilbestrol and ^{17}beta-estradiol, gave positive and equivocal results, respectively, in the pituitary with p53-deficient mice showing a greater incidence of proliferative lesions than wild type. None of the 22 nongenotoxic rodent carcinogens that have been tested produced a positive response but two compounds in this category, chloroform and DEHP, were judged equivocal based on effects in liver and kidney respectively. Negative results were seen with four genotoxic noncarcinogens and six nongenotoxic, noncarcinogens. In total (excluding the compounds with equivocal results), 42/48 compounds (88%) gave results that were concordant with expectations.

5. The neonatal mouse

The neonatal mouse model has been used experimentally since 1959 (Pietra et al., 1959). The neonatal model is known to be very sensitive for the detection of genotoxic carcinogens but not responsive to non-genotoxic carcino-

gens and non-carcinogens, yielding a high sensitivity and specificity in its response. Dose selection for the neonatal model is based on the maximum tolerated or feasible dose. Traditionally, compounds have been tested via the intraperitoneal route of administration which limits the amount of material that can be administered. For the ILSI HESI project, the model was adapted for oral administration, the same route by which most pharmaceuticals are administered (McClain et al., 2001). Also, this permits much higher doses to be used as compared to the IP route of administration. Administration of the test agent at the maximum tolerated or feasible dose occurs on postnatal days 8 and 15. The study is terminated when the animals reach one year of age. The spontaneous tumors in the neonatal model occurred mainly in the liver of male mice and lung of male and female mice with a few tumors observed in the Harderian gland. The positive control, DEN, produced a robust, uniform, and reproducible tumor response with the target organs essentially limited to liver and lung. A total of 13 compounds out of the 21 compounds were evaluated in the neonatal model involving 18 studies with duplicate studies for some compounds. The genotoxic carcinogens including those used as positive controls were clearly positive. The non-genotoxic rodent carcinogens were clearly negative. The non-genotoxic human carcinogen (cyclosporin) was clearly negative. The two other human carcinogens (phenacetin and DES) were negative but estradiol was negative in one of the two oral studies, but was clearly positive in the other. Considering the mode of action for three of the human carcinogens (DES, cyclosporin and phenacetin), which were negative in this model, the mode of action in humans is likely to be epigenetic. Overall, for the three clearly genotoxic chemicals, all were positive. For the nine clearly non-genotoxic chemicals, all nine were negative. The two human carcinogens for which genotoxicity may or may not play a role (DES and phenacetin) were negative and estradiol was positive in one of the two oral studies.

In an earlier evaluation of the model, a total of 45 known carcinogens (rodent and human), including two aromatic hydrocarbons, five aromatic amines, three azo dyes, ten nitroso compounds, three steroids, four tryptophan metabolites and their related compounds, four naturally occurring substances, four pyrolysates of amino acids and ten miscellaneous compounds, were tested (Fujii, 1991). Of the 45, 28 chemicals showed positive results; the remaining 17 chemicals were found to be negative for tumor development. The correlation of the results between the neonatal model and the long-term mouse and/or rat carcinogenesis tests were compared with 37 chemicals tested in the traditional bioassay; the remaining eight chemicals were tested only in the neonatal model. Thirty-one of the 37 chemicals (83.8%) tested in the neonatal model showed similar carcinogenic or non-carcinogenic results obtained in the long-term rodent bioassay. Twenty-nine out of 34 chemicals (85.3%) showed similar results to the adult mouse. Contradictory results were obtained with the following five chemicals: N-hydroxy-acetylaminobiphenyl, 3-hydroxy-anthranilic acid, 3-hydroxy-L-kynurenine, isonicotinic acid hydrazide and phenobarbital. There were 35 chemicals which were comparable with the results of the adult rat, and 32 chemicals showed the same results as the neonatal model (91.4%). Dissimilar results were obtained with estradiol, 3-hydroxy-anthranilic acid and phenobarbital.

3.2.3. Studies to characterize carcinogenic model/mechanism of action

Although still limited, there is an increasing understanding of the modes and/or mechanisms that underlie the carcinogenic process. Such information is becoming ever more important to the entire process of carcinogenicity evaluation, risk assessment and decision-making. It can be of value in predicting whether or not an agent will be carcinogenic when tested in long-term bioassays, in interpreting the human relevance of a positive result in a long-term bioassay and in informing the shape of the dose–response curve and subsequent selection of the appropriate quantitative method for estimating risk to human health, when the mode/mechanism of action is deemed relevant to humans.

Historically, carcinogens have been divided, for pragmatic reasons, into two categories, genotoxic and non-genotoxic, even though there is no universally accepted definition for the two terms. The present discussion defines genotoxic carcinogens as being directly-DNA reactive and non-genotoxic carcinogens (also often referred to as epigenetic) as being not directly-DNA reactive. Most identified human carcinogens usually react with DNA directly through electrophile generation or occasionally through oxidative stress (Clayson and Kitchin, 1999). The approaches to the prediction of risk from exposure to carcinogens have been described for carcinogens with genotoxic and non-genotoxic activities (IARC, 1999; Fan and Howd, 2001; Faustman and Omenn, 2001; Gaylord et al., 1999; Zeise, 2001; USEPA, 2005a).

Increasing interest is being focused on non-genotoxic agents that are determined to be positive in rodent carcinogenicity bioassays. Because their mode/mechanisms may be due to target organ toxicity and subsequent cell proliferation, induction of hormonal imbalance or some other way which does not include direct interaction with DNA, or some other mode/mechanism exhibiting a non-linear dose response, the question has arisen as to whether they should be assumed to be carcinogenic at low doses. For example, saccharin would not be expected to increase bladder tumor production at doses that do not cause cytotoxicity, and phenobarbital would not be expected to cause liver cancer at doses that do not affect cellular growth control (Butterworth, 1990). In the case of saccharin, high concentrations of the compound in the urinary bladder of male rats lead to bladder tumors, following crystalline precipitation and local irritation that produce chronic regenerative hyperplasia (Faustman and Omenn, 2001). Phenobarbital is a non-DNA reactive liver carcinogen that acts by altering growth control, or increasing mitogenesis.

Greater attention is given in risk assessment (and, regulatory decision-making, for that matter) to chemicals acting by a mode/mechanism that is of relevance to human biology (and, by default, to those chemicals whose mode/mechanism is unknown). Many nongenotoxic carcinogens may not present a significant human risk because the effects are either rodent-specific or a high-dose, long-duration exposure would be required that would not represent a feasible human exposure scenario in order to elicit the cellular effects leading to carcinogenicity (Williams and Iatropoulos, 2001). Examples include: male rat urinary bladder tumors with cytotoxicity and reactive hyperplasia from precipitated chemicals, male rat kidney α_{2u}-globulin nephropathy-mediated increases in kidney neoplasms, rodent forestomach tumors, tumors in endocrine organs (thyroid gland tumors with sustained excessive hormonal stimulation), and receptor-mediated carcinogenesis (e.g., rodent peroxisome proliferation-associated with the onset of liver cancer).

Animal studies can be tailored to study the hypothesized mode/mechanism of action. They generally are repeat dose studies of subacute to subchronic duration (i.e., 14–90 days).

Data required to "prove" that a specific mode/mechanism underlies the tumor response observed have been articulated for a number of tumor types: e.g., the alpha 2u-globulin-induced kidney tumor in the male rat (USEPA, 1991), thyroid tumors in rats (USEPA, 1998c), peroxisome proliferator alpha-induced rodent liver tumors (Klaunig et al., 2003), bladder tumors induced by urinary tract calculi (Meek et al., 2003). Frameworks by which an assessor can answer the question as to whether or not the mode(s) of action has been adequately characterized for the tumor observed in the animal bioassay and, if so, whether or not it is relevant to the human have been developed by IPCS (Sonich-Mullin et al., 2001), ILSI Risk Science Institute (Meek et al., 2003) and USEPA (USEPA, 2005).

Knowledge on the mode/mechanism of action and the robustness of the data set affect the choice of the approach to use for hazard characterization (dose–response assessment), e.g., a biologically-based model or a default procedure, linear or non-linear. Probabilistic estimates of low-dose risk or a reference dose/concentration may be calculated. The quality of data, assumptions and uncertainties and possible alternative approaches are to be considered. Continued study of the modes/mechanisms by which chemicals exert their toxicity would help to enhance the mechanistic database and to improve our ability to differentiate among the more and less significant chemical hazards when assessing potential risks to humans. In cases where inadequate data exist on a chemical, a decision may still have to be made regarding the relevance of the tumor etiology to humans based on the existing knowledge and level of certainty when evaluating the data. While mechanistic data can aid in the interpretation and extrapolation of exposure to dose, the full impact of biological plausibility and coherence must be considered in the overall evaluation of the weight of evidence.

3.2.4. Long-term bioassays

For many years, the standard two-year rodent bioassay using both sexes with two rodent species (rat and mouse) has been considered the definitive experimental study in the National Toxicology Program (NTP) for evaluating carcinogenicity of chemicals and the principal means of identifying trans-species carcinogens (Bucher and Portier, 2004). This also remains standard practice for regulatory agencies with testing authorities, although, as noted above, pharmaceutical manufacturers may choose to conduct one or more shorter-term *in vivo* animal studies in lieu of a second long-term bioassay. Some may take issue with describing a long-term rodent bioassay as a "screen," given its cost, complexity and duration. However, carcinogenicity is a case in which observation of the actual endpoint of interest (tumor incidence in humans exposed for sufficient periods to cause cancer) is generally not possible and it is therefore necessary to rely on "screens" of increasing complexity from SAR to *in vitro* tests to animal cancer bioassays. None of these substitute assays is, in fact, a fail-safe predictor of *human* carcinogenic potential, but one of several types of animal studies that serve to shed light on an agent's human hazard profile.

Testing for carcinogenic potential is done in a moderate number of experimental animals (generally 50 or less per sex per dose group). For practical reasons, it is impossible to conduct studies of a size (tens of thousands of animals) that would allow observation of effects following exposure at relatively low doses or to directly estimate the risk at low exposure levels in a large population. Therefore, animal studies have been conducted at relatively high dose levels that are generally orders of magnitude greater than those likely to be encountered in the environment by humans. Because of this, the study results then must be extrapolated from high to low doses and from animals to humans. High level dosing in animals can affect the body's response to the test chemical, e.g., saturation of the metabolic pathways, either detoxification or metabolic activation, and thus affect the final outcome. From a public health standpoint, most of the regulatory agencies in the US have generally regulated carcinogens that are genotoxic or of unknown mode/mechanism at exposure levels that reflect a very low probability of tumor production (e.g., theoretical lifetime excess cancer risk from 1 in 1 million to 1 in 10^5 individuals).

Chemicals tested under USEPA or OECD guidelines are evaluated in two-year studies in the rat and 18-month studies in the mouse. Some researchers have recommended limiting study duration to no more than 18 months in the rat also, based on an evaluation of the results of 210 studies of 24-months' duration in IARC monographs, in which the tumor status of the animals alive at 12 or 18 months was unknown, with the speculation that the magnitude of tumors observed at 24 months suggested that the effect

would have been apparent many months earlier (Davies et al., 2000). However, evidence to support this was not provided. Other researchers have recommended extending the duration of rodent studies to 30 months or a lifetime to increase the sensitivity of the assay (Haseman et al., 2001; Anisimov, 1998; Huff, 1999; Maltoni et al., 1999), unless a shorter duration reflects the natural lifespan of the animal model (Haseman et al., 2001). The recommendation for an extended duration of 30 months or a lifetime was supported by the observation that the majority of neoplasms present prior to 18 months were not visible on gross examination, based on data from interim sacrifices in studies from the NTP for 43 chemicals, with 50 biologically significant carcinogenic effects observed in male (30) and/or female rats (17).

To examine the ability of animal bioassays to predict carcinogenic hazard, data from studies on the carcinogenicity of 471 test substances published in technical reports by NTP as of July 1, 1998, were reviewed for the frequency of positive outcomes, and the interspecies correlation in carcinogenic response between rats and mice (Haseman, 2000). Of the 471 studies, 250 (53%) were positive in one or more sex-species groups, 64 (14%) gave equivocal results, 150 (32%) were negative in all sex-species groups, and 7 (1%) were considered inadequate for evaluation. For those 385 studies in which adequate results were obtained for all four sex-species groups, concordant results in rats and mice (i.e., both species were positive or negative) were found for 283 studies (74%), whereas 47 (12%) were carcinogenic only to rats and 55 (14%) were carcinogenic only to mice. This concordance rate of 74% in rats and mice was reported to be in agreement with earlier estimates based upon smaller data sets (Huff et al., 1991; Gold and Slone, 1993; Lai et al., 1994; Haseman and Huff, 1987).

Haseman and Lockhart (1993) concluded that most target sites in cancer bioassays showed a strong correlation between males and females (65%, 97/146 for rats, 103/159 for mice) especially for forestomach, liver and thyroid tumors. However, the exact target organ correlation is considered less critical than whether or not a chemical is a potent rodent carcinogen (Benigni and Giuliani, 1999). Therefore, tumors observed at the more unusual sites, e.g., the pituitary gland, the eighth cranial nerve or the zymbal gland, should not be dismissed as irrelevant, since organ correlation is often lacking, i.e., between rodent species and between rodents and humans (NRC, 1994; cited by Faustman and Omenn, 2001).

For most known human carcinogens that have been tested adequately in laboratory animals, the findings of carcinogenicity are concordant (Huff, 1993, 1994, 1999). One long-standing exception was arsenic, for which attempts to test for carcinogenic potential in adult animals have yielded primarily negative results. Transplacental carcinogenesis in mice following prenatal exposure to inorganic arsenic was reported many years after human evidence of carcinogenicity (Waalkes et al., 2003). Huff (1999) listed 46 agents for which the evidence of carcinogenicity was first observed in experimental animals and subsequently by human epidemiologic evidence. The list is limited by the fact that humans may be exposed to agents many years before any testing or epidemiologic studies have been initiated or completed, and/or that the agents might not be human carcinogens under any real-world exposure scenario. In some cases, such as commodity (industrial) chemicals, bioassays may not be done until after many years of production and use, if at all.

A comparison of carcinogenic potencies estimated from human epidemiologic data was made with those estimated from animal carcinogenesis bioassays. The 23 chemicals selected had reasonably strong evidence of carcinogenicity in humans or animals and suitable data for quantifying carcinogenic potencies in both animals and humans. The potencies were highly correlated with correlation coefficients ranging as high as 0.9 (Allen et al., 1988).

3.2.5. Carcinogenic risk characterization and decision-making

Risk characterization is the (final) step in risk assessment in which all of the information available on hazard and exposure is synthesized and the nature and magnitude of the risk in humans is described. It also should include a description of the strengths and weaknesses of the data, and the uncertainties and assumptions (including default options) applied. Its ultimate use is as a tool to guide managers in their decision-making (e.g., chemical product testing or production, or regulatory decision making).

3.2.5.1. Case examples – decisions in the absence of 'complete' data. There may be a need to make a decision before all of the desired data, including those from the long-term bioassay, are available. The challenge becomes one of using incomplete or inadequate data without compromising quality and predictivity. These situations may involve new product development (and products in use) for production-type companies and regulatory decision-making by government agencies.

The following examples describe situations in which carcinogenicity and other relevant toxicity or exposure data were considered incomplete for the chemicals involved, but a risk management decision had to be made. The question common to the three examples described was the identification of the occurrence of residues of the chemical in food or drinking water, and whether that food or drinking water was safe for consumption. Existing carcinogenicity data were evaluated using (then) current and evolving scientific practice and used in the conduct of risk assessments, which in turn provided a guide for decision-making. In some cases, a range of susceptibilities and assumptions were considered to reflect the assessment of the risks to susceptible individuals (e.g., infants and children).

3.2.5.1.1. Ethylene dibromide (EDB) in cereal grains and bakery products. Problem formulation: In 1984, residues of the fumigant EDB were found in stored cereal grains and bakery products. No tolerances had been established

during its pre-market approval process because the chemical was initially thought not to leave a residue because of its high volatility. The risk assessment question to be answered was: "What is the nature and extent of the potential health risks associated with consumption of products containing defined levels of EDB residues?"

Review of existing data: Two oral gavage carcinogenicity studies of EDB had been conducted by the NCI (1978). In rats, there were significantly increased incidences of neoplasia at several sites, i.e., squamous cell carcinoma of the stomach, hepatocellular carcinoma and hemangiosarcoma of the circulatory system. None of these tumors occurred in the controls. In mice, there was a significant increase in the incidence of squamous cell carcinoma of the forestomach and of alveolar/bronchiolar adenomas of the lung. Inhalation bioassays in rats and mice showed statistically significant increases in adenocarcinomas and adenomas of the nasal cavity in low dose male and female rats, and alveolar/bronchiolar adenomas and carcinomas in the high dose female rats and mice of both sexes (NTP, 1982). Other tumors observed included hemangiosarcomas of the circulatory system and mesotheliomas of the tunica vaginalis in the high dose male rats, and fibroadenomas of the mammary gland in females. A follow-up inhalation study in mice confirmed the earlier observation of nasal tumors in female mice and nasal hyperplasia in both sexes.

US EPA and others were challenged on the decision to use the data on the forestomach tumors from the gavage study as the basis for the cancer risk estimate. Data analysis showed that tumors had also occurred at other sites. A mathematical model to incorporate early-in-life exposure was attempted but not fully developed and used.

Genotoxicity data also were available at the time this situation arose. EDB was shown to be positive in several strains of *Salmonella* (Barber et al., 1981; Carere and Morpurgo, 1981). Metabolic activation was not required. It also was shown to produce positive responses in chromosomal aberrations and sister chromatid exchanges in CHO cells (Tan and Hsie, 1981; Brimer et al., 1982) and Drosophila (Kale and Baum, 1979; Kale and Baum, 1983).

Development and implementation of strategic plans: US EPA (and others) focused on the carcinogenic potential of EDB in developing the risk assessment. The body of evidence included an inadequate epidemiologic study that showed no evidence of increased tumor or mortality rates in occupationally-exposed workers, positive rodent bioassays via different routes of exposure and some knowledge of mode of action (positive genotoxicity/mutagenicity studies). The question of age-related susceptibility was not fully resolved.

Outcome: EDB bioassay data showed that the chemical was a multi-site, multi-species rodent carcinogen. These results along with the positive genotoxicity data prompted a high degree of concern about potential human risk from consumption of products found to contain EDB residues, particularly since many of these products were consumed by children. In its initial action, the US EPA established three (temporary) permissible levels (tolerances) for previously-treated grains, bakery products and baby foods, to allow these products to move through the channels of trade, while disallowing its continued use prospectively. These temporary tolerances also were adopted by many states although California set a zero tolerance for baby foods. The US EPA subsequently cancelled all uses of EDB.

3.2.5.1.2. Methyl tertiary-butyl ether (MTBE) in drinking water. Problem formulation: MTBE has been used as a gasoline additive and a synthetic solvent primarily as an oxygenate in unleaded gasoline to boost octane and improve combustion efficacy by oxygenation (OEHHA, 1999). The US EPA first allowed the use of MTBE as a gasoline additive in February 1979 at 7% (v/v) or less. This ruling was amended in October 1980 to permit 11% (v/v) MTBE in the finished gasoline. MTBE is highly-water soluble and persistent. MTBE was found to be an increasingly frequent contaminant of groundwater, including that which serves as a source of drinking water. There is no national drinking water standard for this substance. The risk assessment question to be answered was: "What are the nature and extent of the potential health risks associated with consumption of, and contact with, drinking water containing MTBE?"

Review of existing data: Between 1991 and 1999, there was considerable discussion by various scientific bodies regarding the human carcinogenic potential of MTBE. The controversies focused on the quality and human relevance of the existing animal carcinogenicity data. Inhalation exposure to MTBE produced increased incidences of kidney and testicular tumors in male rats and liver adenomas in female mice. Oral gavage exposure produced leukemia and lymphoma (combined) in female rats and testicular tumors in male rats. While there were varying degrees of uncertainty and differences of opinion as to the relevance to human cancer causation for each of the tumor types induced by MTBE in rodents, some would argue that the occurrence of the tumors at all of these sites adds considerably to the weight of evidence supporting the conclusion that MTBE should be considered a possible human carcinogen. Five of the six assays in three studies were positive, showing tumors in two species, in both sexes, by oral and inhalation exposures. A report prepared by the University of California mandated under SB521 concluded that MTBE is an animal carcinogen (UC, 1998). The California Environmental Protection Agency concluded that MTBE is an animal carcinogen and developed a public health goal for MTBE in drinking water based on carcinogenic risk. In 1997, the Office of Water at US EPA developed a Drinking Water Advisory based upon the organoleptic (taste and odor) characteristics of MTBE. The US EPA has not yet developed an oral Reference Dose or performed a cancer risk assessment for MTBE, although in December 1998, it was said to be undergoing (re)assessment in the IRIS program.

Development and implementation of strategic plans: The US EPA's Office of Water concluded that the available data were not adequate to estimate potential health risks of MTBE at low exposure levels in drinking water, but that the data support the conclusion that MTBE is a potential human carcinogen at high doses. It also indicated that ongoing work by the agency and other researchers was expected to help determine more precisely the potential for health effects from MTBE in drinking water.

Outcome: Two environmental agencies reached a seemingly different determination on the carcinogenicity potential of MTBE at low exposure levels in drinking water, but were in agreement at high doses. No epidemiologic studies that evaluate the potential for human carcinogenicity are available. The mode of action is not fully known. The difference of opinion concerning carcinogenic potential had minimal impact in practice in that the two agencies have developed roughly comparable advisory levels for drinking water (13 ppb vs. 20–40 ppb) (OEHHA, 1999).

3.2.5.1.3. Chloroform in drinking water. Problem formulation: US EPA's Maximum Contaminant Level Goal (MCLG) of zero for chloroform established in 1998 was removed from the final Stage 1 Disinfectant and Disinfectant By-products Rule, in accordance with a court order, effective May 30, 2000. The MCLG was challenged successfully based on the assertion that the US EPA had not used the best available peer reviewed science to set the MCLG as required by the Safe Drinking Water Act. The court found the scientific evidence to indicate that chloroform produces its carcinogenic effects only when exposure exceeds a minimum or threshold dose. The risk assessment questions to be answered were: "Is/are the mode(s) of action for the carcinogenic activity in animals adequately understood?" "Is/are the mode(s) of action defined in the test animals relevant to humans?" "Can the shape(s) of the dose response curve(s) be characterized adequately?" "What is the appropriate approach to dose extrapolation to characterize the risk to humans at known/estimated levels of exposure?"

Review of existing data: Chloroform is a carcinogen in rats and mice, producing kidney and liver tumors following ingestion. The US EPA originally concluded that chloroform was a Group B2 probable human carcinogen and derived an oral cancer slope factor of 0.0061 (mg/kg day^{-1}) and a drinking water unit risk of 1.7×10^{-7} (μg/L)$^{-1}$. More recently (2001), the Agency reassessed the original data along with newer data which characterized the mode(s) of carcinogenic action in the test species. It concluded that the mode(s) of action were understood and that they were relevant to humans. In addition, EPA concluded that the scientific evidence supported the conclusion that there would be no differences in expected responses based upon life stage, and that the margin-of-exposure approach, rather than a linear-at-low-dose extrapolation procedure, was more appropriate for the estimation of carcinogenic risk in humans. Further, they stated that a dose of up to 0.01 mg/kg bw /day (equal to the oral RfD) can be considered protective against cancer risk. IARC has classified chloroform as a Group 2B chemical, meaning that it is possibly carcinogenic to humans and that there is sufficient evidence in experimental animals for carcinogenicity. The carcinogenic evaluation and prediction were based on animal data in the absence of human data.

Mode of action: It has been proposed by some scientists that tumors occurring in the kidney and liver in chloroform-treated animals were due to (or secondary to) cytotoxicity and tissue regeneration (ILSI, 1997; USEPA, 1998d). This would mean that no tumors would occur at exposure levels where no cytotoxicity and tissue regeneration occurs. If this is the case, then a dose–response assessment should focus on determining the dose where cytotoxicity, tissue regeneration and tumor formation do or do not occur.

Others have argued that chloroform carcinogenicity may involve multiple mechanisms. The finding of a 3-fold increase in micronucleated kidney cells in rats exposed orally to a high dose of chloroform is of particular interest because the effect reported was in the species and tissue most relevant to cancer risk assessment (Robbiano et al., 1998). Other data include: Dose dependent findings of sister chromatid exchange (*in vivo*) in mouse bone marrow (Morimoto and Koizumi, 1983), chromosomal aberrations *in vivo* in rats treated orally or intraperitoneally with chloroform (Fujie et al., 1990), positive mouse micronuclei assay with chloroform (San Agustin and Lim-Sylianco, 1978), dose-related induction of intrachromosomal recombination in yeast (Brennan and Schiestl, 1998) and reduction in recombination when the assay was performed in the presence of a free radical scavenger, and DNA binding *in vivo* (Colacci et al., 1991). It should be noted, however, that while the available data on mutagenic potential are mixed (both positive and negative), the majority were negative, and some of the positives occurred only at extreme exposure conditions. EPA has concluded that the weight of evidence indicates that chloroform is not a strong mutagen and that neither it nor its metabolites readily bind to DNA (USEPA, 2001).

In 1996, the US EPA cosponsored an International Life Sciences Institute project in which an expert panel reviewed data on the carcinogenicity of chloroform and considered how endpoints related to its mode of carcinogenic action can be applied in the hazard and dose–response assessment (ILSI, 1997). In 1998, the US EPA conducted a reassessment of chloroform and concluded that the chloroform dose–response should be considered nonlinear (USEPA, 1998e). In its dose–response assessment, the US EPA used two approaches to estimate the MCLG for chloroform: the LED_{10} for tumor response, and the reference dose for hepatotoxicity. The linearized multistage approach was considered but not used. A level of 0.3 mg/L was considered to be the appropriate MCLG based on hepatotoxicity which would be equivalent to a 5×10^{-5} cancer risk level, if calculated using a linearized multistage model. However, as noted above, in its final rule, the US EPA's risk managers

considered application of this approach to be premature, ostensibly because no EPA-sponsored external technical peer review had yet occurred. The proposed rule retained the proposed MCLG of zero.

Development and implementation of strategic plans: In 2001, EPA drafted a proposed Stage 2 Disinfectants and Disinfectant By-products Rule. It was published for public comment on August 18, 2003. One element in the draft rule includes a proposed MCLG for chloroform of 0.7 mg/L. This value is based upon the Reference Dose of 0.01 mg/kg bw /day and the assumption that a person drinks 2 L of water per day (90th percentile), weighs 70 kg, and has a relative source contribution of 20 percent. This calculation reflects the implementation of the margin-of-exposure approach to cancer risk estimation noted above. EPA finalized this rule in 2005.

Outcome: Following the court challenge, the US EPA removed the zero MCLG for chloroform from its National Primary Drinking Water Regulations, effective May 30, 2000, and proposed a non-zero MCLG of 0.7 mg/L in its Stage 2 Disinfectants and Disinfectants By-products Rule in 2003.

3.3. Immunotoxicity

The immune system is a potential 'target organ' for toxic damage. Toxic responses may occur when the immune system is the target of chemical insults, resulting in altered immune function; this in turn can result in decreased resistance to infection, certain forms of neoplasia, or immune dysregulation or stimulation which exacerbates allergy or autoimmunity. Alternatively, toxicity may arise when the immune system responds to the antigenic specificity of the chemical as part of a specific immune response (i.e. allergy or autoimmunity) (IPCS, 1996). The major routes of exposure that elicit an immune reaction to a chemical are skin contact, pulmonary exposure (inhalation), and exposure by the oral route (gastrointestinal tract). Animal studies can be used to predict potential toxic effects of chemicals on the human immune system.

3.3.1. Case example: allergic contact dermatitis

Problem formulation: There is a need to be able to predict the potential of chemical agents (such as industrial chemicals, consumer products, and drugs) to produce Allergic Contact Dermatitis (ACD) in humans before these compounds are marketed. For decades, guinea pig (GP) assays were the standard used to assess the ACD potential of these agents. Assays like the Guinea Pig Maximization Test (GPMT) and the Buehler Assay (BA) were recognized as the standard assays for the determination of ACD using laboratory animals. However, the GP assays have a number of shortcomings associated with their use as effective screening assays for ACD. Therefore the question to be addressed was "Can we develop a more efficient quantitative screening assay for ACD which will provide the characteristics of a good screening tool?"

Review of existing data: In highly experienced hands GP assays have considerable credibility, but they are subject to relatively high false positive and false negative results. Interpretation of the results requires experience and expertise and follow-up testing in humans is sometimes required. Because the GP assays evaluate the elicitation phase of ACD, they have an experimental time frame consisting of weeks, and require substantial resources.

Development and implementation of new screening assay: The ultimate goal of any animal toxicity test is to correlate with and be predictive of toxicity in humans. The murine local lymph node assay (LLNA) was developed to evaluate the potential of dermally applied compounds to induce ACD in humans. The LLNA generates quantitative data and provides a dose–response assessment as a routine part of the standard protocol. Because it assesses the induction phase of ACD and not the elicitation phase, as evaluated by the GPMT and BA, the LLNA reduces animal distress. The LLNA potentially reduces animal numbers and is conducted in a relatively short time frame (5 days) as compared to the GP assays (weeks). Additional advantages of the LLNA are that it is more cost effective and less difficult to assess the outcome than are the GP assays.

Many studies have reported that the LLNA is a reliable method for evaluating the potential of dermally applied compounds to produce ACD in the mouse (Kimber et al., 1995; Loveless et al., 1996; Basketter et al., 1996; Kimber et al., 1998). The LLNA produces quantitative results and is capable of identifying compounds that are moderate to strong sensitizers (Kimber et al., 1995). The ICCVAM report indicated that when the LLNA was compared with the GPMT/BA for 97 chemicals, there was a concordance rate of 89% (ICCVAM, 1999). In this analysis, the study considered sensitivity, specificity, positive prediction value, and negative prediction value. When the LLNA was compared to all GP tests using a total of 126 chemicals, the concordance was 86%. Comparisons of both the GPMT and BA with human studies using a total of 57 and 62 chemicals, gave concordance values of 72% and 73%, respectively, while comparisons of the LLNA with studies on humans, using 74 chemicals, gave a similar concordance of 72%. Thus, the LLNA produces reproducible, reliable, and quantitative results in a relatively short time frame using fewer resources when compared with the assays that it was designed to replace. Therefore, the LLNA possesses numerous characteristics that are required of a good screening assay. When comparing the LLNA with the traditional GP methods, the ICCVAM concluded that it appears to provide an equivalent prediction of the risk for human ACD.

While the LLNA has proven to be a robust and reliable screening assay, it does have a number of limitations. Strong irritants such as sodium lauryl sulfate can produce positive results in the LLNA (Gerberick et al., 1992). Thus, the ability to differentiate between weak sensitizers and strong irritants cannot consistently be accomplished with the LLNA alone. There is increasing evidence that with the addition of phenotypic analysis of the draining lymph

node cells, strong irritants and weak sensitizers may be differentiated with the LLNA (Gerberick et al., 2002). Another limitation of the LLNA is that the compound of interest must be soluble in a solvent that is also innocuous in the LLNA. Numerous solvents have been used in the LLNA and shown to be non-reactive. However, one of the most versatile solvents dimethylsulfoxide (DMSO) can cause skin irritation and produces a weak signal in the LLNA. When DMSO is chosen as a vehicle for the LLNA, classification of a compound as a sensitizer when the results are close to the agreed-upon standard for a positive response (a threefold increase in proliferative cells) may be in question. A compound may produce a false positive result and could not be classified as a sensitizer or an irritant because the irritation induced by DMSO could produce an effect on the response induced by the compound of interest. Additionally, compounds that are only soluble in water may not be readily absorbed and care must be taken during the administration of test material to ensure that the test material has been absorbed.

Outcome: The LLNA was the first assay to be considered by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) in a report that was published in 1999 (NIH Publication No. 99-4494). The LLNA was recommended by ICCVAM as a standalone assay for the determination of contact hypersensitivity and is now accepted by the Organization for Economic Cooperation and Development (OECD), USEPA, and USFDA. Acceptance of a newly developed screening assay by international regulatory organizations is testimony to a screening assay's abilities. Additionally, the incorporation of the murine LLNA into the human risk assessment process for potency prediction of sensitizers in addition to its current role as a hazard identification tool has been discussed (Gerberick et al., 2001; Kimber et al., 2002; Scott et al., 2002). While the correlation with human data is high, the LLNA may not predict human sensitizers that have been reported to induce ACD after a prolonged exposure period, as in transdermally delivered pharmaceuticals (Llewellyn et al., 1999). In the cases of intentional prolonged exposure, additional parameters such as murine lymph node cell phenotyping and/or cytokine production or human skin patch testing may provide evidence that a compound has the potential to induce ACD (Gerberick et al., 2002; Humphreys et al., 2003; Dearman et al., 2003). The murine LLNA is a success story in the context of the development of a screening assay with exceptional hazard identification properties, short time frame duration, and effective use of resources. Additionally, with modification, the LLNA may prove to be a useful tool for more intensive investigative analyses of contact sensitizers.

3.4. Endocrine disruptor screening program (EDSP)

The Endocrine Disruptor Screening Program arose from recommendations provided to the US EPA by the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC), a multi-stakeholder committee convened by EPA to help the Agency develop a program to comply with the endocrine disruptor screening requirements of the Food Quality Protection Act and the Safe Drinking Water Act of 1996. EDSP is intended to be a screening level assessment of the potential of an agent to affect estrogen-, androgen- or thyroid hormone-mediated function. The examples provided in this report show how an agent may be identified as an endocrine disrupter, but they do not address how the EDSP will be used in risk assessment, as this is still under discussion.

EDSTAC (1998) recommended that EPA take a tiered approach to endocrine disruptor assessment, with decisions being made at each tier as to whether to continue into the next tier, or whether the information from the completed tier is sufficient to support a decision on whether the agent does or does not have the potential to be an endocrine disruptor. Given that it would be unlikely to further evaluate a compound that had no discernible endocrine activity in a screening tier, the screening tier is weighted towards producing false positives.

The EDSP consists of three tiers: prioritization, screening, and testing. Prioritization consists of determining, from the universe of chemicals, which ones should be advanced more quickly into the assessment program. The criteria for prioritization fall into the categories of potential for exposure and potential to produce an effect. Exposure considerations include production volume, potential for widespread exposure (i.e., a large fraction of the population may be exposed), environmental persistence, and the potential to be exposed to a lot of the material, either through use or bioaccumulation. Potential for biological activity may be determined by a review of the literature, limited *in vitro* screening, or structure–activity relationships. A great deal of work has gone into developing QSARs for estrogen receptor binding (Serafimova et al., 2002; Shi et al., 2001). If they, and companion programs for androgen receptor interaction, can be shown to be relatively predictive, they will become the key elements of the prioritization tier.

Chemicals identified as being of high priority for further assessment enter the screening tier. The screening tier was designed to identify agents acting via any known mechanism of action of environmental estrogens, androgens or thyroid toxicants. The known mechanisms include:

- Estrogen receptor agonism or antagonism.
- Androgen receptor agonism or antagonism.
- Steroid synthesis inhibition.
- Aromatase inhibition.
- 5-Alpha-reductase inhibition.
- Decreased circulating thyroid hormone concentration.
- Interference with the hypothalamic–pituitary–gonadal axis.

There may be additional mechanisms (some have been hypothesized) but these are the only ones that have been documented.

A battery of tests was identified to cover these mechanisms, and the possibility exists of developing alternative screening methods as long as the spectrum of mechanisms is evaluated. The minimum data set would be one in which all of the known mechanisms are covered, but it is preferred that the battery of tests have some overlap such that most of the known mechanisms are evaluated more than once. There were a number of reasons for this preference. The first was to overcome the dilemma as to whether screening level assays should be apical or reductionist. Apical tests, in which one evaluates one or more outcomes that may be caused by a particular mechanism, but are not necessarily specific to that mechanism, have the advantage of evaluating higher order functions and are closer to traditional toxicology tests. However, depending on the endpoints that are measured, they may have little relevance to the question being asked. For example, decreased body weight gain may be an effect of developmental exposure to thyroid-active agents, but it is clear that only a small fraction of the agents that have been shown to decrease body weight do so by interfering with thyroid hormone levels. More fundamental assays, such as receptor binding assays, are highly specific, but do not necessarily inform about the potential for an effect on reproduction *in vivo*. The second reason was to make it possible to include assays from multiple species, an important consideration given that endocrine disruption is a potential concern for ecological health as well as human health. A related reason is that employing multiple, overlapping assays allows one to cover additional life stages, although the original EDSP screening system is not optimized to cover mammalian development.

The prototype EDSP battery consists of receptor binding or transcriptional activation assays for estrogen and androgen receptors; an *in vitro* assay for steroidogenesis; a rodent uterotrophic assay; a Hershberger assay, with and without testosterone implants (to assess the androgenic and anti-androgenic potential of the test agent); a rodent female peripubertal assay; a frog metamorphosis assay; and an apical fish reproductive toxicity assessment.

The following matrix indicates the mode(s) of action covered by each assay:

Different assay systems have been proposed as alternatives to some or all of the assays above.

The complexity and redundancy of the screening battery makes it necessary to evaluate the overall results with a weight-of-evidence approach. EDSTAC recommended a number of heuristic rules for this approach, such as:

- *In vivo* results outweigh *in vitro* results.
- *In vitro* results from assay systems with a metabolic capacity outweigh results from those without a metabolic capacity.
- Apical assays outweigh specific ones.
- Consistent results outweigh a single positive result.
- Biological plausibility must be demonstrated.

Compounds that are deemed to be positive in the screening battery are considered to be prospective endocrine disruptors and proceed to the next assessment tier, which consists of more traditional reproductive toxicity evaluations. For mammalian toxicology, this would consist of the two-generation assay, which has undergone some updating to make it more sensitive to effects on androgens and estrogens, and may be further altered to increase its sensitivity to thyroid toxicants. Work is underway to determine whether a modified one-generation assay could provide sufficient information, and replace the two-generation assay.

3.4.1. Examples of how the tiered screening system might work

Because the EDSP has not officially begun, it is not possible at this point to provide retrospective case studies. However, enough has been published about the process that it is possible to present examples of the data sets that would be generated and decisions that might be made for a number of chemicals. Below are four examples of chemicals that have already been well studied so that we know (or can reasonably surmise) the outcome of screening level assessments.

MOA	E	Anti E	A	Anti A	5-a-R Inhib	Ster	Anti T	Arom	HPGA
Assay									
E receptor	x	x							
A receptor			x	x					
Steroid synthesis						x			
Uterotrophic	x	x							
Hershberger			x	x					
Fem. Pubertal.	x	x				x		x	x
Frog							x		
Fish	x	x	x	x	?	x		x	x

E = estrogenic; Anti E = anti-estrogenic; A = androgenic; Anti A = anti-androgenic; 5-a-R Inhib = 5-alpha-reductase inhibitor; Ster = steroidogenic; Anti T = anti-thyroid; Arom = aromatase inhibitor; HPGA = hypothalamic–pituitary–gonadal axis impact.

3.4.1.1. Example 1: Diethylstilbestrol (DES). DES is known to be a potent estrogen; its estrogenic activity is regarded to be the basis for its teratogenicity (reproductive malformations and predisposition for vaginal cancer in humans) (Mittendorf, 1995). How would DES fare if it were subjected to the EDSP tiered screening system?

If simple structure–activity criteria are used in the prioritization tier, DES would probably be flagged as a candidate for screening because it has several chemical characteristics that indicate it might interact with the estrogen receptor. These include a phenolic moiety (two, actually) that bears an aliphatic moiety in the para position to the hydroxyl group, hydrophobicity, and a molecular weight and size that would permit it to interact with the ligand binding region of the estrogen receptor.

DES would be characterized in the screening tier as a potent and relatively specific estrogen agonist. It would be shown to have a high affinity for the estrogen receptor in the receptor binding assay and in the transcriptional activation assay, but essentially no affinity for the androgen receptor. It would be positive in the uterotrophic assay but not in the Hershberger assay. It would accelerate puberty in the female peri pubertal assay.

The weight-of-evidence conclusions from the screening battery would be that DES is an estrogen and a potential reproductive and developmental toxicant. It gives consistent results in *in vitro* and *in vivo* assays for estrogenicity. Its chemical structure supports the conclusion. The results of the screening tier would support a recommendation that further, more comprehensive testing be done to characterize the hazard and derive dose–response data for risk assessment.

3.4.1.2. Example 2: Bisphenol A (BPA). BPA is another well-studied compound. BPA has been shown to have very weak affinity for the estrogen receptor. Although it is clear that BPA has estrogenic potential, this potential does not appear to play a role in its toxicity. Authoritative multigeneration studies have found no estrogen-like effects (Tyl et al., 2002). This is attributable to the fact that BPA is a very weak estrogen, and that toxicological properties of the chemical limit the highest dose that can be achieved in a state-of-the-art toxicity study such that it is below a level that is functionally estrogenic.

If BPA were a new compound, it might be handled in a screening system as follows: The prioritization phase would identify it as a candidate for screening based on its having a para-substituted phenolic moiety, although the nature of the para substituent is not consistent with the aliphatic character of the substituents for other estrogens.

The screening tier would show results that are qualitatively similar to those for DES, but with much lower potency. BPA binds to the estrogen receptor with four or five orders of magnitude less affinity than estradiol. It is uterotrophic; it slightly accelerates the onset of puberty at high dosage in the peripubertal assay. It would have no effects in the androgenicity assays. The conclusion to be drawn from the screening data is that the results are consistent with estrogenic activity and would indicate the need to characterize hazard in a more convincing toxicity assay. As noted above, the results of that higher level testing would be negative despite the fact that the screening results were consistent and reliable. This should not be regarded as a failure of the screen, but a recognition that it is weighted towards sensitivity.

3.4.1.3. Example 3: Acetic acid. This is an example of a compound that should not make it out of the prioritization tier. It has no characteristics that are reminiscent of an estrogen or androgen. It is too small and too polar to interact with a steroid receptor binding domain. The most appropriate decision for acetic acid would be to conclude that no bench-level screening is needed.

3.4.1.4. Example 4: Dexamethasone. Dexamethasone is a synthetic corticosteroid. Its action is on the glucocorticoid receptor. If it were a new compound, it likely would be flagged for screening because of its steroidal structure. Evaluation *in vitro* would indicate that dexamethasone has limited or no affinity for the estrogen or androgen receptor in receptor binding assays. However, the transcriptional activation assays that have been developed to date to detect androgens also respond to glucocorticoids, so dexamethasone may be positive in this *in vitro* assay for androgens. It would be negative in the *in vivo* assays for estrogenicity and androgenicity. The most appropriate weight-of-evidence conclusion from the screening battery is that dexamethasone is unlikely to be an androgen (or estrogen) and would be a low priority for further testing on that basis. The rationale for this conclusion is that the single positive *in vitro* result is inconsistent with data from other *in vitro* assays or *in vivo* assays. The knowledge that glucocorticoids produce false positives in the transcriptional activation assay supports this conclusion.

3.5. Assessment of the potential allergenicity of foods derived by biotechnology

Modern agricultural biotechnology involves the transfer of DNA that encodes a desired trait to food and feed crops. The safety of foods and feeds derived from such crops is established using the concept of "substantial equivalence." Application of the principle of substantial equivalence involves identifying the similarities and any differences between a product and its closest traditional counterpart that has a history of safe use, and subjecting the differences to a rigorous safety assessment. The similarities noted between the new and traditional crops are not subject to further assessment since this provides evidence that those aspects of the newly developed crop are as safe as crops with a history of safe consumption. The identified differences are subjected to further analysis, to clarify whether any adverse effects on health are likely to result from the intended compositional change. The analysis assesses: (1) the agronomic/

morphological characteristics of the plant; (2) macro- and micro-nutrient composition and content of important anti-nutrients and toxicants; (3) molecular characteristics and expression and safety of any proteins new to the crop; and (4) the toxicological and nutritional characteristics of the novel product compared with its conventional counterpart in appropriate animal models (ILSI, 2004). By following this process, the safety assessment strategies for genetically-modified crops have been shown to provide a level of safety assurance that is at least comparable with that available for conventional crops (ILSI, 2004).

Problem formulation: The process of transferring genes that confer a desired trait from one species to another is one of the advantages that biotechnology has provided to food crops. Prior to the development and marketing of this type of technology in the 1990's, the transfer of genes from different species was limited to very closely related species. Additionally, the ability to select a single gene to transfer was not possible. However, this new technology has raised concerns about the transfer of genes that encode for proteins that may transfer or create an allergen in a food that was previously well tolerated by the majority of the human population. Thus, the issue of a possible introduction of a food allergen in to the food supply must be addressed prior to making the product available on the open market for human consumption.

Review of pre-existing data: Food allergy is defined as an immune response characterized by the production of immunoglobulin E (IgE), to a protein or proteins in food (Hefle et al., 1996). The incidence of food allergic individuals has been estimated to be between 3 and 4% of the general population and 6–8% of children in the United States. While reports exist for allergic reactions to over 160 different foods, greater than 90% of reported reactions are due to eight foods. The "Big 8" is comprised of cow's milk, crustacea, eggs, fish, peanuts, soy, tree nuts, and wheat. Unlike for many other agents, dose–responses have not been established for food allergens. Efforts are underway, including human clinical challenge studies utilizing a consensus protocol, to identify potential threshold doses of food allergens in humans.

Development and implementation of strategic plans: Without a defined threshold for the elicitation of an allergic response by a known or suspected food allergen, it is difficult to assess the safety of foods derived from biotechnology. Thus, the postulated characteristics associated with known food allergen proteins, which are believed to induce the human immune system of susceptible individuals to mount an IgE response, have been relied upon heavily to evaluate foods derived through biotechnology for potential allergenicity. These characteristics include heat stability, protease stability, amino acid sequence homology of known food allergens (usually major proteins in the food) and molecular weights between approximately 10 and 70 kD. However, all of these characteristics do not necessarily apply to all food allergens. Animal models are highly relied upon by toxicologists to predict responses in humans. Numerous animal models, including mice, rats, swine, and canines have been suggested as appropriate model(s) for predicting the allergenicity of proteins in foods derived from biotechnology. However, scientific agreement on the best animal model has yet to be reached. Despite this, animal models are relied upon for the evaluation of proteins to be inserted into foods. Human clinical studies similar to those used in the assessment of pharmaceutical agents are not used to evaluate novel proteins to be expressed in foods derived through biotechnology. Instead, skin prick tests and the radioallergosorbent tests (RAST) are utilized to evaluate the protein to be transferred or the food into which the protein has been transferred. The RAST is limited by the lack of data regarding the human allergenic response to the majority of proteins to be evaluated, and will reveal only if the IgE in the sera is cross-reactive to the protein of interest. Because of the number of shortcomings in each of the aforementioned methods to evaluate foods derived through biotechnology, a decision tree approach that included all of the above methods has been devised.

Outcome: In the late 1980's, the International Food Biotechnology Council (IFBC) was formed to develop an approach to evaluating the safety of foods produced through biotechnology. A decision tree was developed that included assessment of the allergenicity potential of the inserted proteins. The decision tree was accepted by numerous worldwide food safety organizations including the Food and Agriculture Organization of the United Nations/World Health Organization (FAO/WHO), the US FDA, and the Organization for Economic Co-operation and Development (OECD). Because of continuing concerns about the potential allergenicity of foods derived through biotechnology, the decision tree was re-evaluated in 1996 by an International Life Science Institute/International Food Biotechnology Council (ILSI/IFBC) expert committee. The committee produced a detailed decision tree to evaluate the potential allergenicity of foods derived from biotechnology (Metcalfe et al., 1996; Fig. 3). The 1996 ILSI/IFBC decision tree also was recognized worldwide by numerous food safety organizations and then further re-evaluated by an FAO/WHO expert committee in 2001 (FAO/WHO, 2001; Fig. 4). The 2001 decision tree is currently being used to evaluate the potential allergenicity of foods with genes introduced through biotechnology (FAO/WHO, 2001).

The 2001 FAO/WHO decision tree differs from the 1996 ILSI/IFBC decision tree in a number of ways. The 1996 ILSI/IFBC decision tree called for the number of consecutive amino acid sequence homology to be eight amino acids; the 2001 FAO/WHO decision tree reduced the sequence homology number to six amino acids. The 2001 FAO/WHO decision tree indicates that sera from 25 allergic individuals be evaluated versus the 1996 ILSI/IFBC decision tree value of 14. Additionally, the 2001 FAO/WHO decision tree includes a box for "targeted" allergens; these are allergens from sources in addition to food allergens such as animals, pollens, and molds: this element

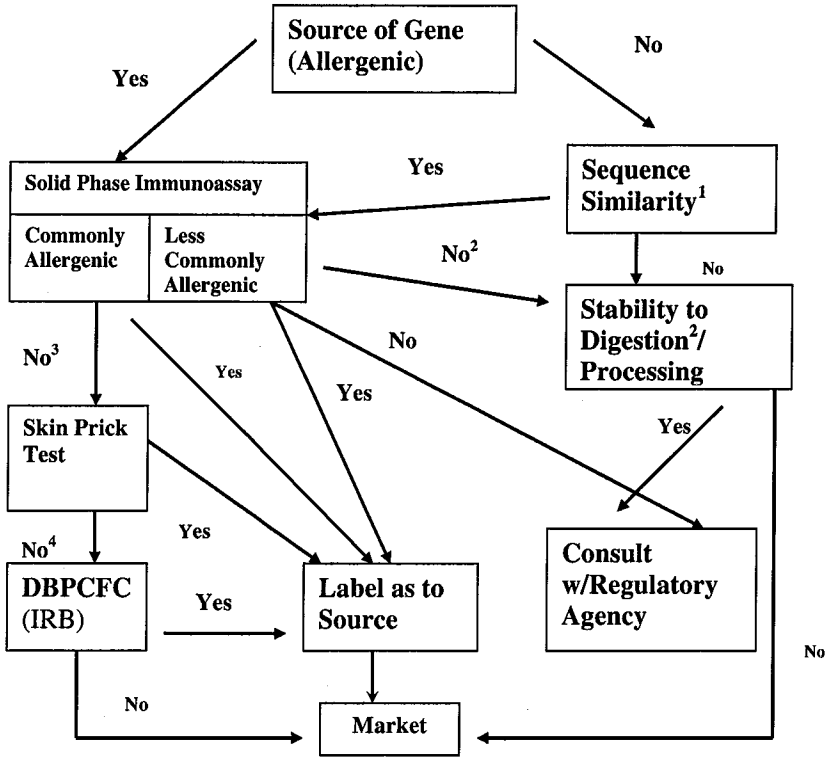


Fig. 3. The ILSI/IFBC decision tree (Metcalfe et al., 1996).

was not included in the 1996 ILSI/IFBC decision tree. The 2001 FAO/WHO decision tree also includes a box for animal models and a standardized protocol for a pepsin resistance assay to evaluate the protease stability of the protein of interest. In a similar fashion to animal models of food allergy, the pepsin resistance characteristics of food allergens have not been agreed upon yet scientifically. While uncertainty concerning the potential allergenicity of foods derived from biotechnology still exists, the acceptance of the decision tree approach by a number of food safety organizations worldwide speaks to the value of a number of assays which are generally regarded as screening assays for the protection of human health.

3.6. Use of limited or surrogate toxicity data for assessing potential risks to human health

Problem formulation: The process of determining the potential of a compound to induce adverse effects on human health, in many cases, can be a prescribed exercise that is conducted with few constraints on the timing of completion. However, issues can arise that require that an assessment be made in a very limited amount of time. These situations can involve compounds that have not been evaluated to the extent required for a robust risk assessment. In some cases, the available toxicity data are limited to a small number of studies, possibly only a single study. Additionally, issues can arise in which data for the specific compound are not available and one must rely on data generated on surrogate compounds that are chemically and structurally close relatives to reach a decision concerning the potential to produce adverse effects on human health. Other situations may arise in which data generated in experiments evaluating unrelated routes of exposure and/or toxicity endpoints can be used to evaluate the potential for adverse human health effects. An example of such could be the use of data derived from an inhalation study to evaluate potential risks associated with ingestion of the compound of interest. While the aforementioned situations are far from ideal for performing risk assessments, time can

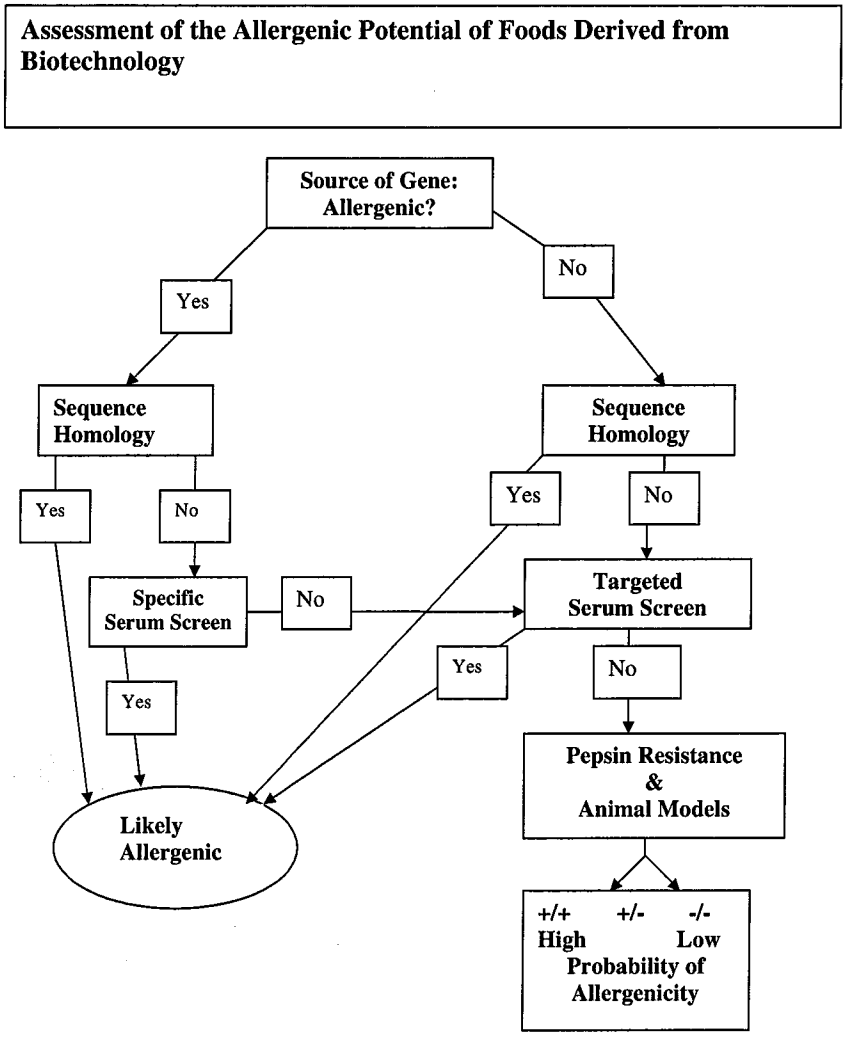


Fig. 4. The FAO/WHO decision tree (FAO/WHO, 2001).

be a constraining factor that cannot be altered to allow for the generation of new data or to search for obscure reference materials. Well-designed screening assays can prove to be invaluable tools in situations as described above. In cases in which ample time exists to perform short-term assays, a screening assay with sufficient specificity can provide much needed data to aid in the risk assessment process.

In many cases, the toxicologist can use limited resources and provide a good approximation of the potential risks to human health. The lack of relevant data can be overcome by relying on a number of assumptions, an understanding of physiology, and knowledge of structure activity relationships. However, as the amount of appropriate toxicity data decreases, the amount of uncertainty in a risk assessment rises. Uncertainty can rise to such a level that the meaningfulness of the risk assessment is compromised. Once this occurs, scientific judgment must be applied to the particular situation. Other factors such as dose and exposure time/occurrences play important roles in the risk assessment process and can affect the amount of uncertainty that one brings into a risk assessment. In many cases, as dose and exposure time/occurrences decrease the uncertainty assigned to a risk assessment decreases as well. Dose and exposure time/occurrences can be utilized as the major factor when situations of high uncertainty indicate the use of scientific judgment to assess for potential risks. A single short-term exposure scenario warrants less concern than an exposure that may occur for a substantial portion of an individual's lifetime. Another important factor to consider is the chemical properties of the compound of interest and the matrix in which the compound of interest is encountered. High or low chemical reactivity may in certain situations preclude the compound of interest from entering humans in a toxic form. The route of exposure is also a major factor to consider in the determination of potential toxicity. The determination of potential risks to human health can be evaluated and successfully predicted using available science, basic assumptions, and toxicological expertise.

The principles and practices described above have successfully been used in a number of instances such as specific cases of unexpected occurrence of contaminants in a food product or, generically, in developing the concept of Threshold of Toxicological Concern (TTC) which is employed by the US Food and Drug Administration and certain other national food regulatory agencies as well as the Joint FAO/WHO Expert Committee on Food Additives (JECFA). These groups have used this concept to distinguish indirect food additives (contact materials) (e.g., USFDA, 1995) and flavoring agents (JECFA, 2005) that pose a significant level of concern from those that pose negligible concern.

The concept of TTC is an approach that, in certain defined situations, can be a useful screening tool. TTC refers to a process of identifying an exposure level below which there is no significant risk to health. This concept is based upon the widely-held tenet of toxicology that, for every agent, there is a threshold of exposure for eliciting effects, such that an exposure at a dose lower than the threshold will have virtually no likelihood of an adverse health effect. As noted above, the USFDA uses the TTC approach in their regulatory framework for indirect food additives. Indirect food additives are substances used in food-contact articles including adhesives and components of coatings, paper and paperboard components, polymers, and adjuvants and production aids. These are substances that may come into contact with food as part of packaging or processing equipment but are not directly added to foods deliberately. The USFDA employs the TTC concept to rapidly and effectively set priorities in order to differentiate agents and circumstances that have the potential to pose a relatively high concern from those that pose negligible concern to human health. The USFDA created the Threshold of Regulation (ToR) for indirect food additives in 1995, a principle that states if a substance used in a food contact article (e.g., food packaging or food processing equipment) passes a cursory exposure and hazard screening, no food additive tolerance (i.e., maximum residue limit (MRL)) need be promulgated. The ToR allows for a relatively quick evaluation by the USFDA of compounds that migrate from packaging materials into foods in very minute quantities ($\leqslant$0.5 ppb) and that may not have been evaluated using oral exposure toxicology studies. This process ensures the preservation of valuable USFDA resources that could be better used to evaluate other issues that may have an impact on human health. In developing the ToR, the USFDA reviewed data that had been previously generated from thousands of feeding studies on hundreds of compounds, including both carcinogens and noncarcinogens.

The results of the US FDA's evaluations revealed that the vast majority of noncarcinogenic compounds, with the exception of five pesticides, did not produce toxicity on an acute or chronic basis at doses below 1 mg/kg bw (ppm). The analysis of data from carcinogens identified in oral feeding studies determined that less than a 1 in 1,000,000 risk for a lifetime of cancer causation would be associated with exposure levels at or below 0.5 mcg/kg (ppb). In addition to evaluating the toxicity potential of numerous compounds, the USFDA provided guidance on the method to estimate exposure from migration to foods. The USFDA assumed a total dietary intake of 3000 g (divided among liquid and solid foods), an upper bound lifetime excess cancer risk of 1 in 1,000,000 and back-calculated the daily intake (dose) of a compound.

Using a highly health-protective approach, USFDA was able to determine that indirect food additives which result in dietary doses of less than or equal to 1.5 μg per person per day are below the ToR (equivalent to 0.5 mcg of an agent per kg of diet). However, known carcinogens or agents suspected of being carcinogenic, based upon structural correlates to known carcinogens are exempted from qualifying for the ToR. The ToR employed by the USFDA is another good example of the utility of "minimal" data sets, such as those generated from carcinogenesis screening assays, to provide a tool to evaluate other compounds of interest and still maintain protection of human health.

As noted above, the FAO/WHO Joint Expert Committee on Food Additives has adopted the TTC principles in their assessment of the safety of flavoring agents. For example, at their meeting in June, 2005, the Committee evaluated the safety of over 130 such agents, many of which lacked the toxicity data set ordinarily considered necessary for developing an acceptable daily intake (ADI). As with food contact substances, the flavoring agents were assessed based first upon the structural class they are in, then determining whether or not the conditions of use would result in an intake greater than the TTC for that class. If the toxicity data on the substance or structurally-related substances were insufficient to derive an ADI, then the judgment regarding presence or lack of a safety concern was based upon whether or not the conditions of use would result in intakes of greater than 1.5 ug/day.

In addition, although the TTC concept was first developed to address the safety assessment of non-pesticide food additives, with an emphasis on their carcinogenic potential, more recent effort has been mounted to examine its value in the assessment of other chemicals used or found in other settings and for additional endpoints of toxicity (e.g., USEPA, 1999a; Kroes et al., 2000, 2004, 2005; Kroes and Kozianowski, 2002; Renwick, 2004, 2005; Blackburn et al., 2005).

The example provided demonstrates how failure to make full use of screening assessment tools such as threshold of concern (TTC) and/or regulatory tools such as Threshold of Regulation (ToR) can lead to an unsatisfying characterization and management of the potential risk. The example also highlights the fact that different authorities may reach different conclusions when the same toxicity data set is available to their assessors. The judgment made about the adequacy of the available data to contribute to a risk assessment and subsequent management decision depends, in part, upon the exposure scenario for which

the assessment is being conducted. It should be noted that, in the example, the exposure scenario is different which could account for some of those differences.

3.6.1. Example: Diisopropylphenylnaphthalene (DIPN) as a food contaminant

Problem formulation: In the late 1990's, diisopropylphenylnaphthalene (DIPN) was discovered to be present as an inadvertent contaminant in recycled paper and board food packaging in the United Kingdom. A judgment was required by the then-named Ministry of Agriculture, Fisheries and Food (MAFF) as to whether or not the levels of contamination constituted a human health risk should the food products be consumed. There was only limited time available for an assessment to be developed and communicated to the public. There existed only limited exposure and chemical-specific or surrogate toxicity data to evaluate the potential risks that DIPN may pose to human health.

DIPN is a solvent that is used in the production of specialty papers, such as carbonless and thermal copy paper, ink-jet printer inks, and as a replacement for polychlorinated biphenyls (PCB) (Sturaro et al., 1994.) There have been instances in which the use of recycled paperboard in primary or secondary food packaging has resulted in the transfer of DIPN from the packaging material to the food (Sturaro et al., 1994; UKMAFF, 1999; Summerfield and Cooper, 2001). Under these circumstances, the chemical would be evaluated by food regulatory authorities in accordance with principles and practices associated with indirect food additives (Begley, 1997). However, 2,6-DIPN also is a biochemical pesticide registered in the US to inhibit the sprouting of stored potatoes (USEPA, 1999b.) As a pesticide, it is evaluated by the pesticide regulatory authorities in accordance with principles and practices associated with these agents. Potentially, then, different toxicity and exposure data may be available to the regulatory authorities. However, it should be noted that, depending upon the stage of development of a regulatory position on the agent, these data may not be publicly available, or even available cross-agency. As noted below, this was the case at the time the UK risk assessment was being developed.

Development and implementation of strategic plans: Key to the development of a risk assessment in the short term is the evaluation of all data available to the assessors, their paucity notwithstanding. To the extent possible, both toxicity and exposure data must be reviewed. The UK regulatory system incorporates the convening of advisory committees to assist in the assessment of scientific information. These advisory committees communicate their conclusions to the sponsoring agency. In this case, the Committee on Toxicity of Chemicals in Food, Consumer Products and the Environment (COT) contributed the assessment for DIPN.

The COT proceeded by (1) reviewing and commenting on a survey of DIPN residues measured in recycled paper and board food packaging, (2) reviewing the limited number of reports examining the potential toxicity of DIPN that existed in the peer-reviewed scientific literature, and later (3) reviewing the data that had been developed in response to regulatory requirements as a pesticide in the US.

The published papers described the use of *in vitro* yeast assays which were negative for estrogenic activity; i.p. administration to mice, focusing on pulmonary damage; biliary and urinary metabolites of 2,6-DIPN in rats; *in vitro* carp hepatocyte metabolism of 2,6-DIPN; data on urinary metabolites in rabbits; and absorption, distribution, and excretion of 2,6-DIPN in rats. (Vinggaard et al., 2000; Honda et al., 1990; Kojima et al., 1978, 1982, 1985a,b). While the utilization and extrapolation of rodent and other animal data to human health risk assessment is common, the paucity of data on the potential toxicities associated with DIPN exposure, in particular, the lack of toxicity data from oral exposure in animals, suggests that the task of performing a credible risk assessment for humans following oral exposure would be very difficult.

A report on the toxicity of DIPN by Hoke and Zellerhoff (1998) used a number of the principles described earlier in this section to conclude that DIPN and other alkylated naphthalenes would not share the toxicities associated with naphthalene. The authors used the information that had been generated on known toxicities and metabolism associated with naphthalene in mammalian species and compared it with the mammalian metabolism data known to exist for DIPN. They observed that, unlike naphthalene, the metabolic pathway for DIPN in rats proceeded nearly exclusively through oxidation of the isopropyl side chain, rather than ring oxidation as occurs with naphthalene. The lack of ring oxidation correlates well with the lack of lung toxicity while the extent of enzymatic oxidative attack at the aromatic ring structure results in a pattern of toxicity characteristic of naphthalene and its monomethyl derivatives. While Hoke and Zellerhoff (1998) focused on the potential for pulmonary toxicity to occur, the understanding of the metabolic consequences, absorption, and distribution of 2,6-DIPN also would allow for an extrapolation to the potential for toxicity to occur at other organs or tissues in the body. The lack of human metabolism data would suggest that the toxicologist would have to rely on the metabolism data available from other mammalian species (rats and rabbits) when assessing potential risks of DIPN to humans, a not uncommon situation.

On the exposure side, it was shown that very low concentrations of DIPN were found to transfer to food (UKMAFF, 1999; Sturaro et al., 1994) and would result in very low dietary intakes (<10 μg/kg bw/day) Another factor is that the number of exposures an individual would experience is most likely low. In addition, the human sense of smell and taste can detect a number of compounds at lower concentrations than can current analytical tools. This fact can also act to limit ingestion of some tainted foods. An unpleasant or out of the ordinary taste of food which has been consumed in the past usually is a deterrent to continued ingestion of the food in question.

Outcome: There is no universal agreement on whether the existing data on DIPN were sufficient to perform a risk assessment. Initially, the COT stated that the toxicological data on DIPN were inadequate and recommended that further studies be conducted, including mutagenicity tests and long-term studies (UKMAFF, 1999) In February, 2000, COT issued its review of newly available mutagenicity data, determining that DIPN was nonmutagenic. The Committee reiterated the need for further safety studies, but stated that 28- and 90-day studies would suffice, since DIPN was unlikely to be a genotoxic carcinogen and dietary intake would be expected to be very low. In October, 2002, the COT issued its evaluation of a newly-submitted 90-day study of 2,6-DIPN and a report prepared by a consultant toxicologist of available data on DIPN. The COT concluded that the 90-day and chronic data (a 1977 long-term carcinogenicity study) cited in the report were not adequate to allow for the derivation of a tolerable daily intake (TDI). Additionally, the 90-day study could not be used to set a no-observed-adverse-effect-level (NOAEL) because the test agent used, 2,6-DIPN, was not the isomeric DIPN mixture used by the ink industry, and, therefore, not the actual substance that may leach into food from packaging. The COT continues to be of the opinion that the toxicity data remain inadequate for use in the derivation of a tolerable daily intake (TDI) and retain their original position that levels of DIPN in recycled paper and board food packaging be kept as low as reasonably practicable to minimize migration into food (COT, 2003). They have not made a statement as to whether or not they believe exposure to DIPN at levels predicted to be in foods constitutes a safety concern.

In contrast, in 1999, US EPA granted a one-year temporary exemption from the requirement to establish a tolerance (MRL) for residues on stored potatoes (USEPA, 1999b). The Agency provided a very good estimation of exposure including dose, frequency, and duration of exposure via stored potatoes to DIPN and concluded it would not constitute a risk to human health. The Agency concluded that 2,6-DIPN did not pose a significant risk to humans or the environment, under the conditions of the proposed use, based on data submitted for the establishment of a temporary tolerance for 2,6-DIPN for the inhibition of sprouting in stored potatoes. In 2003, a temporary tolerance was granted for the use of 2,6-DIPN at 0.5 ppm in or on potatoes and 3.0 ppm in or on potato peels which was to expire on May 31, 2006. The US EPA did have more data than the UK authorities to use in its determination, including a rat sub-chronic 90-day study, which it deemed adequate to derive a no-observed-effect-level (NOEL) and a lowest-observed-adverse-effect-level (LOAEL) and a rat developmental study (USEPA, 2003b). In late 2005, the registrant petitioned the Agency to raise the residue limits on potatoes to 2 ppm and on potato peels to 6 ppm. A decision by the Agency is pending (USEPA, 2005b). [It should be noted that, while US EPA is required to conduct aggregate exposure assessments when deciding whether or not to grant a tolerance, the Agency did not include potential exposure to DIPN leached from food packaging in their aggregate assessment].

Using a weight of the evidence approach, the risks posed to humans following ingestion of low levels of DIPN in foods as a result of transfer from packaging materials can be hypothesized with a relatively low level of uncertainty. Based upon consideration of the information on the comparative toxicology and toxicokinetics of the alkylated naphthalenes, their organoleptic properties and the estimates of exposures that could occur as a result of leaching from food packaging (and, stored potatoes), one could conclude that potential exposures were unlikely to pose an unacceptable risk to humans. This conclusion is supported further by the US EPA decision to issue a temporary tolerance for DIPN based upon estimates of exposure and some additional, more substantive, data on toxicity (USEPA, 2003b).

As noted above, one caveat is that not all of the data available to US EPA were available to the UK authorities and none of the information is published in the peer-reviewed literature. This situation is an example of when valuable and informative data exist but are not readily available for use in a time-constrained risk assessment. Thus, using available data, which in some cases is limited to data generated from an assay that may be classically defined as a screening tool, can provide valuable insights into the potential risks to human health associated with a compound.

4. The future: scientific advances that may improve screening strategies

The biotechnology revolution has created a scientific environment in which all fields of biology are being transformed by the powerful technologies that have enabled the enormous recent advances in understanding of cellular and genetic function. These unprecedented advances, coupled with the availability of technologies that made them possible, are certain to lead to improved screening tools and screening strategies. It will soon be possible to simultaneously monitor the expression of the entire genome. It has already become possible to rapidly identify genetic sequence variations associated with biological outcomes, and identify proteins and intermediary metabolites associated with pathological processes. Some of the promising developments and their potential application to biological screening are discussed below.

4.1. Specific biomarkers

Efficient screening requires biomarkers that are easily monitored and that reflect a specific pathology of concern. The global approaches made possible by "-omics" technologies will facilitate the identification of such biomarkers (Aardema and MacGregor, 2002; Inoue and Pennie, 2003; MacGregor, 2003). As a more comprehensive

knowledge of cellular and molecular responses to specific forms of damage is elucidated, the appropriate biomarkers that reflect these processes, and that are accessible in tissue and body fluids, can be selected and assembled into suitable formats for rapid and efficient screening. Appropriate markers for particular applications can be selected, and will facilitate both *in vivo* and *in vitro* screening. For example, transcription arrays, although extremely powerful, are not suitable for *in vivo* monitoring in humans and in many situations in laboratory models because they require cellular components (messenger RNAs) that are not accessible *via* non- or minimally-invasive techniques. This does not significantly limit their application in *in vitro* cellular model systems, but precludes their application in many circumstances *in vivo*.

Screening approaches need to be designed around specific biomarkers that best reflect the event for which it is desired to screen. For example, in clinical cardiovascular screening programs, certain types of biomarkers are used to identify patients at risk of developing disease and other types are used to diagnose the occurrence of acute cardiovascular events such as myocardial infarction (MI). Cholesterol and lipoprotein, which are mechanistically linked to the risk of arterial plaque formation, and C-reactive protein, which is an indicator of vascular inflammatory response, are examples of biomarkers that reflect risk categories (Ridker et al., 2002; Pearson et al., 2003). Cardiac troponin T or I and cardiac creatine phosphokinase, which are specific myocardial components released from injured or dying myocytes, are used to determine if cardiac damage has in fact occurred (Wallace et al., 2004). These well-established clinical practices illustrate the need to define the objective of a given screening approach and to select those markers most appropriate to a particular objective. It can be expected that the exciting new technologies now available will facilitate the development of comprehensive sets of biomarkers that predict and report cellular and tissue injury. Some of the classes of biomarkers that are expected to become available are summarized below.

4.2. Biomarkers of cellular integrity

Since their introduction in the 1950's, specific cellular constituents that leak passively from cells with compromised membrane integrity, or are lost from cells that have died or are undergoing death processes, have been an important class of biomarker used in both nonclinical and clinical testing. Many of the principal serum biomarkers used in routine toxicology testing are of this class (MacGregor, 2003). These include serum transaminases (AST and ALT), cardiac troponins (cTnT, cTnI), alkaline phosphatase (ALP), glutathione transferase (GST), and creatine phosphokinase (CPK). These markers have been extremely useful as screening assays for various forms of organ and tissue damage, and for monitoring specific types of toxicity. These biomarkers are essentially "reporters" of damage that appear in peripheral blood (or in culture medium in the case of *in vitro* studies) when cells undergo injury or death. They are not necessarily "predictive" biomarkers that indicate early stages in a process that leads to pathology.

Those in common use today were introduced individually as our scientific knowledge of their high content in particular tissues accumulated. The advent of proteomic technologies and immunological analysis techniques specific to individual protein isoforms now makes possible the systematic identification of molecules with appropriate properties to serve as biomarkers of cellular damage in specific cell types. The identification of specific forms of small to moderate weight proteins unique to important cell populations beyond those currently monitored would be extremely valuable, as they would allow comprehensive screening for injury to the cell types in which they occur. Once identified, they could be formatted as a simple screen based on determinations of their levels in plasma, serum, or cell culture medium. New technologies such as antibody arrays would provide very effective formats for such screening tools (Cahill, 2001; Huang et al., 2001).

4.3. Biomarkers of damage response

Defensive and protective molecular systems have co-evolved with the functional molecular pathways necessary for carrying on the chemistry of life (MacGregor et al., 1995). Many of these defensive molecular systems are inducible, involving increases in the activity and/or concentration of protective molecules and pathways. Examples include inducible DNA repair enzymes, molecular "chaperones" and "proteasomes" that associate with and either refold or destroy improperly folded proteins and enzymes, and antioxidants that protect against cellular oxidants (Lindahl and Wood, 1999; Wickner et al., 1999).

These damage-specific molecular responses can be used to screen for the occurrence of damage of particular types. Conceptually, this approach to screening is not new. One established example is the measurement of unscheduled DNA synthesis (repair synthesis), which occurs in response to DNA damage, as a method of screening for damage to DNA (Mirsalis and Butterworth, 1980). Our expanded knowledge of the role of the specific defense of molecules, coupled with advances in technology that permit rapid and inexpensive multiplex measurements, promise new screening formats that should allow simultaneous screening for many different types of specific damage.

4.4. Biomarkers of pathway perturbations

Powerful genomic, proteomic, and other "-omic" technologies now available have greatly expanded our ability to monitor functional pathways and their perturbation. These technologies provide the ability to simultaneously monitor large numbers, or even the complete complement of particular classes of molecules. For example, gene expression arrays currently are capable of simultaneously

screening gene transcription products of 15,000 genes simultaneously–approximately half of the entire complement of cellular genes. Specific technologies currently available include oligonuceleotide or cDNA arrays for monitoring gene transcription, metabonomic methodologies involving the use of high resolution NMR to monitor metabolic products of intermediary metabolism, antibody and protein binding arrays that allow multiplex analysis of specific proteins, and mass spectroscopic techniques combined with multi-dimensional separation technologies for protein analysis (MacGregor, 2003). These technologies all hold enormous potential for multiplex analysis of very large numbers of different classes of cellular molecules. The tools of proteomics and metabonomics allow efficient monitoring of individual proteins and metabolic products of intermediary metabolism (Anderson et al., 2000; Bichsel et al., 2001; Nicholson et al., 1999; Yates, 2000), and these technologies, together with transcription arrays, will likely allow the toxicologist to increase the efficiency of identifying pathway perturbations by several orders of magnitude. With these advances, it is now possible to monitor many thousands of biomarkers simultaneously rather than testing hypotheses with assays that focus on one or a few cellular components.

These technologies will undoubtedly be exploited to identify suitable markers for screening approaches, and then to build these markers into specific formats that are amenable to screening programs. Already, such technologies have revolutionized the pharmaceutical industry by providing ultra-high throughput screening of binding to important cellular receptors and other ligands. Most would agree that a similar revolution of toxicology testing and screening for adverse events is already on the near horizon. This should be expected to greatly facilitate the first step in the risk assessment process, hazard identification, as it will become possible to use laboratory model systems to rapidly identify perturbations in molecular pathways that lead to, or result from, cellular damage. As many specific cellular defense responses to specific forms of injury are now known, this powerful technology should also allow "fingerprinting" classes of cellular damage *via* the monitoring of the induced responses associated with specific types of damage.

4.5. The future

Undoubtedly, the future will bring even more exciting unforeseen opportunities and challenges. However, those discussed above already promise a radical transformation and exponential increase in the efficiency of screening approaches. As these exciting opportunities are realized, it will be necessary to keep in mind the fundamental principles discussed above. Although the available technologies have greatly simplified the analytical aspects of multiplex screening, the multiplex nature of the anticipated screens of the future add greatly to the complexity of establishing the necessary linkage between the markers selected in a screening program and the desired outcome. The analytical capabilities and biological knowledge available to us are truly exciting, but their successful application will still require careful attention to the basic elements of any screening program. These elements include: a clear definition of objectives of the screening program, definition of the mechanistic association between individual biomarkers selected and the specific events for which it is desired to screen, and identification of potential confounding factors that may lead to false negative or false positive results.

Acknowledgement

This report is based on the deliberations of the ILSI Risk Science Institute Expert Working Group on Screening Tools for Toxic Chemicals. The entire working group is acknowledged as active contributors to, and therefore co-authors of, this document. The project was overseen by a steering committee whose members served either on the Working Group or as peer reviewers for the document: Dr. Richard A. Becker, American Chemistry Council; Dr. Mildred S. Christian, Argus International, Inc.; Dr. George P. Daston, Procter & Gamble Company; Dr. John Doull, University of Kansas Medical Center; Dr. David Eastmond, University of California; Dr. Penelope Fenner-Crisp, ILSI Research Foundation, Risk Science Institute; Dr. Joseph Holson, WIL Research Laboratories; Dr. Michael Luster, CDC/NIOSH; Dr. John Moore, Hollyhouse, Inc., Dr. Jennifer Seed, US Environmental and Protection Agency; Dr. Isabel Walls, ILSI Research Foundation, Risk Science Institute.

This project was funded under a cooperative agreement between the US Environmental Protection Agency, Office of Pollution Prevention and Toxics, Washington, D.C., and the ILSI Research Foundation, Risk Science Institute. Although the document represents the consensus of the working group members, it does not necessarily represent the policies, positions or opinions of their respective agencies and organizations.

Mrs. Rita Johnson, Ms. Audrey Ray and Ms. Stephanie Carter are gratefully acknowledged for providing administrative support for this project.

References

Aardema, M.J., MacGregor, J.T., 2002. Toxicology and genetic toxicology in the new era of toxicogenomics: Impact of -omics technologies. Mutat. Res. 499, 13–25.

Allen, B.C., Crump, K.S., Shipp, A.M., 1988. Correlation between carcinogenic potency of chemicals in animals and humans. Risk Anal. 8 (4), 531–544.

Anderson, N.L., Matheson, A.D., Steiner, S., 2000. Proteomics: Applications in basic and applied biology. Curr. Opin. Biotechnol. 11, 408–412.

Anisimov, V.N., 1998. Ageing and the mechanisms of carcinogenesis: some practical implications. J. Exp. Clin. Cancer Res. 17, 263–268.

Ashby, J., Paton, D., 1993. The influence of chemical structure on the extent and sites of carcinogenesis for 522 rodent carcinogens and 55 different human carcinogen exposures. Mutat. Res. 286 (1), 3–74.

Ashby, J., Tennant, R.W., 1991. Definitive relationships among chemical structure, carcinogenicity and mutagenicity for 301 chemicals tested by the US NTP. Mutat. Res. 257, 229–306.

ATSDR, 1999. Agency for Toxic Substances and Disease Registry. Toxicological Profile for Chlorophenols. PB/99/166639, Atlanta, GA.

Barber, E.D., Donish, W.H., Mueller, K.R., 1981. A procedure for the quantitative measurement of volatile liquids in the Ames Salmonella/microsome assay. Mutat. Res 90, 31–48.

Bartsch, H., Malaveille, C., 1990. Screening assays for carcinogenic agents and mixtures: an appraisal based on data in the IARC Monograph series. IARC Sci. Publ. 104, 65–74.

Basketter, D.A., Gerberick, G.F., Kimber, I., Loveless, S.E., 1996. The local lymph node assay: viable alternative to currently accepted skin sensitization tests. Food Chem. Toxicol. 34, 985–997.

Begley, T.H., 1997. Methods and approaches used by the FDA to evaluate the safety of food packaging materials. Food Add. Contam. 14 (6–7), 545–553.

Benigni, R., Giuliani, A., 1999. Tumor profiles and carcinogenic potency in rodents and humans: Values for cancer risk assessment. Environ. Carcinogenesis Ecotoxicol. Rev. – Part C of Environ. Sci. Health 17 (1), 45–67.

Benigni, R., Giuliani, A., 2003. Putting the predictive toxicology challenge into perspective: reflections on the results. Bioinformatics 19, 1194–1200.

Benigni, R., Passerini, L., 2002. Carcinogenicity of the aromatic amines: from structure–activity relationships to mechanisms of action and risk assessment. Mutat. Res. 511, 191–206.

Benigni, R., Zito, R., 2004. The second National Toxicology Program comparative exercise on the prediction of rodent carcinogenicity: definitive results. Mutat. Res. 566 (1), 49–63.

Bichsel, V.E., Liotta, L.A., Petricoin III, E.F., 2001. Cancer proteomics: From biomarker discovery to signal pathway profiling. Cancer J. 7, 69–78.

Biomarkers Definitions Working Group, 2001. Biomarkers and surrogate end- points: preferred definitions and conceptual framework. Clin. Pharmacol. Ther. 69, 89–95.

Blackburn, K., Stickney, J.A., Carlson-Lynch, H.L., McGinnis, P.M., Chappell, L., Felter, S.P., 2005. Application of the threshold of toxicological concern approach to ingredients in personal and household care products. Regul. Toxicol. Pharmacol. 43 (3), 249–259.

Boutwell, R.K., Bosch, D.K., 1959. The tumor-promoting action of phenol and related compounds for mouse skin. Cancer Res. 19, 413–424.

Brennan, R.J., Schiestl, R.H., 1998. Chloroform and carbon tetrachloride induce intrachromosomal recombination and oxidative free radicals in Saccharomyces cerevisiae. Mutat Res. 397 (2), 271–278.

Brimer, P.A., Tan, E.L., Hsie, A.W., 1982. Effect of metabolic activation on the cytotoxicity and mutagenicity of 1,2-dibromoethane in the CHO/HGPRT system. Mutat. Res. 95, 377–388.

Bucher, J.R., Portier, C., 2004. Human carcinogenic risk evaluation, Part V: The national toxicology program vision for assessing the human carcinogenic hazard of chemicals. Toxicol. Sci. 82 (2), 363–366.

Butterworth, B.E., 1990. Consideration of both genotoxic and nongenotoxic mechanisms in predicting carcinogenic potential. Mutat. Res. 239, 117–132.

Cahill, D.J., 2001. Protein and antibody arrays and their medical applications. J. Immunol. Methods 250, 81–91.

Carere, A., Morpurgo, G., 1981. Comparison of the mutagenic activity of pesticides *in vitro* in various short term assays. Prog. Mutat. Res. 2, 87–104.

Clayson, D.B., Kitchin, K.T., 1999. Interspecies differences in response to chemical carcinogens. In: Kitchin, Kirk T. (Ed.), Carcinogenicity. Testing, Predicting, and Interpreting Chemical Effects. Marcel Dekker Inc., New York, pp. 837–880.

Colacci, A., Bartoli, S., Bonora, B., Guidotti, L., Lattanzi, G., Mazzullo, M., Niero, A., Perocco, P., Silingardi, P., Grilli, S., 1991. Chloroform bioactivation leading to nucleic acids binding. Tumori 77 (4), 285–290.

COT (United Kingdom Food Standards Agency, Committee on Toxicity), 2003, Annual Report 2002. Available from: <http://www.food.gov.uk/multimedia/pdfs/COT_COM_COC.pdf#page=12>.

Cronin, M.T.D., Jaworska, J.S., Walker, J.D., Comber, M.H.I., Watts, C.D., Worth, A.P., 2003. Use of QSARs in international decision-making frameworks to predict health effects of chemical substances. Environ. Health Persp. 111, 1391–1401.

Davies, T.S., Lynch, B.S., Monro, A.M., Munro, I.C., Nestman, E.R., 2000. Rodent carcinogenicity tests need be no longer than 18 months: an analysis based on 210 chemicals in the IARC monographs. Food Chem. Toxicol. 38, 219–235.

Dearman, R.J., Betts, C.J., Humphreys, N., Flanagan, B.F., Gilmour, N.J., Basketter, D.A., Kimber, I., 2003. Chemical allergy: considerations for the practical application of cytokine profiling. Toxicol. Sci. 71 (2), 137–145.

Duchin, K.L., McKinstry, D.N., Cohen, A.I., Migdalof, B.H., 1988. Pharmacokinetics of captopril in healthy subjects and in patients with cardiovascular diseases. Clin. Pharmacokinet. 14 (4), 241–259.

Eastin, W.C., Mennear, J.H., Tennant, R.W., Stoll, R.E., Branstetter, D.G., Bucher, J.R., McCullough, B., Binder, R.L., Spalding, J.W., Mahler, J.F., 2001. Tg.AC genetically altered mouse: assay working group overview of available data. Toxicol. Pathol. 29 (Suppl.), 60–80.

EDSTAC, 1998. Endocrine Disruptor Screening and Testing Advisory Committee Final Report. Available from: <http://www.epa.gov/scipoly/oscpendo/history/finalrpt.htm>.

Eriksson, L., Jaworska, J., Worth, A.P., Cronin, M.T.D., McDowell, R.M., Gramatica, P., 2003. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. Environ. Health Persp. 111, 1361–1375.

Fan, A.M., Howd, R., 2001. Quantitative Risk Assessment of Non-genotoxic Carcinogens. In: Choy, W.N. (Ed.), Genetic Toxicology and Cancer Risk Assessment. Marcel Dekker, New York, pp. 299–320.

FAO/WHO, 2001. Evaluation of Allergenicity of Genetically Modified Foods. Available from: <http://www.who.int/foodsafety/publications/biotech/en/ec_jan2001.pdf>.

Faustman, E.M., Omenn, G.S., 2001. Risk Assessment. In: Klaassen, Curtis D. (Ed.), Casarett and Doull's Toxicology. The Basic Science of Poisons, 6th ed. McGraw-Hill, San Francisco, pp. 83–104.

Fielden, M.R., Zacharewski, T.R., 2001. Challenges and limitations of gene expression profiling in mechanistic and predictive toxicology. Toxicol. Sci. 60, 6–10.

Franke, R., Gruska, A., 2003. General introduction to QSAR. In: Benigni, R. (Ed.), Quantitative Structure-Activity Relationship (QSAR) Models of Mutagens and Carcinogens. CRC Press, pp. 1–40.

French, J., Storer, R.D., Donehower, L.A., 2001. The nature of the heterozygous Trp53 knockout model for identification of mutagenic carcinogens. Toxicol. Pathol. 29 (Suppl.), 24–29.

Fujie, K., Aoki, T., Wada, M., 1990. Acute and subacute cytogenetic effects of the trihalomethanes on rat bone marrow cells *in vivo*. Mutat. Res. 242 (2), 111–119.

Fujii, K., 1991. Evaluation of the newborn mouse model for chemical tumorigenesis. Carcinogenesis 12 (8), 1409–1415.

Gaylord, D.W., Kodell, R.L., Chen, J.J., Krewski, D., 1999. A unified approach to risk assessment for cancer and noncancer endpoints based on benchmark doses and uncertainty factors. Regul. Toxicol. Pharmacol. 29, 151–157.

Gerberick, G.F., House, R.V., Fletcher, E.R., Ryan, C.A., 1992. Examination of the local lymph node assay for use in contact sensitizer risk assessment. Fundam. Appl. Toxicol. 19, 438–445.

Gerberick, G.F., Robinson, M.K., Ryan, C.A., Dearman, R.J., Kimber, I., Basketter, D.A., Wright, Z., Marks, J.G., 2001. Contact allergenic potency: correlation of human and local lymph node assay data. Am. J. Contact. Dermat. 12 (3), 156–161.

Gerberick, G.F., Cruse, L.W., Ryan, C.A., Hulette, B.C., Chaney, J.G., Skinner, R.A., Dearman, R.J., Kimber, I., 2002. Use of a B cell marker (B220) to discriminate between allergens and irritants in the local lymph node assay. Toxicol. Sci. 68 (2), 420–428.

Gold, L.S., Slone, T.H., 1993. Prediction of carcinogenicity from two versus four sex-species groups in the carcinogenic potency database. J. Toxicol. Environ. Health 39, 143–157.

Greene, N., 2002. Computer systems for the prediction of toxicity: an update. Dv. Drug Delivery Rev. 54, 417–431.

Hardcastle, J.D., Chamberlain, J.O., Robinson, M.H., Moss, S.M., Amar, S.S., Balfour, T.W., James, P.D., Mangham, C.M., 1996. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. Lancet 348, 1472–1477.

Haseman, J.K., 2000. Using the NTP database to assess the value of rodent carcinogenicity studies for determining human cancer risk. Drug Metabolism Rev. 32 (2), 169–186.

Haseman, J.K., Huff, J.E., 1987. Species correlation in long term carcinogenicity studies. Cancer Lett. 37, 125–132.

Haseman, J.K., Lockhart, A.M., 1993. Correlations between chemically related site-specific carcinogenic effects in long-term studies in rats and mice. Environ. Health Persp. 101, 50–54.

Haseman, J., Melnick, R., Tomatis, L., Huff, J., 2001. Carcinogenesis bioassays: Study duration and biological relevance. Food Chem. Toxicol. 39 (7), 739–744.

Hefle, S.L., Nordlee, J.A., Taylor, S.L., 1996. Allergenic Foods. Critical Rev. Food Sci. Nutri. 36 (S), S69–S89.

Helma, C., Kramer, S., 2003. A survey of the Predictive Toxicology Challenge 2000-2001. Bioinformatics 19, 1179–1182.

Hoke, H., Zellerhoff, R., 1998. Metabolism and toxicity of diiospropylnaphthalene as compared to naphthalene and monoalkyl naphthalenes: a minireview. Toxicology 126, 1–7.

Honda, T., Kiyozumi, Kojima, S., 1990. Alkylnaphthalene. XI. Pulmonary toxicity of naphthalene, 2-methylnaphthalene, and isopropylnaphthalenes in mice. Chem. Pharm. Bull. (Tokyo) 38 (11), 3130–3135.

Huang, R.-P., Huang, R., Fan, Y., Lin, Y., 2001. Simultaneous detection of multiple cytokines from conditioned media and patients sera by an antibody-based protein array system. Anal. Biochem. 294, 55–62.

Huff, J.E., 1993. Chemicals and cancer in humans. First evidence in experimental animals. Environ. Health Perspect. 100, 201–210.

Huff, J.E., 1994. Chemicals causally associated with cancers in humans and in laboratory studies. A perfect concordance. In: Waalkes, M.P., Ward, J.M. (Eds.), Carcinogenesis. Ravens Press, New York, pp. 25–37.

Huff, J.E., 1999. Value, Validity and Historical Development of Carcinogenesis Studies for Predicting and Confirming Carcinogenic Risks to Humans 21-123. In: Kitchin, Kirk T. (Ed.), Carcinogenicity. Testing, predicting, and interpreting chemical effects. Marcel Dekker, Inc., New York.

Huff, J., Cirvello, J., Haseman, J., Bucher, J., 1991. Chemicals associated with site-specific neoplasia in 1394 long-term carcinogenesis experiments in laboratory rodents. Environ. Health Perspect. 93, 247–270.

Hulzebos, E.M., Janssen, P.A.H., Malankiewicz, L., Meijerink, M.C.M., Muller, J.J.A., Pelgrom, S.M.G., Verdam, L. Vermeire, T.G. 2001. The Application of Structure-Activity Relationships in Human Hazard Assessment: a First Approach. RIVM Report 601516 008, Rijksinstituut voor volksgezondheid en milieu, BA Bilthoven, The Netherlands.

Humphreys, N.E., Dearman, R.J., Kimber, I., 2003. Assessment of cumulative allergen-activated lymph node cell proliferation using flow cytometry. Toxicol. Sci. 73 (1), 80–89.

IARC, 1999. Quantitative Estimation and Prediction of Human Cancer Risks. In: Moolgavkar, S., Krewski, D., Zeise, L., Cardis, E., Moller, H., (Eds.), IARC Scientific Publication No. 131. International Agency for Research on Cancer. Lyon, France.

ICCVAM, 1999. The murine local lymph node assay: A test method for assessing the allergenic contact dermatitis potential of chemicals/compounds. Interagency Coordinating Committee on the Validation of Alternative Methods. National Toxicology Program. National Institutes of Health. Department of Health and Human Services. NIH Publication No. 99-4494.

ICH, 1997a. International Conference on Harmonization. ICH harmonized Tripartite Guideline. Genotoxicity: A standard battery for genotoxicity testing of pharmaceuticals S2B. Geneva, Switzerland.

ICH, 1997b. In: International Conference on Harmonization. ICH harmonized Tripartite Guideline. Testing for Carcinogenicity of Pharmaceuticals S1B. Geneva, Switzerland.

ILSI, 1997. An evaluation of US EPA's proposed guidelines for carcinogen risk assessment using chloroform and dichloroacetate as case studies: Report of an expert panel. International Life Sciences Institute, Health and Environmental Sciences Institute.

ILSI International Food Biotechnology Committee, 2004. Nutritional and Safety Assessments of Foods and Feeds Nutritionally Improved through Biotechnology. Comprehensive Reviews in Food Science and Food Safety. Available from: <www.ift.org/pdfs/crfsfs/crfsfsv3n2p0035-0104ms20040106.pdf>.

Inoue, T., Pennie, W.D. (Eds.), 2003. Toxicogenomics. Springer-Verlag, Tokyo, pp. 171–193.

International Programme On Chemical Safety (IPCS), 1996. Principles And Methods For Assessing Direct Immunotoxicity Associated With Exposure To Chemicals. World Health Organization Geneva, Switzerland. Available from: <www.Inchem.Org/Documents/Ehc/Ehc/Ehc180.Htm#Partnumber:1>.

Ito, N., Tsuda, H., Tatematsu, M., Inoue, T., Tagawa, Y., Aoki, T., Uwagawa, S., Kagawa, M., Ogiso, T., Masui, T., et al., 1988a. Enhancing effect of various hepatocarcinogens on induction of preneoplastic glutathione S-transferase placental form positive foci in rats–an approach for a new medium-term bioassay system. Carcinogenesis 9 (3), 387–394.

Ito, N., Imaida, K., Tsuda, H., Shibata, M., Aoki, T., de Camargo, J.L., Fukushima, S., 1988b. Wide-spectrum initiation models: possible applications to medium-term multiple organ bioassays for carcinogenesis modifiers. Jpn. J Cancer Res. 79 (4), 413–417.

Ito, N., Imaida, K., de Camargo, J.L., Takahashi, S., Asamoto, M., Tsuda, H., 1988c. A new medium-term bioassay system for detection of environmental carcinogens using diethylnitrosamine-initiated rat liver followed by D-galactosamine treatment and partial hepatectomy. Jpn. J Cancer Res. 79 (5), 573–575.

Ito, N., Shirai, T., Hasegawa, R., 1992. Medium-term bioassays for carcinogens. IARC Sci. Publ. 116, 353–388.

Ito, N., Hasegawa, R., Imaida, K., Takahashi, S., Shirai, T., 1994. Medium-term rat liver bioassay for rapid detection of carcinogens and modifiers of hepatocarcinogenesis. Drug Metab. Rev. 26 (1-2), 431–442.

Ito, N., Hasegawa, R., Imaida, K., Hirose, M., Shirai, T., 1996. Medium-term liver and multi-organ carcinogenesis bioassays for carcinogens and chemopreventive agents. Exp. Toxicol. Pathol. 48 (2–3), 113–119.

Ito, N., Imaida, K., Tamano, S., Hagiwara, A., Shirai, T., 1998a. Medium-term bioassays as alternative carcinogenicity test. J. Toxicol. Sci. 23 (Suppl. 2), 103–106.

Ito, N., Imaida, K., Hirose, M., Shirai, T., 1998b. Medium-term bioassays for carcinogenicity of chemical mixtures. Environ. Health Perspect. 106 (Suppl. 6), 1331–1336.

Ito, N., Imaida, K., Asamoto, M., Shirai, T., 2000. Early detection of carcinogenic substances and modifiers in rats. Mutat. Res. 462 (2–3), 209–217.

Ito, N., Tamano, S., Shirai, T., 2003. A medium-term rat liver bioassay for rapid *in vivo* detection of carcinogenic potential of chemicals. Cancer Sci. 94 (1), 3–8.

Jackson, E.K., 2001. Chapter 31 Renin and Angiotensin. In: Hardman, J.G., Limbird, L.E. (Eds.), Goodman and Gilman's The Pharmacological Basis of Therapeutics. McGraw Hill, New York, pp. 821–822.

JECFA, 2005. Joint FAO/WHO Expert Committee on Food Additives. Sixty-fifth Meeting. Summary and Conclusions. www.who.int/ipcs/food/jecfa/summaries/summary65.pdf (Accessed on June 7–16, 2005).

Kale, P., Baum, J.W., 1979. Sensitivity of Drosophila melanogaster to low concentrations of the gaseous 1,2-dibromoethane: I. Acute exposures. Environ. Mutagen 1, 15–18.

Kale, P.G., Baum, J.W., 1983. Sensitivity of Drosophila melanogaster to low concentrations of gaseous mutagens. 4. Mutations in embryonic spermatogonia. Mutat. Res. 113, 135–144.

Kimber, I., Hilton, J., Dearman, R.J., Gerberick, G.F., Ryan, C.A., Basketter, D.A., Scholes, E.W., Ladics, G.S., Loveless, S.E., House, R.V., Guy, A., 1995. An international evaluation of the murine local lymph node assay and comparison of modified procedures. Toxicology 103, 63–73.

Kimber, I., Hilton, J., Dearman, R.J., Gerberick, G.F., Ryan, C.A., Basketter, D.A., Lea, L., House, R.V., Ladics, G.S., Loveless, S.E., Hastings, K.L., 1998. Assessment of the skin sensitization potential of topical medicants using the local lymph node assay: an international evaluation. J. Toxicol. Environ. Health Part A 53, 563–579.

Kimber, I., Dearman, R.J., Basketter, D.A., Ryan, C.A., Gerberick, G.F., 2002. The local lymph node assay: past, present and future. Contact Dermatitis 47 (6), 315–328.

Kirkland, D., Aardema, M., Henderson, L., Muller, L., 2005. Evaluation of the ability of a battery of three *in vitro* genotoxicity tests to discriminate rodent carcinogens and non-carcinogens I. Sensitivity, specificity and relative predictivity. Mutat. Res. 584 (1-2), 1–256.

Kirsch-Volders, M., Sofuni, T., Aardema, M., Albertini, S., Eastmond, D., Fenech, M., Ishidate Jr., M., Kirchner, S., Lorge, E., Morita, T., Norppa, H., Surralles, J., Vanhauwaert, A., Wakata, A., 2003. Report from the *in vitro* micronucleus assay working group. Mutat. Res. 540 (2), 153–163.

Klaunig, J.E., Babich, M.A., Baetcke, K.P., Cook, J.C., Corton, J.C., David, R.M., DeLuca, J.G., Lai, D.Y., McKee, R.H., Peters, J.M., Roberts, R.A., Fenner-Crisp, P.A., 2003. PPAR alpha agonist-induced rodent tumors: modes of action and human relevance. Crit. Rev. Toxicol. 33 (6), 655–780.

Kojima, S., Nakagawa, M., Suzuki, R., Horio, M., Tanaka, Y., 1978. Alkylnaphthalenes. I. Absorption, tissue distribution and excretion of 2,6-diiospropylnaphthalene in rats. Chem. Pharm. Bull. (Tokyo) 26 (10), 3007–3009.

Kojima, S., Honda, T., Nakagawa, M., Kiyozumi, M., Takadate, A., 1982. Urinary metabolites of 2,6-diiospropylnaphthalene in rats. Drug Metab. Dispos. 10 (4), 429–433.

Kojima, S., Honda, T., Kiyozumi, M., 1985a. Biliary metabolites of 2,6-diiospropylnaphthalene in rats. Bull. Environ. Contam. Toxicol. 35 (6), 745–749.

Kojima, S., Tanaka, M., Yoshida, T., 1985b. Studies on the sequence of 2,6-diiospropylnaphthalene metabolite formation using carp hepatocyte. Ecotoxicol. Environ. Saf. 9 (3), 364–368.

Kroes, R., Kozianowski, G., 2002. Threshold of toxicological concern (TTC) in food safety assessment. Toxicol. Lett. 127 (1–3), 43–46.

Kroes, R., Galli, C., Munro, I., Schilter, B., Tran, L., Walker, R., Wurtzen, G., 2000. Threshold of toxicological concern for chemical substances present in the diet: a practical tool for assessing the need for toxicity testing. Food Chem. Toxicol. 38 (2–3), 255–312.

Kroes, R., Renwick, A., Cheeseman, M., Kleiner, J., Mangelsdorf, I., Piersma, A., Schilter, G., Schlatter, J., vam Schothorst, F., Vos, J.G., Wurtzen, G., 2004. Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet. Food Chem. Toxicol. 42 (1), 65–83, European branch of the International Life Sciences Institute.

Kroes, R., Kleiner, J., Renwick, A., 2005. The threshold of toxicological concern concept in risk assessment. Toxicol. Sci. 86 (2), 226–230.

Lai, D.Y., Baetcke, K.P., Vu, V.T., Cotruvo, J.A., Eustis, S.L., 1994. Evaluation of reduced protocols for carcinogenicity testing of chemicals: Report of a joint EPA/NIEHA workshop. Regul. Toxicol. Pharmacol. 19, 183–201.

Levin, B., Brooks, D., Smith, R.A., Stone, A., 2003. Emerging technologies in screening for colorectal cancer: CT colonography, immunochemical fecal occult blood tests, and stool screening using molecular markers. CA Cancer J. Clin. 53 (1), 44–55.

Lindahl, T., Wood, R.D., 1999. Quality control by DNA repair. Science 286, 1897–1905.

Llewellyn, G.C., Lockwood, J.F., Wierda, D., 1999. Evaluation of transdermally-delivered drugs using the local lymph node assay. The Toxicol. 48, 802.

Loveless, S.E., Ladics, G.S., Gerberick, G.F., Ryan, C.A., Basketter, D.A., Scholes, E.W., House, R.V., Hilton, J., Dearman, R.J., Kimber, I., 1996. Further evaluation of the local lymph node assay in the final phase of an international collaborative trial. Toxicology 108, 141–152.

MacGregor, J.T., 2003. The future of regulatory toxicology: impact of the biotechnology revolution. Toxicol. Sci. 75, 236–248.

MacGregor, J.T., Farr, S., Tucker, J.D., Heddle, J.A., Tice, R.R., Turteltaub, K.W., 1995. New molecular endpoints and methods for routine toxicity testing. Fund. Appl. Toxicol. 26, 156–173.

Macri, J.N., Weiss, R.R., 1982. Prenatal serum alpha-fetoprotein screening for neural tube defects. Obstet Gynecol. 59, 633–639.

Maltoni, C., Soffritti, M., Belpoggi, F., 1999. The scientific and methodological basis of experimental studies for detecting and quantifying carcinogenic risks. Ann. New York Acad. Sci. 895, 10–26.

Mandel, J.S., Bond, J.H., Church, T.R., Snover, D.C., Bradley, G.M., Schuman, L.M., Ederer, F., 1993. Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study. N. Engl. J. Med. 328, 1365–1371.

McClain, R.M., Keller, D., Casciano, D., Fu, P., MacDonald, J., Popp, J., Sagartz, J., 2001. Neonatal mouse model: review of methods and results. Toxicol. Pathol. 29 (Suppl), 128–137.

Meek, M.E., Bucher, J.R., Cohen, S.M., Dellarco, V., Hill, R.N., Lehman-McKeeman, L.D., Longfellow, D.G., Pastoor, T., Seed, J., Patton, D.E., 2003. A framework for human relevance analysis of information on carcinogenic modes of action. Crit. Rev. Toxicol. 33 (6), 591–653.

Metcalfe, D.D., Astwood, J.D., Townsend, R., Sampson, H.A., Taylor, S.L., Fuchs, R.L., 1996. Assessment of the allergenic potential of foods derived from genetically engineered crop plants. Crit. Rev. Food Sci. Nut. 36 (Suppl.), S165–S186.

Milunsky, A., Alpert, E., 1984. Results and benefits of a maternal serum alpha-fetoprotein screening program. JAMA 252, 1438–1442.

Mirsalis, J.C., Butterworth, B.E., 1980. Detection of unscheduled DNA synthesis in hepatocytes isolated from rats treated with genotoxic agents: an *in vivo–in vitro* assay for potential carcinogens and mutagens. Carcinogenesis 1, 621–625.

Mittendorf, R., 1995. Teratogen update: carcinogenesis and teratogenesis associated with exposure todiethylstilbestrol (DES) in utero. Teratology 51, 435–445.

Morimoto, K., Koizumi, A., 1983. Trihalomethanes induce sister chromatid exchanges in human lymphocytes *in vitro* and mouse bone marrow cells *in vivo*. Env. Res. 32, 72–79.

Moss, G.P., Gullick, D.R., Cox, P.A., Alexander, C., Ingram, M.J., Smart, J.D., Pugh, W.J., 2006. Design, synthesis and characterization of captopril prodrugs for enhanced percutaneous absorption. J. Pharm. Pharmacol. 58 (2), 167–177.

NCI, 1978. Bioassay of 1,2-dibromoethane. National Cancer Institute. Technical Report No. 86. US Department of Health, Education, and Welfare. Bethesda, Maryland. NCI-CG-TR-86.

Nicholson, J.K., Lindon, J.C., Holmes, E., 1999. "Metabonomics": Understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica 29, 1181–1189.

NIEHS (National Institute of Environmental Health Sciences), 1997. Validation and regulatory acceptance of toxicological test methods: A report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM). NIH publication No. 97-3981. NIEHS, Research Triangle Park, NC.

NRC, 1994. Science and Judgment in Risk Assessment. National Research Council, National Academy Press, Washington, DC.

NTP, 1982. Carcinogenicity bioassay of 1,2-dibromoethane in F344 and B6C3F1 mice (inhalation study). Technical Report No. 210. National Toxicology Program, US Department of Health and Human Services.

NTP, 2006. National Toxicology Program. Descriptions of NTP Study Types. Available from: <http://ntp-server.niehs.nih.gov/ntpweb/index.cfm?objectid=12F22374-B4D0-2B5B-B021C109307FCD12>.

OEHHA, 1999. Public Health Goals for Chemicals in Drinking Water. Methyl tertiary-butyl ether (MTBE). Office of Environmental Health Hazard Assessment, California Environmental Protection Agency, Oakland/Sacramento.

Pearson, T.A., Mensah, G.A., Alexander, R.W., Anderson, J.L., Cannon, R.O., Criqui, M., Fadl, Y.Y., Fortmann, S.P., Hong, Y., Myers, G.L., Rifai, N., Smith, S.C., Traubert, K., Tracy, R.P., Vinicor, F., 2003. Markers of inflammation and cardiovascular disease. Application to clinical and public health practice. Circulation 107, 499–511.

Pietra, G., Spencer, K., Shubik, P., 1959. Response of newly born mice to a chemical carcinogen. Nature 183 (4676), 1689.

Renwick, A., 2004. Toxicology databases and the concept of thresholds of toxicological concern as used by the JECFA for the safety evaluation of flavouring agents. Toxicol. Lett. 149 (1-3), 223–234.

Renwick, A., 2005. Structure-based thresholds of toxicological concern-guidance for application to substances present at low levels in the diet. Toxicol. Appl. Pharmacol. 207 (Suppl. 2), 585–591.

Richard, A.M., Williams, C.R., 2003. Public sources of mutagenicity and carcinogenicity data: use in structure–activity relationship models. In: Benigni, R. (Ed.), Quantitative Structure–Activity Relationship (QSAR) Models of Mutagens and Carcinogens. CRC Press, pp. 145–173.

Ridker, P.M., Rifai, N., Rose, L., Buring, J.E., Cook, N.R., 2002. Comparison of C-reactive protein and low-density lipoprotien cholesterol levels in the predicition of first cardiovascual events. New Engl. J. Med. 347, 1557–1565.

Robbiano, L., Meretoe, E., Migliazza-Morando, A., Pastore, P., Brambilla, G., 1998. Increased frequency of micronucleated kidney cells in rats exposed to halogenated anaesthetics. Mutat. Res. 413 (1), 1–6.

Robinson, D.E., MacDonald, J.S., 2001. Background and framework for ILSI's collaborative evaluation program on alternative models for carcinogenicity assessment. International Life Sciences Institute. Toxicol. Pathol. 29 (Suppl.), 13–19.

Rosenkranz, H.S., 2003. SAR in the assessment of carcinogenesis: The MultiCASE approach. In: Benigni, R. (Ed.), Quantitative Structure-Activity Relationship (QSAR) Models of Mutagens and Carcinogens. CRC Press, 2003, pp. 175–206.

Rozman, K.K., Doull, J., 2000. Dose and time as variables of toxicity. Toxicology 144 (1-3), 169–178.

San Agustin, J., Lim-Sylianco, C.Y., 1978. Mutagenic and clastogenic effects of chloroform. Bull. Phil. Biochem. Soc. 1, 17–23, cited in US EPA, 1994.

Scott, A.E., Kashon, M.L., Yucesoy, B., Luster, M.I., Trinkle, S.S., 2002. Insights into the quantitative relationship between sensitization and challenge for allergic contact dermatitis reactions. Toxicl. Appl. Pharmacol. 183 (1), 66–70.

Serafimova, R., Walker, J., Mekenyan, O., 2002. Androgen receptor binding affinity of pesticide "active" formulation ingredients: QSAR evaluation by COREPA method. SAR QSAR Environ. Res. 13, 127–134.

Shelby, M.D., 1988. The genetic toxicity of human carcinogens and its implications. Mutat. Res. 204, 3–15.

Shelby, M.D., Zeiger, E., 1990. Activity of human carcinogens in the Salmonella and rodent bone-marrow cytogenetics tests. Mutat. Res. 234, 257–261.

Shi, L.M., Fang, H., Tong, W., Wu, J., Perkins, R., Blair, R.M., Branham, W.S., Dial, S.L., Moland, C.L., Sheehan, D.M., 2001. QSAR models using a large diverse set of estrogens. J. Chem. Inf. Comput. Sci. 41, 186–195.

Sonich-Mullin, C., Fielder, R., Wiltse, J., Baetcke, K., Dempsey, J., Fenner-Crisp, P., Grant, D., Hartley, M., Knaap, A., Kroese, D., Mangelsdorf, I., Meek, E., Rice, J.M., Younes, M., 2001. International Programme on Chemical Safety., 2001. IPCS conceptual framework for evaluating a mode of action for chemical carcinogenesis. Regul. Toxicol. Pharmacol. 34 (2), 146–152.

Storer, R.D., French, J.E., Haseman, J., Hajian, G., LeGrand, E.K., Long, G.G., Mixson, L.A., Ochoa, R., Sagartz, J.E., Soper, K.A., 2001. P53± hemizygous knockout mouse: overview of available data. Toxicol. Pathol. 29 (Suppl.), 30–50.

Sturaro, A., Parvoli, G., Rella, R., Bardati, S., Doretti, L., 1994. Food contamination by diisopropylnaphthalenes from cardboard packages. Int. J. Food Sci. Tech. 29, 593–603.

Summerfield, W., Cooper, I., 2001. Investigation of migration from paper and board into food—development of methods for rapid testing. Food Addit. Contam. 18 (1), 77–88.

Tamano, S., Hagiwara, A., Kawabe, M., Sano, M., Tanaka, H., Imaida, K., Shirai, T., Ito, N., 2001. Summary of test results of medium-term carcinogenesis bioassays targeting the liver or multi-organs. Toxicol. Pathol. 29 (Suppl.), 320–322.

Tamaoki, N., 2001. The rasH2 transgenic mouse: nature of the model and mechanistic studies on tumorigenesis. Toxicol. Pathol. 29 (Suppl.), 81–89.

Tan, E.-L., Hsie, A.W., 1981. Mutagenicity and cytotoxicity of haloethanes as studied in the CHO/HGPRT system. Mutat. Res. 90, 183–191.

Tennant, R.W., Margolin, B.H., Shelby, M.D., Zeiger, E., Haseman, J.K., Spalding, J., Caspary, W., Resnick, M., Stasiewicz, S., Anderson, B., et al., 1987a. Prediction of chemical carcinogenicity in rodents from *in vitro* genetic toxicity assays. Science 236 (4804), 933–941.

Tennant, R.W., Spalding, J.W., Stasiewicz, S., Caspary, W.D., Mason, J.M., Resnick, M.A., 1987b. Comparative evaluation of genetic toxicity patterns of carcinogens and noncarcinogens: strategies for predictive use of short-term assays. Environ. Health Perspect 75, 87–95.

Tennant, R.W., Stasiewicz, S., Eastin, W.C., Mennear, J.H., Spalding, J.W., 2001. The Tg.AC (v-Ha-ras) transgenic mouse: nature of the model. Toxicol. Pathol. 29 (Suppl.), 51–59.

Tyl, R.W., Myers, C.B., Marr, M.C., Thomas, B.F., Keimowitz, A.R., Brine, D.R., Veselica, M.M., Fail, P.A., Chang, T.Y., Seely, J.C., Joiner, R.L., Butala, J.H., Dimond, S.S., Cagen, S.Z., Shiotsuka, R.N., Stropp, G.D., Waechter, J.M., 2002. Three-generation reproductive toxicity study of dietary bisphenol A in CD Sprague–Dawley rats. Toxicol. Sci. 68, 121–146.

UC, 1998. Health and Environmental Assessment of MTBE. Report to the Governor and Legislature of the State of California as Sponsored by SB 521. Five Volumes, 874pp. Available from: <http://tsrtp.ucdavis.edu/mtberpt/>.

UK MAFF, 1999. UK Ministry of Agriculture, Fisheries and Food. Diisopropylnaphthalenes in food packaging made from recycled paper and board, Food Surveillance Information Sheet, Number 169.

USEPA, 1991. US Environmental Protection Agency. Alpha$_{2u}$-globulin: Association with chemically-induced renal toxicity and neoplasia in the rat. EPA/625/3-91/019F, Washington, DC.

USEPA, 1998a. US Environmental Protection Agency. Guidelines for ecological risk assessment. Office of Research and Development. EPA/630/R-95/002F, Washington, DC.

USEPA, 1998b. US Environmental Protection Agency. Hazardous waste management system. Identification and listing of hazardous waste: land disposal restriction, organobromine production waste. 40 CFR, Part 261.

USEPA, 1998c. US Environmental Protection Agency. Assessment of thyroid follicular cell tumors. Risk Assessment Forum. Office of Research and Development. EPA/630/R-97/002. Washington, DC.

USEPA, 1998d. US Environmental Protection Agency. Health risk assessment/characterization of the drinking water disinfection byproduct chloroform. Office of Science and Technology, Office of Water, US Environmental Protection Agency, Washington, DC.

USEPA, 1998e. US Environmental Protection Agency. National Drinking Water Regulations: Disinfectants and Disinfection Byproducts; Final Rule. Federal Register 63 (241), 69389-69476, December 16.

USEPA, 1999a. US Environmental Protection Agency. 2,6-Diiospropylnaphthalene; Temporary exemption from the requirement of a tolerance. Federal Register 64(183):51245-51248, September 22.

USEPA, 1999b. US Environmental Protection Agency. Threshold of regulation policy-deciding whether a pesticide with a food use pattern

needs a tolerance. October 18. Available from: <www.epa.gov/fedrgstr/EPA-PEST/1999/October/Day-27/6042.pdf>.

USEPA, 2001. US Environmental Protection Agency. Toxicologic review of chloroform (CAS No.67-66-3). In support of summary information on the Integrated Risk Information System. EPA/635/R-01/001.

USEPA, 2003a. US Environmental Protection Agency. A Framework for a Computational Toxicology Research Program in ORD, Draft. Office of Research and Development, USEPA, Washington, DC. Available from: <www.epa.gov/comptox/comptox_framework.html>.

USEPA, 2003b. US Environmental Protection Agency. 2,6-Diiospropylnaphthalene; Temporary tolerances. Federal Register 68 (153), 47246–47353, August 8.

USEPA, 2005a. US Environmental Protection Agency. Guidelines for Carcinogen Risk Assessment. EPA/630/P-03/001F, Washington, DC.

USEPA, 2005b. US Environmental Protection Agency. Notice of filing of a pesticide petition amendment to regulations for residues of the biochemical pesticide 2,6-diisopropylnaphthalene (2,6-DIPN) in or on potatoes. Federal Register 70 (236): 73234–73236. December 9.

USEPA, 2006. OPPTS Harmonized Test Guidelines. Series 870 Health Effects Test Guidelines-Final Guidelines. Available from: <http://www.epa.gov/opptsfrs/publications/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/index.html>.

USFDA, 1995. US Food and Drug Administration. Food Additives; Threshold of regulation for substances used in food-contact articles. Federal Register 60(136): 36582–36596, July 17.

Usui, T., Mutai, M., Hisada, S., Takoaka, M., Soper, K.A., McCullough, B., Alden, C., 2001. CB6F1-rasH2 mouse: Overview of available data. Toxicol. Pathol. 29 (Suppl.), 90–108.

van Kreijl, C.F., McAnulty, P.A., Beems, R.B., Vynckier, A., van Steeg, H., Fransson-Steen, R., Alden, C.L., Forster, R., van der Laan, J.W., Vandenberghe, J., 2001. Xpa and Xpa/p53+/− knockout mice: overview of available data. Toxicol. Pathol. 29 (Suppl.), 117–127.

van Steeg, H., de Vries, A., van Oostrom, C.Th., van Benthem, J., Beems, R.B., van Kreijl, C.F., 2001. DNA repair-deficient Xpa and Xpa/p53+/- knock-out mice: nature of the models. Toxicol. Pathol. 29 (Suppl.), 109–116.

Vinggaard, A.M., Korner, W., Lund, K.H., Bolz, U., Petersen, J.H., 2000. Identification and quantification of estrogenic compounds in recycled and virgin paper for household use as determined by *in vitro* yeast estrogen screen and chemical analysis. Chem. Res. Toxicol. 13 (12), 1214–1222.

Waalkes, M.P., Ward, J.M., Liu, J., Diwan, B.A., 2003. Transplacental carcinogenicity of inorganic arsenic in the drinking water: induction of hepatic, ovarian, pulmonary, and adrenal tumors in mice. Toxicol. Appl. Pharmacol. 186 (1), 7–17.

Wagner, P.M., Nabholz, J.V., Kent, R.J., 1995. The new chemicals process at the Environmental Protection Agency (EPA): structure–activity relationships for hazard identification and risk assessment. Toxicol. Lett. 79, 67–73.

Wald, N.J., Cuckle, H., Brock, J.H., Peto, R., Polani, P.E., Wordford, F.P., 1977. Maternal serum alpha-fetoprotein measurement in antenatal screening for anencephaly and spina bifida in early pregnancy. Report of U.K. collaborative study on alphafetoprotein in relation to neural tube defects. Lancet 1, 1323–1332.

Wallace, K.B., Hausner, E., Herman, E., Holt, G.D., MacGregor, J.T., Metz, A.L., Murphy, E., Rosenblum, I.Y., Sistare, F.D., York, M.J., 2004. Serum troponins as biomarkers of drug-induced cardiac toxicity. Toxicol. Pathol. 32, 106–121.

Waters, M.D., Stack, H.F., Jackson, M.A., 1999a. Genetic toxicology data in the evaluation of potential human environmental carcinogens. Mutat. Res. 437 (1), 21–49.

Waters, M.D., Stack, H.F., Jackson, M.A., 1999b. Short-term tests for defining mutagenic carcinogens. IARC Sci. Publ. 146, 499–536.

Waters, M., Boorman, G., Bushel, P., Cunningham, M., Irwin, R., Merrick, A., Olden, K., Paules, R., Selkirk, J., Stasiewicz, S., Weis, B., Van Houten, B., Walker, N., Tennant, R., 2003. System biology and the chemical effects in biological systems (CEBS) knowledge base. Environ. Health Persp. Toxicogenom. 111 (1T), 15–28.

Weinberg, H.S., Krasner, S.W., Richardson, S.D., Thruston, A.D. Jr., 2002. The Occurrence of Disinfection By-Products (DBPs) of Health Concern in Drinking Water: Results of a Nationwide DBP Occurrence Study, EPA/600/R-02/068, US Environmental Protection Agency, Athens, GA.

WHO, 2005. World Health Organization. Concise International Chemical Assessment Document 66: 2,4,6-Tribromophenol and other simple brominated flame retardants, Geneva, Switzerland.

Wickner, S., Maurizi, M.R., Gottesman, S., 1999. Posttranslational quality control: folding, refolding, and degrading proteins. Science 286, 1888–1893.

Williams, G.M. Iatropoulos, M.J., 2001. Principles of testing for carcinogenic activity. In: Principles and Methods in Toxicology. A Wallace Hayes, E. Fourth Edition, Taylor and Francis, Philadelphia, pp. 959–1000.

Witt, K.L., Knapton, A., Wehr, C.M., Hook, G.J., Mirsalis, J., Shelby, M.D., MacGregor, J.T., 2000. Micronucleated erythrocyte frequency in peripheral blood of B6C3F$_1$ mice from short-term, prechronic, and chronic studies of the NTP Carcinogenesis Bioassay Program. Environ. Mol. Mutagen. 36, 163–194.

Woo, Y.T., Lai, D., 2003. Mechanisms of action of chemical carcinogens and their role in structure–activity relationships (SAR) analysis and risk assessment. In: Benigni, R. (Ed.), Quantitative Structure–Activity Relationship (QSAR) Models of Mutagens and Carcinogens. CRC Press, pp. 42–80.

Woo, Y.T., Lai, D., McLain, J.L., Manibusan, M.K., Dellarco, V., 2002. Use of mechanism-based structure–activity relationships analysis in carcinogenic potential ranking for drinking water disinfection byproducts. Environ. Health Persp. 110 (Suppl. 1), 75–87.

Yamamoto, S., Urano, K., Koizumi, H., Wakana, S., Hioki, K., Mitsumori, K., Kurokawa, Y., Hayashi, Y., Nomura, T., 1998a. Validation of transgenic mice carrying the human prototype c-Ha-ras gene as a bioassay model for rapid carcinogenicity testing. Environ. Health Perspect. 106 (Suppl. 1), 57–69.

Yamamoto, S., Urano, K., Nomura, T., 1998b. Validation of transgenic mice harboring the human prototype c-Ha-ras gene as a bioassay model for rapid carcinogenicity testing. Toxicol. Lett. 102–103 (473-8).

Yates III, Y.R., 2000. Mass spectrometry. From genomics to proteomics. Trends Genet. 16, 5–8.

Zeiger, E., Haseman, J.K., Shelby, M.D., Margolin, B.H., Tennant, R.W., 1990. Evaluation of four *in vitro* genetic toxicity tests for predicting rodent carcinogenicity: confirmation of earlier results with 41 additional chemicals. Environ. Mol. Mutagen. 16 (Suppl. 18), 1–14.

Zeise, L., 2001. Risk assessment of Genotoxic Carcinogens. In: Choy, W.N. (Ed.), Genetic Toxicology and Cancer Risk Assessment. Marcel Dekker, New York, pp. 299–320.