Toxicology and Applied Pharmacology 233 (2008) 34–44

Contents lists available at ScienceDirect

Toxicology and Applied Pharmacology

journal homepage: www.elsevier.com/locate/ytaap

Databases applicable to quantitative hazard/risk assessment—Towards a predictive systems toxicology

Michael Waters *, Marcus Jackson

ILS, Inc., P.O. Box 13501, Research Triangle Park, NC 27709, USA

ARTICLE INFO

Article history:
Available online 2 July 2008

Keywords:
Data integration
Toxicogenomics
Predictive toxicology
Carcinogenicity
Mode-of-action
Genotoxicity

ABSTRACT

The Workshop on The Power of Aggregated Toxicity Data addressed the requirement for distributed databases to support quantitative hazard and risk assessment. The authors have conceived and constructed with federal support several databases that have been used in hazard identification and risk assessment. The first of these databases, the EPA Gene-Tox Database was developed for the EPA Office of Toxic Substances by the Oak Ridge National Laboratory, and is currently hosted by the National Library of Medicine. This public resource is based on the collaborative evaluation, by government, academia, and industry, of short-term tests for the detection of mutagens and presumptive carcinogens. The two-phased evaluation process resulted in more than 50 peer-reviewed publications on test system performance and a qualitative database on thousands of chemicals. Subsequently, the graphic and quantitative EPA/IARC Genetic Activity Profile (GAP) Database was developed in collaboration with the International Agency for Research on Cancer (IARC). A chemical database driven by consideration of the lowest effective dose, GAP has served IARC for many years in support of hazard classification of potential human carcinogens. The Toxicological Activity Profile (TAP) prototype database was patterned after GAP and utilized acute, subchronic, and chronic data from the Office of Air Quality Planning and Standards. TAP demonstrated the flexibility of the GAP format for air toxics, water pollutants and other environmental agents. The GAP format was also applied to developmental toxicants and was modified to represent quantitative results from the rodent carcinogen bioassay. More recently, the authors have constructed: 1) the NIEHS Genetic Alterations in Cancer (GAC) Database which quantifies specific mutations found in cancers induced by environmental agents, and 2) the NIEHS Chemical Effects in Biological Systems (CEBS) Knowledgebase that integrates genomic and other biological data including dose–response studies in toxicology and pathology. Each of the public databases has been discussed in prior publications. They will be briefly described in the present report from the perspective of aggregating datasets to augment the data and information contained within them.

Introduction

National and international regulatory and public health organizations, and industries that utilize modern technologies in drug and chemical safety assessment, recognize the value of the toxicological triage process that accompanies hazard classification and risk assessment. These organizations have traditionally utilized the published literature and public data repositories to provide the data and information needed for this work. These data and information have been based on the application of conventional testing methodologies, most of which were developed more than two decades ago. More recently most of these same organizations have recognized the potential utility of toxicogenomics data in hazard identification and classification, but because of the costs involved in toxicogenomics experiments, the sheer volume of data generated, and the difficulty of properly interpreting the data, the utilization of toxicogenomics data, particularly by the regulatory community, has progressed slowly.

Global technologies including cDNA and oligonucleotide microarrays, protein chips, mass spectrometry and nuclear magnetic resonance (NMR)-based molecular profiling can simultaneously measure the expression of numerous genes, proteins, and metabolites, respectively, thus providing the potential to accelerate the discovery of toxicant pathways, modes of action, and specific chemical and drug targets. Toxicogenomics combines conventional toxicology study designs with global genomics technologies and appropriate genetic, pharmacological and toxicological models (Fig. 1) to provide a comprehensive view of the function of the genome and the biochemical machinery of the cell under stress. Examples of anticipated regulatory applications of molecular expression technologies include predictive gene signatures for toxicity and carcinogenicity, support for low dose extrapolation (e.g. threshold vs. non-threshold), and classification of

* Corresponding author.
E-mail address: mwaters@ils-inc.com (M. Waters).

0041-008X/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.taap.2007.12.036

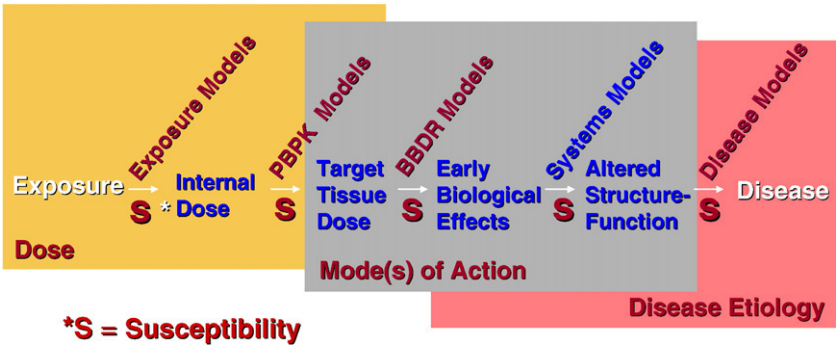

Fig. 1. The Exposure to Disease Paradigm-modeling dose, mode of action, and disease etiology (Waters and Fostel, 2004).

human disease susceptibility. There have been numerous proposals and some successes in applying toxicogenomics methods in the assessment and prediction of chemical toxicity (Steiner et al., 2004; Waters and Fostel, 2004; Ruepp et al., 2005). Within the past year, four groups of investigators (Nakayama et al., 2006; Nie et al., 2006; Tsujimura et al., 2006; Thomas et al., 2007) have suggested and appear to have demonstrated that it is also possible to use gene expression profiling and gene expression classification methods to accurately predict chemical carcinogenicity.

There are a number of paths forward to improve the utility of toxicogenomics in hazard and risk assessment and some of them will be identified in this manuscript. Proof-of-concept experiments and datasets on the effects of prototypic chemicals are needed to determine the diagnostic and predictive utility of "molecular signature" data and corresponding clinical and pathophysiological phenotype data. Toxicogenomic datasets and conventional toxicity data available within public databases such as those described here could be used in the determination of a chemical's mode of action by reference to a prototypic member of a mode of action class. In principle, the molecular signatures displayed by a particular chemical could be used together with data on corresponding pathophysiological phenotypes to classify a toxicogenomic mode of action.

Comprehensive data collection and integration, as well as iterative biological modeling, is required to realize the full potential of toxicogenomics as a component of a predictive systems toxicology (Waters and Fostel, 2004). The full sequence of events between initial exposure and final disease outcome is shown from left to right in Fig. 1. Following exposure, the body's "ADME" [absorption, distribution, metabolism, excretion] systems control local concentrations of a chemical stressor in various body compartments. The impact of genetics is felt in specific alleles encoding various transporters, xenobiotic metabolizing enzymes, etc. Mathematical models such as exposure models, physiologically based pharmacokinetic (PB/PK) and biologically-based dose–response (BBDR) models can be used to approximate these processes. PB/PK models are a set of differential equations structured to provide a time course of a chemical's mass-balance disposition (wherein all inputs, outputs, and changes in total mass of the chemical are accounted for) in pre-selected anatomical compartments. BBDR models are dose–response models based on underlying biological processes. Once the target tissue is exposed to a local stressor the cells respond by either adaptation or by undergoing a toxic response; this process can be modeled with systems toxicology approaches. Finally, the disease outcome itself can be mimicked by genetic or chemically-induced models of particular diseases, e.g., cancer. Both conventional toxicology and toxicogenomics data and information must be assembled and made publicly available as a first step.

The Workshop on The Power of Aggregated Toxicity Data addressed the requirement for distributed databases to support quantitative hazard and risk assessment. The authors have conceived, and/or constructed with federal support several databases that have been used in hazard identification and risk assessment. The EPA Gene-Tox Database (Waters and Auletta, 1981; Waters, 1994) developed for the EPA Office of Toxic Substances, and currently hosted by the National Library of Medicine is based on the first and only literature evaluation of its kind in the field of toxicology. This public resource represents the collaborative evaluation, by government, academic, and industry scientists, of all major short-term tests for the detection of mutagens and presumptive carcinogens. Completed in 1987 (Ray et al., 1987), the two-phased evaluation process considered more than 200 tests and resulted in more than 50 peer-reviewed Gene-Tox Program publications on test system performance, and a qualitative database on thousands of chemicals. Results for some short-term tests have been updated to 1995.

Subsequently, the graphic and quantitative EPA/IARC Genetic Activity Profile (GAP) Database was developed in collaboration with the International Agency for Research on Cancer (IARC). Whereas the Gene-Tox Database represents an evaluation of the qualitative performance of short-term tests and chemicals studied in these assays, the GAP database attempts to record quantitative genotoxicity test results (e.g., the lowest effective dose) for each given chemical. GAP has served IARC for many years in support of hazard classification of potential human carcinogens (Waters et al., 1988, 1991, 1999). GAP includes approximately 500 chemical agents in *IARC Monographs* evaluated by IARC Working Groups in Monograph Volumes 44–73 and Supplement 6, about 250 EPA priority chemicals, including pesticides, hazardous air pollutants, and Superfund toxicants. GAP represents the abstraction of approximately 9000 published primary articles with peer review by IARC Working Groups.

The Toxicological Activity Profile (TAP) prototype database was patterned after GAP and utilized acute, subchronic, and chronic data from the Office of Air Quality Planning and Standards. TAP demonstrated the flexibility of the GAP format for air toxics, water pollutants and other environmental agents. The GAP format was also applied to developmental toxicants (Kavlock et al., 1991) and was modified to represent quantitative results from the rodent carcinogen bioassay (Jackson et al., 1997).

Data and information from the databases described above have been used for many years to evaluate genetic damage and toxicity associated with exposure to chemical, biological, and physical agents, and to provide summaries of relevant results for carcinogenicity evaluations. Thus the genetic toxicity data have been invaluable for hazard identification in risk assessment and for identifying genotoxic modes of action. The capability to readily access high quality evaluated data published in multiple studies in different species and target cell types according to the specific types of genetic lesions also helps to elucidate putative carcinogenic mode(s) of action of agents involved in tumor induction and progression as will be shown below.

To improve the knowledgebase of structural changes in the DNA of cancer cells and the toxicological events leading to such changes, the