## 5.0 CHEMICAL DATA SETS FOR VALIDATION OF *IN VITRO* TOXICITY TESTS

### 5.1 Introduction

Breakout Group 4 discussed the selection of chemical data sets for validation of *in vitro* toxicity tests. The Breakout Group agreed that it would not develop specific lists of chemicals but would concentrate upon principles for the development of a database of chemicals that could be used in validation of individual tests or prediction models, and strategies for selection of the chemicals to be included in the database. Primary database development will most likely come from existing databases such as those available at the U.S. EPA, FDA, NCI, NTP, DOT, Galileo, Euclid, and others that are to be identified.

In addition to establishing criteria for primary database development, a set of criteria was developed for selecting chemicals for subset development. The chemicals in the subsets will be chosen from the primary database and will be used to validate individual tests or prediction models. The primary assumption in establishing criteria for subset development is that the purpose and proposed use of the test, the endpoint measured, the range of testable chemicals, and the prediction model must be clearly defined before chemical selection begins. Criteria that were considered important in selecting a set of reference chemicals were developed, as well as a set of fields considered relevant for the chemical reference database.

Lastly, the Breakout Group assembled a list of recommended actions that was divided into two parts: one that was database specific and one that was human toxicity specific.

### 5.2 Objectives

Before beginning a discussion of the primary database development, the Breakout Group defined some common points of reference and some points of agreement that would serve as the basis for discussions during the meeting. These are presented in the next sections.

#### 5.2.1 *Points of Reference*

(1) The main function of the Breakout Group was to develop a set of general principles that would be useful for choosing test chemicals for validation.

(2) The Breakout Group would attempt to identify databases, and other sources that contain the information necessary to choose the test chemicals, and define their uses and limitations.

(3) The Breakout Group agreed that it would not identify specific chemicals or develop lists of chemicals at this time.

#### 5.2.2 *Points of Agreement*

In addition to the three reference points, several items were set out by the Breakout Group to ensure that all members understood the exact aim of the discussion and their charge to the Breakout Group.

(1) It was agreed that the aim of the Breakout Group was to identify chemicals and supporting chemical information that can be used to validate replacement test(s) for acute toxicity tests.

(2) The chemicals used to validate a replacement test should cover the entire range of responses of the LD50 values. They should not be chosen to bracket just the range of classification used in the internationally agreed upon classification scheme(s).

(3) In addition to covering the entire range of responses, the chemicals chosen for use in a validation study should be uniformly distributed across that range, (i.e., there should not be a preponderance of either very toxic or non-toxic chemicals among those used).

(4) Identification of “chemical classes” is problematic. The basis for classification is the most significant issue. There was an unresolved discussion within the Breakout Group as to whether

classification should be done on the basis of chemical structure or mechanism of biological action. There was some discussion also about classifying according to use, such as "pesticide" or "food additive".

(5) The Breakout Group agreed that it is not necessary to be restricted to only one classification scheme. Chemicals could be classified by structure and by biological activity and/or use class. The classification approach would, by necessity, vary according to the type of test and its proposed uses.

(6) There are many public databases from which to draw information. These databases contain chemicals of concern to society. Investigators may not need, therefore, to use the proprietary databases such as the U.S. EPA OPP pesticides database or the FDA drug database to get the information and identify chemicals for use in tests for validation, but it would be helpful if information from those databases could be made available.

(7) There is a need for training sets of chemicals that can be used for method development, and validation sets of chemicals that can be used for confirming the predictive capacity of the tests.

(8) In selecting chemicals for use in validation studies, investigators need to consider the user community(ies) and assure that chemicals are chosen that meet their needs.

(9) The performance parameters of the *in vivo* tests must be clearly defined prior to chemical selection if the results of these tests are to serve as a baseline for judging success.

#### 5.2.3 *Definition of Responsibility*

Breakout Group 4 defined its responsibility as follows:

- To define what chemical data sets are required for validation studies;
- To define the information to be included as part of the data set;
- To identify existing resources;
- To recommend approaches for using existing data sets;
- To recommend approaches for developing new data sets.

The Breakout Group explored the possible use of such databases as the HPV database, the U.S. EPA pesticides database, the NTP chemical database, the FDA database of drugs and food additive chemicals, and the use of QSAR to predict toxicity of chemicals.

### 5.3 Current Status: Discussions Regarding the Use of the NTP and HPV Databases, and the Use of QSAR

#### 5.3.1 *The NTP Database*

The NTP chemicals were not tested for acute toxicity and therefore no LD50 data were developed. However, many were tested in 90-day studies, and some in 14-day studies, and these have associated target-organ toxicity data, as do the 2-year carcinogenicity studies. This information would be useful in validating *in vitro* tests for target-organ toxicity. The NTP database would be a useful component of any primary database of chemicals for validation.

Both the U.S. EPA pesticides database and the FDA drugs and food additive databases have associated LD50 data of good quality. However, there was some question about the ultimate accessibility of these data because of claims of confidentiality by the sponsors. Ease of access was a concern even where the data are not claimed to be confidential. Access through the Freedom of Information Act (FOIA) was discussed as a possibility, but this is a slow process and members of the Breakout Group expressed the desire that sources of unencumbered data should be used if they were available. Also, this approach may not provide the supporting information deemed necessary by the Breakout Group.

#### 5.3.2 *The HPV Database*

There was a short presentation of the classification of the chemicals that are part of the HPV Program of the U.S. EPA OPPT. Using only

696 pure chemicals on the list and classifying them according to chemical structure, a list of 45 chemical classes with from 4 to 72 chemicals per class was developed. This classification is based solely upon chemical structure and each chemical is assigned to one class only. There is no indication of how many of these chemicals fall into more than one class. There is also no indication of which of these chemicals have LD50 data, the quality of these data where they exist, or the range of responses that is covered. Without this information, it is impossible to tell which of the HPV chemicals would be useful as validation chemicals. In addition, the chemicals on the HPV list are primarily industrial chemicals and their use as validation chemicals might not meet the needs of all user communities.

#### 5.3.3 *QSAR Methods and Structure-Activity Methods for Toxicity*

QSAR methods can be applied to the problem of developing models to predict toxicity endpoints or toxic classes given sufficient quantity and quality of data.

The basis for the prediction of toxicity from chemical structure is that the properties of a chemical are implicit in its molecular structure. Biological activity can be expressed as a function of partition and reactivity. For a chemical to be able to express its toxicity, it must be transported from its site of administration to its site of action and then it must bind to or react with its receptor or target. This process may also involve metabolic transformation(s) of the chemical and its metabolites.

The application of QSAR principles to the prediction of the toxicity of new or untested chemicals has been achieved in a number of different ways and covers a wide range of complexity. The common feature of these approaches is that their starting point is a mechanistic hypothesis linking chemical structure and/or functionality with the toxicological endpoint of interest. A number of such "*in silico*" methodologies have also been applied with varying degrees of success to the evaluation of LD50 values and MTDs, and some are available commercially (e.g., DEREK, MCASE, and TOPKAT).

The prediction of toxicity from chemical structure and physical properties can make a valuable contribution to the reduction of animal usage in the screening out of potentially toxic chemicals at an early stage and in providing data for making positive classifications of toxicity. However, such methods should also be validated, using protocols similar to those described in these pages, so as to assess their potential effectiveness in assessing acute toxicity.

### 5.4 Identification of Needs

#### 5.4.1 *Selection of Test Chemicals for Validation of In Vitro Tests*

In the context of using *in vitro* tests to replace or reduce animal usage, the performance of an *in vitro* test or an *in silico* test is assessed by its capability of correctly predicting the *in vivo* response. However, it is unreasonable to expect that the *in vitro* test will be able to predict the result of an *in vivo* test with any more accuracy than would a repeat *in vivo* test.

The assessment of any new test would be best accomplished by selecting a series of reference chemicals that cover the full range of responses, from negative, to weak, to intermediate, to strong. Selection of only strongly active chemicals will not provide information on the discriminating ability of a test, or its ability to detect the weakly active chemicals. The absence of chemicals known to be inactive will not allow a determination of the ability of the test to identify chemicals without activity, or of the false positive rate of the test.

#### 5.4.2 *Evaluating the Quality of Data Used to Develop the Chemical Data Set*

A major challenge facing researchers developing either *in vitro* or *in silico* models is the sparse availability of high quality data derived from experiments with animals, or from human monitoring studies and clinical reports. Biological data which do not meet today's stringent requirements of acceptability,

particularly historical data generated prior to the advent of standardized test guidelines, but which are nevertheless of acceptable quality, can be used to validate newly developed test methods.

The Breakout Group discussed the establishment of a primary database from which sets of chemicals could be drawn for use as validation chemicals for specific tests or prediction models. In addition to the need to establish criteria for primary database development, a set of criteria for selecting chemicals for subset development should be developed.

## 5.5 Conclusions

### 5.5.1 *Primary Assumption for Data Set Development*

The primary assumption in establishing criteria for data set development is:

- The purpose and proposed use of the test, the endpoint measured, the range of testable chemicals, and the prediction model must be clearly defined before chemical selection begins.

Such information is used as the guide for choosing the most appropriate materials for evaluating whether or not the test method would satisfy its proposed uses.

### 5.5.2 *Criteria for Data Set Development*

The following criteria were established for data set development.

(1) The chemicals selected must be consistent with the test protocol and its prediction model.
   - The chemicals selected must be physically and chemically compatible with the test system.
   - The relevant chemical classes must be included.
     — The definition of chemical class is context-specific.
     — The developers of the test must specify the parameters that define the class.
     — The chemicals must be independently chosen.

(2) The toxicity must cover the range of response with uniform distribution.

(3) The number of chemicals used in the subset will depend on the nature of the test and the questions being asked, and should be determined with statistical advice.

### 5.5.3 *Primary Data Base Development*

Primary database development will most likely come from existing databases such as those available at the EPA, FDA, NCI, NTP, DOT, Galileo, Euclid, and others that are to be identified. As noted above, the more publicly available the database, the easier it will be to access the data. The problem, of course, is quality control of the data that goes into the database. The two most important considerations in assembling the primary set of reference chemicals are: (a) *in vivo* data must be of high quality, cover the range of response, and be uniformly distributed over that range and (b) the chemicals selected must be commercially available and their specifications (including purity) must be available.

The Breakout Group noted that there were some unresolved questions surrounding the issue of quality control. The first concerned protocol and, specifically, route of administration. There was some discussion about whether to accept tests done by all routes of administration or to limit the database to the oral route. It was decided that oral and inhalation routes were acceptable and that the dermal route while important for some purposes, was not of primary concern for most acute toxicity studies. However, the Breakout Group agreed, that if data were available from all routes, such data should be included in the database.

The Breakout Group agreed that, where possible, the data used should be derived from generally recognized test guidelines, such as those from the U.S. EPA, OECD, ICH, etc., because data from these guidelines carry a higher degree of assurance than data from an undefined or novel protocol. An issue that was not resolved was whether or not to require that the data used in the

database be from a study done according to Good Laboratory Practices (GLPs).

### 5.5.4 *Criteria for Choosing Reference Chemicals: Reference Test Data*

The following criteria were considered of prime importance in selecting a set of reference chemicals.

(1) The reference data for the endpoint predicted are available.

(2) The performance characteristics of the reference test must be defined.
- Variation will be introduced by protocol (including animal strain) differences.
- Different agencies use different protocols.
- The between-laboratory reproducibility of the test must be determined.
- The limitations of the reference test must be known.

(3) The reference test data must be of high quality.

(4) The protocol used must be available for review.

(5) Generally accepted methods (e.g., OECD, EPA, FDA, ICH guidelines) should have been used to generate the data.

(6) Details of the study should be available and ideally should satisfy ICCVAM and ECVAM Submission Guidelines.

(7) Study has sufficient supporting information. Ideally, GLPs should have been followed in study development.

(8) Other important considerations:
- The chemicals should be drawn from a wide range of structural and use classes.
- They should not be highly reactive, corrosive, or controlled substances.

### 5.5.5 *Database Fields*

The Breakout Group defined some of the information fields it considered relevant for the chemical reference database. These fields should include information about the identity, purities, and properties of the chemicals, and detailed reference test data.

(1) Chemical Information
- Name and Chemical Abstract Service (CAS) Number;
- Structure (coded, e.g., using Simplified Molecular Input Line Entry Specification [SMILES] nomenclature);
- Physical chemical characteristics (e.g., $K_{ow}$, pKa, water solubility, molecular weight., physical state);
- Purity;
- Chemical class (e.g., The International Union of Pure and Applied Chemistry [IUPAC] and use).

(2) Reference Test Data
- Specifications of chemical used in reference test;
- Information concerning the protocol used to generate the data;
- Endpoint value (e.g., LD50) and variance term (e.g., confidence interval), if available;
- Species, strain, sex;
- Route of exposure; duration of exposure;
- Information needed by Breakout Groups 2 and 3 should also be included.

## 5.6 Recommended Actions

### 5.6.1 *Rodent Toxicity Database*

(1) A study should be undertaken of existing databases to determine:
- The variation in the rodent LD50 introduced by differences in protocols;
- The within- and between-laboratory reproducibility of the rodent LD50 test and other acute toxicity tests that will be used as reference tests.

(2) An expert committee should be convened that will assemble a reference set of test chemicals from existing databases according to the criteria specified.

### 5.6.2 *Human Toxicity Database*

(1) There is a need to build upon the foundations of the MEIC and MEMO exercises.

(2) An expert panel should review the MEIC/MEMO approach for measuring acute toxicity parameters in humans.

(3) A consensus standard approach for measuring acute toxicity parameters is necessary.

(4) Existing sources of information need to be carefully searched in order to assure all relevant human data are obtained.

(5) A mechanism prospectively should be established to: (a) gather human toxicity data from hospital/Poison Control Center (PCC) sources; (b) retrieve existing human toxicity data; (c) collect and organize human toxicity data as accidents occur. Biomonitoring data should also be collected. Such information could define sub- or non-toxic levels, and be used to see if they overlap with the range of reported toxic levels.