

Introduction and Summary. Genotoxicity and Carcinogenicity Databases: An Assessment of the Present Situation

by Silvio Parodi* and Michael D. Waters†

A central purpose of this meeting was to review the present status of genotoxicity and carcinogenicity databases. We wanted to analyze and discuss the current level of development of databases in these specialty areas and to determine whether they are satisfactory for the ways in which they are being used. To answer the question as to the state of development of these kinds of databases, we have to address first the most critical and most frequent ways in which such databases are employed.

We envisage three major types of use: *a*) for purposes of toxicological review and/or regulation, *b*) for chemical structure-activity evaluations, and *c*) for assessing correlative and mechanistic relationships between genotoxicity and carcinogenicity. To build a reasoned and articulated judgment about the genotoxic and carcinogenic hazard presented by a given chemical, two types of databases are useful: a database that includes expert assessment and a database that summarizes essential experimental data but without expert assessment.

The International Agency for Research on Cancer (IARC) database is an example of an evaluated database of selected chemicals considered important in terms of the extent and intensity of human exposure and of potential hazard. This internationally peer-reviewed database was described in this meeting by H. Vainio. It offers not only the basic information about the genotoxicity or carcinogenicity of a given compound, but also the judgments and the overall conclusions of a panel of experts that have attempted to synthesize the available data. Such a database is extremely valuable for the evaluative and/or regulatory purposes mentioned previously.

A recommendation for the future is that efforts to build this type of database be extended to new chemicals for which human exposure is likely. Of special interest are genotoxins and carcinogens present in our diet, to which we often have significant levels of exposure.

An analogous evaluated database on genotoxicity and cancer, the Gene-Tox database, was described by A. Auletta. In addition to providing published comprehensive analyses of test systems and chemicals, this database is currently available on the National Library of Medicine's TOXNET system.

A major evaluated cancer database on long-term studies in mice and rats is the National Cancer Institute/National Toxicology Program summarized by J. Huff. This important database is a major factor in the toxicological evaluation of chemicals as well as in the analysis of genotoxicity and related tests for their ability to predict cancer in experimental animals.

If a chemical has not been included in a more articulated type of database such as the NCI/NTP database, then a second type of database may be consulted. This type of database summarizes the most essential data elements related to long-term carcinogenicity bioassays or to short-term tests for genotoxicity. Also for this type of database it may be important to use the judgment of a group of experts to evaluate the technical quality or adequacy of the performance of a given test or experiment. However, the main purpose of such a database is to present in some detail all the essential elements of experimental data.

It is evident that such factual databases ideally should be as comprehensive as possible. Obviously, there are practical limitations. These include database building costs such as abstracting and programming and user costs as search time and charges. The ideal situation for the user would be to have a comprehensive factual database for carcinogenicity and a similar one for genotoxicity available in personal computer format.

It was apparent from the meeting that existing databases of this second type already display some similarities in data organization and content. Two databases discussed in this meeting, the TD₅₀ carcinogenicity database described by L. S. Gold and the EPA/IARC genetic activity profile database described by M. Waters are examples of this second type of database. These two databases may be sufficiently close in information content that a common design structure could be contemplated. Future interaction among individuals participating in this field (builders and users of these databases) could lead to a common consensus about the detailed structure of genotoxicity and carcinogenicity databases. Perhaps an initial focus on critical data elements would enhance prospects for the realization of the desired compatibility of such databases.

A concerted effort involving the coordinated contribution of experts from different countries could have as its future goal the building of a comprehensive and universal database on genotoxicity and carcinogenicity. Progress in this direction could greatly enhance the broad utility of both types of databases.

*Istituto Nazionale per la Ricerca sul Cancro, Genova, Italy.

†U.S. Environmental Protection Agency, Research Triangle Park, NC 27711.

Address reprint requests to S. Parodi, Istituto Nazionale per la Ricerca sul Cancro, Viale Benedetto XV, no. 10, 16132 Genova, Italy.

Among the users of databases, we have mentioned up to now only those seeking specific chemical information for purposes of toxicological evaluation or regulation. There are two other types of individuals interested in genotoxicity and carcinogenicity databases that in principle, need comprehensive data. One group is represented by investigators involved in structure-activity relationship studies. For a given class of chemicals or for a given significant molecular fragment as described by G. Klopman or the concept of a structural alert as described by J. Ashby, this group needs data subsets that are sufficiently large to enable a detailed description of the structural basis of the toxicological behavior of chemical analogues.

Another group of individuals that ideally requires comprehensive databases is represented by investigators involved in assessing the relationships among short-term tests and relationships between these tests and carcinogenicity bioassays. Several publications in the literature and recent results reported in this meeting show that smaller, selective databases can offer different views of the performance characteristics of given tests. Correlations of the results of short-term tests with carcinogenicity can be different for different chemical classes. Additional complications have been introduced with the recognition of the so-called "nongenotoxic carcinogens" as will be discussed later.

Several statistical and mathematical techniques are available for the comparative analyses mentioned above. These approaches were discussed by S. Parodi and R. Benigni. In addition to the more classical parametric and nonparametric statistics, interesting examples of the application of factor analysis and cluster analysis were reported during this meeting. The mathematical aspects of the analyses are obviously relevant, but they require a substratum of sufficiently large data subsets representing families of chemicals in order to exhibit more completely their potential usefulness.

Standing between the "evaluated expert" database of IARC, and the "summary of essential data" databases, two participants, S. Nesnow and D. Brusick, each presented in this meeting a database in which the information about a specific chemical is synthesized in quantitative values taking into account several factors relevant to carcinogenicity or genotoxicity, respectively. This elaborated information could be viewed as a component of future more complex and integrated expert systems, and shows a new, promising avenue in the application of genotoxicity and carcinogenicity data to future studies in hazard evaluation.

Finally, several of the participants described more specialized databases that are devoted to a specific area. We will mention only a few examples. Data bases relative to pesticides or to food additives were described by K. Dearfield and D. Benz, respectively. These databases considered all kinds of toxicological data, not just genotoxicity and carcinogenicity, and the speakers emphasized how these data are used in the regulatory process. An

additional family of data (a potential database) discussed during the meeting by W. Anwar, is that data derived from human biomonitoring. Such data are especially relevant to the processes of carcinogenesis risk assessment.

The possibility of using genotoxicity and carcinogenicity data derived from industry (e.g., data submitted to national and international regulatory bodies) also was addressed during the discussion. It was observed that submitted data tend to be mainly for nongenotoxic/noncarcinogenic chemicals and that data on proprietary chemicals giving positive results are frequently not accessible. No clear consensus about the possibility of enriching genotoxicity and carcinogenicity databases with this type of data was reached during the meeting.

An ultimate limitation in our ability to analyze existing data relates to our lack of understanding of mechanisms of genotoxicity and carcinogenicity. In fact, such an understanding may be derived in part from the construction and analysis of databases. Despite this limitation, important conclusions about the complex relationship existing between genotoxicity and carcinogenicity have been reached and published in the literature with comparative assessments of the type mentioned previously. These assessments tend to use the more comprehensive databases on genotoxicity and carcinogenicity. Indeed, experimental oncology has demonstrated that in the process of carcinogenesis, not only irreversible alterations in the genome, but also epigenetic phenomena involving stimulation of cell proliferation, stimulation of clonal expansion of preneoplastic cells, and modulation of differentiation can play a dramatic role in increasing the incidence of tumors.

Recent experimental investigations using databases have quantitatively measured the practical relevance of this nongenotoxic component of the process of carcinogenesis. The usage of various relatively small databases derived from this type of investigation has sometimes introduced bias and partial discordance among different analyses of the data. This situation is obviously an additional stimulus in the direction of building larger, more comprehensive databases, especially for the so-called nongenotoxic carcinogens.

The general impression of the participants of the meeting was that databases on genotoxicity and carcinogenicity are not only essential instruments for studying human hazards related to cancer and mutation, but are also critical tools for assessing the progress and the problems that lie ahead in this field. For these reasons, it is worthwhile to improve and extend these important sources of information in the near future.

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.