

Predicting Mammalian Mutagenesis by Submammalian Assays: An Application of Database GEN

by F. E. Würgler*

A database containing qualitative information on the genotoxic activity of about 3000 chemicals is described. The initial aim for the construction of the database was to develop an instrument for comparing the performance of different genotoxicity assay systems. One application of the database is the prediction of expected results in any genotoxicity assay for chemicals that were tested in a small number of genotoxicity assays. The Bayesian prediction is calculated based on the sensitivities and specificities between any predictive test and the target test for which the prediction is to be determined. The predictivity of the system for *in vivo* mammalian assays is at present (with the exception of the micronucleus assay and the *in vivo* sister chromatid exchanges) limited, in particular because of the limited number of chemicals tested in the expensive *in vivo* assays and, in addition, due to the lack of sufficient information on negative compounds. A continued updating of the database will possibly help to overcome some of the present difficulties.

Introduction

The aim of genetic toxicology is to detect those chemical compounds capable of inducing genetic damage in man. This category of chemicals includes, as the major components known today, the potential human germ cell mutagens as well as the genotoxic carcinogens. For a number of reasons, direct tests in humans are impossible. Therefore, model systems have to be used. In general, the *in vivo* systems nearest to the human situation that can be used on a large scale are rodent models, in particular the mouse. For a number of reasons, e.g., high costs, long duration of experiments, and reduction of the use of experimental animals for ethical reasons, it has become an ongoing effort to replace at least part of the *in vivo* tests by experiments with submammalian species and/or *in vitro* assays. Before relying on the results of the short-term tests (STTs), these tests have to be validated against the results obtained with *in vivo* systems and where possible against information available for humans.

Every possible effort was made to evaluate the STTs as predictive tests for carcinogenic potential. For this purpose, the results of STTs were compared with the results obtained in rodent lifetime carcinogenicity studies. To our knowledge, no similar, systematic studies have been undertaken so far in order to validate the STTs for their predictivity for genotoxic effects observed in animal models *in vivo*. Such studies may be helpful in deciding which STTs might be optimal substitutes for *in vivo* mammalian genotoxicity tests.

The available genotoxicity assays differ in important biological parameters, such as pharmacokinetics and metabolism, as well

as the genetic end points studied. The database GEN may be used for comparisons of genotoxicity assays. Some experience gained is reported here.

The Database GEN

Each chemical in the database GEN is characterized by the Chemical Abstracts Services Registry Number (CASRN). The assays included and the major sources of information are given in Table 1. Results from any individual assay or subassay are coded as 1 for negative; 2 for inconclusive; and 3 for positive. Missing data are coded as 0. The code 1 + 3 represents the conclusive results; the code 1 + 2 + 3 represents the nonzero results. For future versions of the program, the introduction of 4 for weak positive will be considered.

CASRN and the names of the chemicals are contained in one file; CASRN and assay results are in another file. Table 1 gives the list of the major groups of genotoxicity assays and genetic end points included in the database.

The data contained in the database have been predominantly taken from the following sources: Gene-Tox reports, publications of the U.S. National Toxicology Program (NTP) (in particular the results of *Drosophila* assays), International Collaborative Studies, some reviews, and the database published by Palajda and Rosenkranz (1).

The Software

The software is written in Fortran 77 and runs on a 8700 VAX of the Computing Center of the Swiss Federal Institute of Technology in Zürich. The features of the software are discussed below.

*Institute of Toxicology, Swiss Federal Institute of Technology and University of Zürich, Schorenstrasse 16, CH-8603 Schwerzenbach near Zürich, Switzerland.

Table 1. Major groups of assays included in the database GEN.

Code	Assay	Code	Assay
ECR	<i>E. coli</i> rec assay	DES	<i>Drosophila melanogaster</i> eye assay single spots
pIA	<i>E. coli</i> DNA repair	DET	<i>Drosophila melanogaster</i> eye assay twin spots
Bsr	<i>B. subtilis</i> repair	DEU	<i>Drosophila melanogaster</i> unstable eye system
IND	Induc test (lambda induction in <i>E. coli</i>)	DUE	<i>Drosophila melanogaster</i> unstable eye system (<i>exr</i> ⁻)
ASO	SOS chromotest with activation	DWA	<i>Drosophila melanogaster</i> wing assay all spots
NSO	SOS chromotest without activation	DWS	<i>Drosophila melanogaster</i> wing assay small single spots
PrM	<i>Proteus mirabilis</i> DNA repair	DWL	<i>Drosophila melanogaster</i> wing assay large single spots
UMA	umu test with activation	DWT	<i>Drosophila melanogaster</i> wing assay twin spots
UMN	umu test without activation	DWE	<i>Drosophila melanogaster</i> wing assay all spots (<i>exr</i> ⁻)
UDS	Unscheduled DNA synthesis	CHO	Specific gene mutations in Chinese hamster ovary cells
Drp	DNA repair assay using eukaryotic systems	V79	Specific gene mutations in V79 Chinese hamster cells
Sty	Salmonella mutagenicity assay (all strains)	Mly	Specific gene mutations in mouse lymphoma L5178Y cells
S9A	<i>S. typhimurium</i> TA 98 with activation	Cvt	Mammalian cytogenetics <i>in vitro</i>
S0A	<i>S. typhimurium</i> TA 100 with activation	MCC	Mammalian cell culture aneuploidy (hyperploidy)
S2A	<i>S. typhimurium</i> TA 102 with activation	DLA	Dominant lethal assay
S5A	<i>S. typhimurium</i> TA 1535 with activation	Mnt	Micronucleus test
S7A	<i>S. typhimurium</i> TA 1537 with activation	Htr	Heritable translocations in the mouse
S8A	<i>S. typhimurium</i> TA 1538 with activation	Msl	Mouse specific locus test
S9N	<i>S. typhimurium</i> TA 98 without activation	Msp	Mouse spot test
S0N	<i>S. typhimurium</i> TA 100 without activation	Cbm	Bone marrow cytogenetics <i>in vivo</i>
S2N	<i>S. typhimurium</i> TA 102 without activation	Cle	<i>In vivo</i> leukocyte cytogenetics
S5N	<i>S. typhimurium</i> TA 1535 without activation	Cco	<i>In vivo</i> oocyte cytogenetics
S7N	<i>S. typhimurium</i> TA 1537 without activation	Csg	<i>In vivo</i> spermatogonia cytogenetics
S8N	<i>S. typhimurium</i> TA 1538 without activation	Csp	<i>In vivo</i> spermatocyte cytogenetics
EcW	<i>E. coli</i> WP2 reverse mutation assay	MFG	Mammalian female germ cells aneuploidy (hyperploidy)
SCA	Pooled data from <i>Saccharomyces cerevisiae</i> (all assays)	MMG	Mammalian male germ cells aneuploidy (hyperploidy)
YEF	<i>Saccharomyces cerevisiae</i> forward mutations	MSC	Mammalian somatic cells <i>in vivo</i> aneuploidy (hyperploidy)
YER	<i>Saccharomyces cerevisiae</i> reverse mutations	SAB	Sperm abnormality assay
YEH	<i>Saccharomyces cerevisiae</i> homozygosis (recombination or conversion)	SCE	Sister chromatid exchanges (all assays)
YEC	<i>Saccharomyces cerevisiae</i> gene conversion	SCT	Sister chromatid exchanges <i>in vitro</i>
YMR	<i>Saccharomyces cerevisiae</i> mitotic recombination	SCV	Sister chromatid exchanges <i>in vivo</i>
YEA	<i>Saccharomyces cerevisiae</i> aneuploidy	SCW	Sister chromatid exchanges (all assays including weak responses)
YEM	<i>Saccharomyces cerevisiae</i> mitochondrial mutations	STW	Sister chromatid exchanges <i>in vitro</i> including weak responses
Spo	Genetic effects in <i>Schizosaccharomyces pombe</i>	SVW	Sister chromatid exchanges <i>in vivo</i> including weak responses
NAN	<i>Neurospora crassa</i> aneuploidy	HMA	Host-mediated mutagenicity assay
AAN	<i>Aspergillus nidulans</i> aneuploidy	Bfl	Mutagenicity assay using body fluids
DAA	Pooled data from <i>Drosophila melanogaster</i>	PAN	Plant systems aneuploidy
DRL	<i>Drosophila melanogaster</i> sex-linked recessive lethal test	BHK	Transformation of BHK21 cells
DTR	<i>Drosophila</i> heritable translocations	3T3	Balb/C-3T3 neoplastic transformation assay
DCM	<i>Drosophila melanogaster</i> chromosome mutation	SHE	Transformation of Syrian hamster embryo cells
DCL	<i>Drosophila melanogaster</i> clastogenicity	C3H	Transformation of C3H cells
DAN	<i>Drosophila melanogaster</i> aneuploidy (Dellarco table)	MpR	Mouse prostate transformation assay
Dan	<i>Drosophila melanogaster</i> aneuploidy	Vet	Viral enhancement systems (transformation assay)
DNG	<i>Drosophila melanogaster</i> nondisjunction gain	IME	Inhibition of intracellular molecular exchange
DNL	<i>Drosophila melanogaster</i> nondisjunction loss	PRO	Promoting activity
SMR	Pooled data from SMART assays (eye and/or wing)	ACR	Animal carcinogens
DEA	<i>Drosophila melanogaster</i> eye assay all spots	HCR	Human carcinogens

Applications. DATA RETRIEVAL. Data retrieval routines allow one to obtain, for any chemical defined by its CASRN, a list containing all nonzero assay results stored in the database. Because the primary goal in constructing the database was to develop an instrument for comparisons of assays and to calculate predictions, no references to original publications were included. The program contains, however, with every assay description, a list of the secondary sources from which data have been obtained. In the printout, this list is printed together with the assay description, and it allows, although with limited convenience, one to find the original publication containing the original information by checking the secondary sources.

Another feature of the program allows one to get lists of chemicals showing any predetermined pattern of genotoxic activity (e.g., the chemicals positive with activation in Salmonella TA98 and TA100, but negative in the *in vivo* bone marrow micronucleus test). This feature is helpful if, in a testing program,

unexpected combinations of test results are obtained. A list of the chemicals showing the same or a very similar pattern may help one to develop an experimentally testable hypothesis to explain the basis of the unexpected pattern.

In certain instances, it is interesting to get a list of chemicals giving opposite results in two particular assays (e.g., negative for gene mutation but positive for recombination). In the same basis, it is possible, by selecting pairs of *in vitro* assays with and without metabolic activation, to check which compounds need metabolic activation and which are direct-acting mutagens.

For special purposes, it is possible to get a dump of the whole database with the original codes attached to every individual chemical and a summary for every individual chemical on the number of nonzero, conclusive, positive, negative, and inconclusive assay results present.

INFORMATION ON ASSAYS. By selecting the code for a particular genetic end point in one particular assay and the code for the

assay result (positive, negative, inconclusive, or any combination thereof), one can obtain a list, in CASRN order, of all chemicals tested in the assay leading to the particular test result. At the end of the list, the total number of chemicals found is given.

INFORMATION ON A SET OF ASSAYS. The database can be used to calculate the Hemming distances between any pair of assays. The distance matrix can be structured in such a way that it can serve as input to BMDP. With the BMDP program (2), a cluster analysis can be performed. In one application we compared the different assays used to detect chemically induced recombination in different organisms. Comparing the *Drosophila* and yeast data, the analysis indicates that the results from the *Drosophila* somatic assays appear to be more similar to the yeast gene conversion data than to the yeast mitotic recombination data. Because gene conversion is, so far, only known to occur in meiotic cells of *Drosophila*, this result encouraged us to initiate studies aimed at detecting somatic gene conversion in *Drosophila*. Another technique to get some information on the comparative performance of a set of assays is the determination of kappa values for pairs of assays (3).

PREDICTING ASSAY RESULTS. Theoretically, the information contained in the database can be used to calculate predictions for the outcome of any genotoxicity assay for any chemical for which minimal information on its genotoxic potential is available, e.g., for which at least one genotoxicity assay has been performed. As more experimental information is available, the precision of the predictions improves.

Definitions. Tests are all the assays (with their individual genetic end points) contained in the database. The target test is the particular genetic end point of an assay for which the prediction is to be calculated.

Method. The predictions are calculated by applying the Bayes' theorem to the sensitivities and specificities calculated for any pair of assays available in the database. Sensitivity describes what fraction of chemical found positive in the target assay was also positive in the predictive assay. Specificity describes what fraction of chemicals negative in the target test was also negative in the predictive assay. The Bayesian analysis is described in detail by Pet-Edwards et al. (4,5).

Predictions for Chemicals Present in the Database. For any chemical that has at least one entry in the database, the prediction for any test may be calculated. The procedure is as follows: *a*) Select the chemical by entering the CASRN. Then the system presents on the screen all the assay results for this chemical that are in the database and which might be used to calculate a prediction. *b*) Then the test for which the result should be predicted (target test) is selected. As the system knows the target test, it starts to calculate all the relevant sensitivities and specificities, that is, those between the target test and any test for which the chemical has experimental data. The fact that the sensitivities and specificities are calculated with every application makes sure that we have a learning system making use immediately of any new data entered into the database. For the adjustment of sensitivities and specificities based on small numbers of chemicals and those with numerical values of 1.0 and 0.0, the procedure described by Ennever and Rosenkranz (6) was used. The system now classifies the assays available for the calculation of the predictions according to high and low values for the sensitivity and specificity (4-6). *c*) The prediction for the result expected in the target

test may be calculated based on all predictive tests available or by using only some selected tests, e.g., selected according to their classification.

Applications. The applications approach may be used to predict a nonexistent result, or to check the expectation, if experimentally an inconclusive result was obtained, or to check the degree of confidence for a conclusive result (e.g., an unexpected negative result).

For an exploratory data analysis, the system allows one to change, temporarily, one or more assay results for the chemical under study. This allows one to study the change in prediction if a predictive assay with an inconclusive result would have had a positive or a negative response, for example. Or one may check the change in prediction (e.g., for carcinogenicity as the target test) and how the prediction would be improved by adding an additional conclusive result from a not-yet performed assay. In this way one might decide whether the test would add useful information to the genotoxic profile of the chemical and whether it is worth conducting the actual experiment.

A third application allows one to enter the information on assay results for a chemical not present in the database. This information is not included in the database and disappears as the application is terminated. Upon selection of the target test, the system uses the database to calculate the relevant sensitivities and specificities, and the analysis continues as described above.

Experiences. In connection with the attempts to reduce animal experiments, we became interested in using our system to study the predictivity of STTs not only for carcinogenesis but for genotoxic effects in animals *in vivo* (in particular, the mouse). To make a long story short, we find that for the *in vivo* micronucleus test and the *in vivo* sister chromatid exchange test, fairly reasonable predictions are possible, but for the other *in vivo* assays, such as the mouse specific locus test, the mouse heritable translocation test, and the mouse spot test, predictions have not yet been possible. The reasons for this are *a*) that a relatively small number of chemicals tested in these assays are contained in the database, *b*) these chemicals have not been systematically tested in the STT, and *c*) the number of chemicals reported negative in an *in vivo* assay is limited, thus the calculation of reliable sensitivities and specificities becomes the major problem.

Because the database is predominantly based on the Gene-Tox reports, we expect that an update of the database with the data published after the completion of the Gene-Tox phase 1 might lead to some improvements of the system to study predictions for the *in vivo* assays other than carcinogenicity.

REFERENCES

1. Palajda, M., and Rosenkranz, H. S. Assembly and preliminary analysis of a genotoxicity data base for predicting carcinogens. *Mutat. Res.* 153: 79-134 (1985).
2. BMDP Statistical Software. University of California, Berkeley, CA, 1983.
3. Mendelson, M. L., and Moore, D. H. Kappa, a measure of agreement for data comparing mutagenicity and carcinogenicity (abstract). *Environ. Mutagen.* 6: 480 (1984).
4. Pet-Edwards, J., Chankong, V., Rosenkranz, H. S., and Haines, Y. Y. Application of the CPBS method to the Gene-Tox database. *Mutat. Res.* 153:187-200 (1985).
5. Pet-Edwards, J., Haines, Y. Y., Chankong, V., Rosenkranz, H. S., and Ennever, F. K. *Risk Assessment and Decision Making Using Test Results.* Plenum Press, New York, 1989.
6. Ennever, F. K., and Rosenkranz, H. S. Selection of batteries in an industrial setting. *Environ. Mutagen.* 9: 359-361 (1987).