Computational Systems Pharmacology and Toxicology

Issues in Toxicology

Titles in the Series:

1: Hair in Toxicology: An Important Bio-Monitor
2: Male-mediated Developmental Toxicity
3: Cytochrome P450: Role in the Metabolism and Toxicity of Drugs and other Xenobiotics
4: Bile Acids: Toxicology and Bioactivity
5: The Comet Assay in Toxicology
6: Silver in Healthcare
7: *In Silico* Toxicology: Principles and Applications
8: Environmental Cardiology
9: Biomarkers and Human Biomonitoring, Volume 1: Ongoing Programs and Exposures
10: Biomarkers and Human Biomonitoring, Volume 2: Selected Biomarkers of Current Interest
11: Hormone-Disruptive Chemical Contaminants in Food
12: Mammalian Toxicology of Insecticides
13: The Cellular Response to the Genotoxic Insult: The Question of Threshold for Genotoxic Carcinogens
14: Toxicological Effects of Veterinary Medicinal Products in Humans: Volume 1
15: Toxicological Effects of Veterinary Medicinal Products in Humans: Volume 2
16: Aging and Vulnerability to Environmental Chemicals: Age-related Disorders and their Origins in Environmental Exposures
17: Chemical Toxicity Prediction: Category Formation and Read-Across
18: The Carcinogenicity of Metals: Human Risk Through Occupational and Environmental Exposure
19: Reducing, Refining and Replacing the Use of Animals in Toxicity Testing
20: Advances in Dermatological Sciences
21: Metabolic Profiling: Disease and Xenobiotics
22: Manganese in Health and Disease
23: Toxicology, Survival and Health Hazards of Combustion Products
24: Masked Mycotoxins in Food: Formation, Occurrence and Toxicological Relevance
25: Aerobiology: The Toxicology of Airborne Pathogens and Toxins

26: Chemical Warfare Toxicology, Volume 1: Fundamental Aspects
27: Chemical Warfare Toxicology, Volume 2: Management of Poisoning
28: Toxicogenomics in Predictive Carcinogenicity
29: Human Stem Cell Toxicology
30: The Comet Assay in Toxicology, 2nd edition
31: Computational Systems Pharmacology and Toxicology

How to obtain future titles on publication:
A standing order plan is available for this series. A standing order will bring delivery of each new volume immediately on publication.

For further information please contact:
Book Sales Department, Royal Society of Chemistry, Thomas Graham House, Science Park, Milton Road, Cambridge, CB4 0WF, UK
Telephone: +44 (0)1223 420066, Fax: +44 (0)1223 420247
Email: booksales@rsc.org
Visit our website at www.rsc.org/books

Computational Systems Pharmacology and Toxicology

Edited by

Dale E. Johnson
University of Michigan, USA
Email: daleej@umich.edu

Rudy J. Richardson
University of Michigan, USA
Email: rjrich@umich.edu

Issues in Toxicology No. 31

Print ISBN: 978-1-78262-332-8
PDF eISBN: 978-1-78262-373-1
EPUB eISBN: 978-1-78801-120-4
ISSN: 1757-7179

A catalogue record for this book is available from the British Library

The Royal Society of Chemistry is a charity, registered in England and Wales, Number 207890, and a company incorporated in England by Royal Charter (Registered No. RC000524), registered office: Burlington House, Piccadilly, London W1J 0BA, UK, Telephone: +44 (0) 207 4378 6556.

For further information see our web site at www.rsc.org

Printed in the United Kingdom by CPI Group (UK) Ltd, Croydon, CR0 4YY, UK

Preface

Our motivation for bringing about a book on systems computational pharmacology and toxicology was a natural development from teaching courses on these subjects, first at the University of California in Berkeley and later at the University of Michigan in Ann Arbor. Our courses and this book address a critical need to modernize pharmacology and toxicology—to transform these fields from descriptive disciplines to predictive sciences. This transformation is necessary, because classic descriptive approaches are far too inefficient and expensive to assess the medical efficacy or toxicity of the many thousands of synthetic chemicals or natural products to which humans and other species are or will be exposed.

Not long ago, the approaches set forth in this book were either not possible or performed by specialists using such tools as quantum mechanics and mainframe computers. Now, because of rapid advances in technology, software, and theory, coupled with the public availability of large chemical and biomedical data sets through the internet, it is possible for non-specialist bench scientists to undertake sophisticated molecular modeling, bioinformatics, cheminformatics, and systems biology procedures on desktop computers as well as mobile devices, including to some extent electronic tablets and smart phones. Thus, powerful computational tools have become highly accessible, but knowing how and when to use the right tools in the right way can be a daunting task. This book seeks to make the job easier to understand and implement.

Recognizing that we now have the capability to understand pharmacological and toxicological effects at multiple biological levels, our book highlights the process of integrating the elements of complex phenomena into a systems approach. Thus, whereas inverse docking and pharmacophore mapping can identify molecular targets of candidate drugs or toxicants, intelligent mining of databases can identify networks of genes and proteins involved

Issues in Toxicology No. 31
Computational Systems Pharmacology and Toxicology
Edited by Dale E. Johnson and Rudy J. Richardson

Published by the Royal Society of Chemistry, www.rsc.org

in the system-wide biological responses to chemicals. A pharmacological or toxicological effect may begin with atomic-level binding, but ultimately the intact organism responds in a holistic manner. It is necessary to continually readjust our focus by many orders of magnitude to encompass the spectrum from molecular orbitals to human populations.

The tools and models discussed in the book hold tremendous promise for advancing applied and basic science, streamlining drug efficacy and safety testing, and increasing the efficiency and effectiveness of risk assessment for environmental chemicals. The content of chapters is designed to provide readers with an understanding of the basic principles and current methods of computational pharmacology and toxicology. These principles and approaches are discussed in several chapters in order to show how to connect chemicals with diseases and associated genes, and how to create pharmacology/toxicology connectivity maps or networks.

Vital to these expositions of principles and methods are illustrations of modeling and/or predicting potential pharmacological or toxicological effects from multiple properties. These characteristics include chemical structure, inference from similar compounds, *in silico* target identification, exposure, bioaccumulation, environmental persistence, biomarkers, and networks of biological pathways affected by a chemical.

Systems toxicology approaches used in the safer design of chemicals and identification of safer alternatives, which are major parts of global green chemistry initiatives, are also discussed, along with the concept of the adverse outcome pathway and modeling approaches for hazard identification and risk assessments for large numbers of environmental chemicals for which supporting data are sparse.

The book also expands the conventional boundaries of research and development of pharmaceutical agents. Thus, traditional Chinese medicines that include recipes containing several pharmacologically active phytochemicals are becoming role models of polypharmacy research.

The final chapter describes an inquiry-based computational toxicology course. Students work in small cooperative groups and are given tools, data, and basic concepts to solve toxicity-related environmental, public health, and/or disease-oriented problems in novel ways. Several case studies serve both to educate the reader and to provide material for teaching.

As co-editors, we are each involved in research and education on the topics covered in the book. We have authored or co-authored several of the chapters ourselves, and the other chapters have been written by experts recruited from around the world.

Dale E. Johnson and Rudy J. Richardson

Contents

Issues in Toxicology No. 31
Computational Systems Pharmacology and Toxicology
Edited by Dale E. Johnson and Rudy J. Richardson

Published by the Royal Society of Chemistry, www.rsc.org

CHAPTER 1

Systems Biology Approaches in Pharmacology and Toxicology

DALE E. JOHNSON*[a,b]

[a]University of Michigan, School of Public Health, Department of Environmental Health Sciences, Ann Arbor, MI 48109-2029, USA; [b]University of California, Berkeley, Department of Nutritional Sciences and Toxicology, Morgan Hall, Berkeley, CA 94720-3104, USA
*E-mail: daleej@umich.edu

1.1 Introduction

The science and practical field of toxicology has been changing dramatically over the last 15–20 years, transitioning into a more systems biology and network-based approach.[1–4] Several factors have been involved, including the developing genomics era where the understanding of genetic changes has enhanced the ability to understand diseases and chemically-induced toxicities at the molecular level. The genomics era has also ushered in "omics" technologies and approaches such as transcriptomics, metabolomics, proteomics, and epigenomics, which have changed the way we view mechanisms of toxicity and the perturbation of biological systems that lead to adverse outcomes.[5] These advances have been coupled with the public availability of large datasets of information and new modeling approaches that have enhanced the ability to understand toxicological events and effects at multiple biological levels.[6] Since our scientific approaches, inquiries, and visions aimed at understanding toxicological events and outcomes have

Issues in Toxicology No. 31
Computational Systems Pharmacology and Toxicology
Edited by Dale E. Johnson and Rudy J. Richardson

Published by the Royal Society of Chemistry, www.rsc.org

been broadened tremendously, this reinforces our need for new and better ways to assess toxicity and risk. The large numbers of uncharacterized chemicals already present in the environment and new chemicals that continue to enter it has required hazard and risk assessments to be made with very few data. These factors have had a major influence on the need to accelerate new approaches and move away from an overdependence on *in vivo* animal testing and make better use of computational, molecular, and *in vitro* tools.[6,7] The identification of the majority of toxic events in *in vivo* animal toxicology studies rely on high-dose exposure to the animals and default linear extrapolation procedures,[8] with the incorporation of newer technologies absent in the vast majority of animal studies. This has been considered a shortcoming in risk assessment and several weaknesses in this process include the comparative shape of the dose–response relationship after relevant levels of human exposure, whether biological and/or toxicological thresholds do in fact exist and for what toxicological endpoints, and potential population variability in response.[5]

1.2 Systems Toxicology

Accordingly, research in toxicology has moved into a new systems-oriented phase called systems toxicology, which involves the study of complex molecular response networks initiated by exposure (both intentional and unintentional) to chemical substances. At the heart of systems toxicology approaches are the development and usage of quantitative mechanistic models that create a predictive toxicology aspect relevant to all toxicology fields, including drug research and development and environmental research. The overall approach involves the integration of classical toxicology with the quantitative analysis of large networks of chemically-induced molecular and functional changes, which occur across multiple levels of biological organization.[5] Examples of key influential events in this transition since the year 2000 include the release of human genome sequencing data including specific signal transduction domains, the development and issuance of the report *Toxicity Testing in the Twenty-first Century* by the National Research Council (NRC),[9] which has influenced all sectors of the toxicology field, and the development and publication of the adverse outcome pathway (AOP) approach,[6,10,11] which has highlighted the realities that exist as the science moves away from an overdependence on *in vivo* testing and makes greater use of computational, molecular, and focused *in vitro* tools. Additional drivers of change include the European Union (EU) report from the Scientific Committee on Health and Environmental Risks, the EU's Registration, Evaluation, Authorisation and Restriction of Chemical Substances (REACH) program, and the International Programme on Chemical Safety (IPCS).[7,12] The paradigm shift can also be seen in the drug research and development sector, but rather than focusing on drugs during late stages of development or on marketed drugs, the systems-related efforts are positioned at the front end of research, both on safer chemical design and extensive target research.

While the drug industry is required to conduct animal toxicology studies by regulatory agencies and international guidelines, the major effort underway is to determine chemical liabilities early in the drug discovery pipeline, both to reduce the time and cost of failures later in the process, but also to avoid costly failures once a drug reaches the market.[5] Currently, there is an International Consortium for Innovation and Quality in Pharmaceutical Development (IQ), where several pharmaceutical and biotechnology companies have created a Nonclinical to Clinical Translational Database (WG1) to allow analysis of the reliability and potential limitations of nonclinical data in predicting clinical outcomes, including the evaluation of conventional biomarkers of toxicity.[13] Current screening approaches applied to the front end of drug research are described below.

1.3 Chemical Toxicities

1.3.1 Single-Target Toxicity Concepts

The science and practice of toxicology over the past several decades have consistently used classic toxicological approaches, such as *in vivo* and *in vitro* toxicology studies, combined with predictive toxicological methodologies. The desired endpoints of the *in vivo* animal research efforts have been the determination of a toxic dose where a chemical could be shown to induce pathologic effects after a specified duration of treatment or exposure. Where appropriate, these studies have included the estimate of the lowest observed adverse effect level, the no observed adverse effect level, and the maximally tolerated dose (MTD).[5,14] These adverse effect level estimates are traditionally used in drug research and development to predict the first dose in humans and to predict margins of safety estimates based on delivered dose and/or internal exposure from pharmacokinetic/pharmacodynamic (PK/PD) modeling with extrapolations into clinical trial subjects. By regulatory requirements, all potential drugs undergoing research and development will undergo both *in vitro* and *in vivo* studies, and, if the compound reaches the clinical trial stage successfully, data from human exposure to judge the adequacy of nonclinical data in predicting clinical outcomes. Uncertainties in these estimates include the definition of adverse, which is specific for each organ system in each study and typically determined by the study pathologist; the accuracy of cross-species extrapolations (particularly rodent-to-human); and the true definition of risk–benefit for each individual drug. However, the generation of classical toxicology data does not assure the accurate prediction of potential human toxicity. Sundqvist and colleagues[15] have reported on a human dose prediction process, supplemented by case studies, to integrate uncertainties into simplified plots for quantification. Drug safety is recognized as one of the primary causes of attrition during the clinical phases of development; however, in numerous instances the actual determination of serious adverse effects only occurs after the drug reaches the market.

In the United States, ~2 million patients are affected with drug-mediated adverse effects per year, of which ~5% are fatal.[16] This places drug toxicity as one of the top five causes of death in the United States, and the costs to the health care system worldwide are estimated at US$40–50 billion per year.[16] In drug development there are always risk–benefit considerations, which will weigh any potential toxicity against the benefit expected to be gained by a patient taking the drug. An example of the uncertainty of these estimates can be seen in the methods used for carcinogenicity testing and evaluation for drug approval. The design of these studies rely on high-dose exposure to animals and default linear extrapolation procedures, while little consideration is given to many of the new advances in the toxicological sciences.[17] Carcinogenicity studies are typically 2-year studies in rodents conducted with three dosage groups (low, mid, and high dose) and one or two concurrent control groups. Dose levels are established from previous studies, such as 13-week toxicity studies, where a MTD has been estimated. Each group in the carcinogenicity study has 60–70 animals of each sex, and the analysis of whether there is a potential carcinogenicity concern is based on an analysis of each tumor in each tissue or organ system individually by gender; certain tumors are combined *via* standardized procedures for statistical analysis. The analysis uses the historical database from the laboratory where the studies are conducted to determine whether each tumor is considered common or rare, using the background incidence of 1% as the standard. Common tumors are those with a background incidence of 1% or over and rare tumors are those with a background incidence below 1%. In the statistical analysis, *p*-values for rare and common tumors are evaluated for pair-wise significance at 0.05 (for rare) and 0.01 (for common). The rare *vs.* common tumor classification is an arbitrary tumor threshold and adjustments to the specific classifications by individual tumor, which can occur from laboratory to laboratory and *via* analyses of different control groups, can have consequences in the overall tumor evaluation outcome.[8] Applying a "weight of evidence" approach into the evaluation procedures, particularly during regulatory review, attempts to alleviate some of the uncertainties; however, after more than 50 years of on-going experience, these studies still fail to bring the 21st century mindset to carcinogenicity testing. The classic toxicological process for drug development assumes that a chemical interacts with a higher affinity to a single macromolecule (the toxicological target), and therefore a single biological pathway may be perturbed at the initial target modulation. This would be followed by downstream activation of secondary and possibly tertiary pathways that result in the tissue or organ effect as indicated by key biomarkers.[2] In this concept, the magnitude of toxicological effects are related to the concentration of altered molecular targets (at the site of interest), which in turn is related to the concentration of the active form of the chemical (parent compound or metabolite) at the site where the molecular targets are located. Also included in this concept is the unique susceptibility of the organism exposed to the compound.

1.3.2 Toxicological Profiling for Potential Adverse Reactions

Predictive toxicology efforts in drug research and development involve the use of multiple sources of legacy data including data generated by chemical and pharmaceutical companies and data submitted to regulatory agencies. These efforts have led to the "data warehouse" model which includes data generated through high throughput and targeted screening, and *in vitro* and *in vivo* toxicology studies on thousands of compounds and structural analogues. In a majority of cases these data also include findings from clinical trials where an experimental drug was tested on humans.

The information is applied in a "backward" fashion to predict potential findings where data do not yet exist or where decisions are being made on new potential drug candidates. Bowes and colleagues[18] have described a pharmacological profiling effort by four large pharmaceutical companies: AstraZeneca, GlaxoSmithKline, Novartis, and Pfizer. The companies suggest that ~75% of adverse drug reactions can be predicted by studying pharmacological profiles of candidate drugs. The pharmacological screening identifies primary effects related to the intended action of the candidate drug, whereas identification of secondary effects due to interactions with targets other than the primary (intended) target could be related to off-target adverse events. The groups have identified 44 screening targets including 24 G-protein coupled receptors, eight ion channels, six intracellular enzymes, three neurotransmitter transporters, two nuclear receptors, and one kinase. These types of screening data are used in the data warehouse model, typically configured in a proprietary fashion within each company. Other collaborative efforts have been developed and data from these sources would also be incorporated.

Blomme and Will[19] have reviewed the current and past efforts by the pharmaceutical industry to optimize safety into molecules at the earliest stage of drug research. They conclude that new and emerging technologies in the past two decades have had limited impact on nonclinical attrition rates associated with safety issues. In addition, they point out that front-loading series of toxicology assays to "kill early, kill often" have been challenged due to high false-positive rates and an overall low positive predictive value. The primary issue cited is the lack of information on an efficacious exposures (PK/PD) and the fact that the assays are more likely to represent hazard identification and not risk assessment. Therefore, it is suggested that these data be used as alerts rather than discontinuance criteria. In a more systems toxicology approach, a large effort is now being directed towards understanding the extent of pharmacological modulation of both precedented and unprecedented targets in relation to potential safety liabilities and developing technologies to determine achievable therapeutic windows. Blomme and Will[19] discuss efforts at AbbVie and Pfizer where target safety assessments are explored. The assessments include the biology of the target, tissue expression maps, messenger RNA and proteins, human genetic data, phenotypes from genetically engineered animal models, historical data from on-going

and past clinical trials targeting similar targets and associated pathways, extensive datamining *via* biomedical databases, and *in silico* simulation of the various consequences of target modulation. The majority of systems research in drug safety use specific toxicities as the starting point. Ivanov and colleagues[20] discuss the use of specific methods to counteract ventricular tachyarrhythmia (VT). While the *in vitro* HERG potassium channel assay is used universally as a predictor of VT, this is only one mechanism of action and other targets must also be identified and explored. These researchers have used the following approach: (1) creation of VT positive and negative compound libraries; (2) *in silico* prediction of extensive drug–target interaction profiles of chemical libraries identifying potential VT-related targets; (3) gene ontology and pathway enrichment on these potential VT targets to elucidate potential biological processes; (4) creation of a cardiomyocyte regulatory network based on general and heart-specific signaling and regulatory pathways; and (5) simulation of the changes in the regulatory network caused by the inhibition at each node in order to define potential VT-related targets. These are the type of studies that lead to more refined *in vitro* and *in silico* assessments of potential drug adverse effects at the early stage of drug research.

Verbist and colleagues[21] have outlined another type of systems toxicology proposal at Janssen involving QSTAR (quantitative structure-transcription-assay relationships) by integrating high-throughput gene expression profiling data; chemical information, particularly detailed analogue analysis; and bioassay data. Using several compounds from a single chemical scaffold targeting PDE10A, a target of pharmacological interest at Janssen, changes in tubulin gene expression were identified in a subset of compounds. Therefore a screening process was developed involving multiple cell lines, gene expression profiling, *in vitro* micronucleus assays, and high-content imaging to show microtubule aggregates as compared to other phenotypes. Besides the chemical series of interest, known positive and negative compounds were included in the process. This study presents a valuable proof-of-concept of how to link and potentially improve the risk assessment in early drug discovery using several technologies in a drug research systems toxicology approach.

1.3.3 Toxicological Concepts for Safer Chemical Design

Voutchkova and colleagues[22] have outlined an extensive framework for safer chemical design using multiple data and modeling resources. These types of data generation and modeling approaches are the basis of the process for a green chemistry model for specific series of chemicals used or proposed for use as reagents, solvents, or chemical intermediates in chemical synthesis. The simplified scheme involves a model building process, where chemical structures of interest are evaluated for chemical motifs (structural alerts) known to be associated with human health or environmental hazards, chemicals are clustered into hazard categories and specific high- or higher-throughput

and targeted assays are identified for each hazard category. Analogue series directly applicable to the chemistry under evaluation are prepared or obtained and screened in the relevant assays. With homologous or structurally similar series, local chemical-toxicity models can be developed, validated and incorporated into the initial computational screening process. This general method can be applied to any specific hazard, any series of chemicals, and any assay methodology. Examples of hazard categories include carcinogenicity; reproductive and developmental; mutagenicity; neurotoxicity; endocrine disruption; cardiovascular; dermatotoxicity; digestive system toxicity; hematotoxicity; hepatotoxicity; immunotoxicity; muscular toxicity; nephrotoxicity; ocular toxicity; ototoxicity; respiratory toxicity; persistence in the environment; bioaccumulative in the environment; toxic to water organisms; water contaminant; and air pollutant. Structural motif alert and expert predictions can be achieved using OpenTox,[23] an open access system used to predict potential hazards from chemical structures and known chemical motifs associated with human health and environmental endpoints, and Derek from Lhasa,[24] a rule-based expert system that de-convolutes a chemical structure into sub-structural fragments and addresses potential toxicity consistent with the above hazard categories. The software is also used to create specific local expert predictions from screening data. Meteor (Lhasa) predicts potential metabolites and the metabolite structures can be used in Derek predictions. This type of inquiry is highly useful for establishing basic information for rank-ordering compounds, as in early candidate selection, and in the process of safer chemical synthesis. Multiple screening approaches have been used for evaluation of chemical toxicity using high-throughput technology and multiple assays. These include the United States Environmental Protection Agency (EPA) ToxCast program[25] where over 2000 chemicals have been evaluated in over 700 high-throughput assays. This is a section of the Tox21 testing program, a collaboration among EPA, the National Institutes of Health (NIH), including the National Center for Advancing Translational Sciences at the National Toxicology Program at the National Institute of Environmental Health Sciences, and the United States Food and Drug Administration. The Tox21 program involves high-throughput screening of more than 10 000 environmental chemicals and approved drugs using more than 100 assays. All data are publicly available, as discussed later. Wink and colleagues[26] discuss a quantitative high-content imaging *in vitro* process to elucidate chemical interactions with cellular adaptive stress response pathways to gain a better insight into chemical toxicities at a phenotypic cellular level. The key to their reported technology is a panel of reporter cell lines to monitor multiple key nodes of the adaptive stress response pathways. Examples include cellular redox homeostasis, unfolded protein response, endoplasmic reticulum damage, inflammatory signaling, and DNA damage response. These assays hold the potential to be incorporated into multiple large-scale screens to evaluate health-related chemically-induced biological phenomena in drug research as well as hazard identification.

1.3.4 Biomarkers

Biomarkers are typically used to define the onset, continuation, and either positive or negative characteristics of the induced biological effects of the drug (chemical) under research. Biomarkers have been classified as biomarkers of exposure, susceptibility, and outcome. The definition of biomarker as used in drug discovery and development is a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic response(s) to a therapeutic intervention.[27] In pharmacological studies, where a relevant therapeutic target is identified and pursued, biomarkers are developed that correlate with the proof of concept for the drug candidate. Biomarkers are developed to show (1) that a desired modulation of the target occurs as anticipated by the chemical therapeutic; (2) that the chemical-induced target modulation produces a desired biological effect; (3) that the induced biological effect alters the disease under study; and (4) that there may be increased susceptibility to the therapeutic candidate by certain individuals, such as those based on pharmacogenetic predispositions. In toxicology studies, biomarkers are objectively measured and evaluated as indicators of (1) normal biological processes; (2) pathogenic processes; (3) pharmacologic response(s) to a therapeutic intervention, which in some cases could mean excessive or nonspecific pharmacologic activity; and (4) exposure–response relationships. Pharmacogenetic markers are also studied from a toxicological standpoint, particularly in relation to drug metabolism. In environmental research and risk assessment, biomarkers are frequently referred to as indicators of human or environmental hazards. Discovering and implementing new biomarkers for toxicity caused by exposure to a chemical from a therapeutic intervention or in some cases through unintentional exposure continues to be pursued through the use of animal models to predict potential human effects, from human studies (clinical or epidemiological) or from biobanked human tissue samples, or the combination of these approaches.[27] In addition, several omics technologies such as transcriptomics, metabolomics, and proteomics have added an important aspect to biomarker research.[12] More recently, epigenomics, which is the study of changes in gene activity not attributed to DNA sequence alterations, has been shown to have increased importance in disease causality research.[12] These technologies and data produced, along with large datasets of high-throughput screening data, essentially changed the process of defining biomarkers. The process of discovering or inferring biomarkers through computational means involves the identification or prediction of the molecular target(s) of the chemical, which in many cases can be secondary or undesirable targets and the association of these targets with perturbed biological pathways. The integration of these approaches with the quantitative analysis of chemically induced molecular and functional changes has brought to fruition the goals originally outlined in the 2007 NRC report.[9]

1.4 Environmental Toxicology

The concepts of environmental chemical toxicity are different to the concepts of drug toxicity, as with environmental exposures there are only risk considerations, with virtually no benefit associated with chemical exposure. Currently there are more than 80 000 chemicals on the market and/or reach the environment, and ~2000 new chemicals are introduced each year. Unlike drugs, for the majority of these compounds, there is limited or inadequate toxicological information with which to make rational evaluations of risk.[2] Where information exists or is associated from similar structures of analogous compounds, a risk assessor may arrive at various hazard reference values, such as a derived no-effect level, to estimate an acceptable level of protection of human or wildlife health and the environment.[5] Depending on the context and urgency of the risk assessment, which could include the classification of a chemical in terms of hazard severity and risk management assessment, several assumptions must be made which could cause the level of uncertainty to increase.

1.4.1 Adverse Outcome Pathway

The concept of the AOP was a necessary enhancement to the *Toxicity Testing in the Twenty-first Century* report, made to more adequately support ecological risk assessment.[6] After the first publication in 2010, and subsequent publications by scientists at the EPA,[10,11] the AOP concept and developing case studies have become a primary force in the progression of the computational systems toxicology approach in environmental risk assessment. Unlike work in drug discovery and development, which always has a confidential business information component and therefore results are not fully publically available, AOPs are developed in a fully open-access mode and are supported by publicly available databases updated by EPA scientists. The AOP concept highlights existing knowledge that links the direct molecular initiating event, of which in theory is the interaction or modulation of a molecular biomolecule or target with a xenobiotic, and an adverse outcome at a biological organization that spans multiple levels of biological organization including the following general examples from Ankley and colleagues.[6] These events are outlined below.

(1) Macro-molecular interactions, such as receptor–ligand interactions including agonism and antagonism, DNA binding, and protein oxidation.
(2) Cellular responses, such as gene activation, protein production, alterations in signaling, and protein depletion.
(3) Organ responses, such as disrupted homeostasis, alterations in tissue development and/or function, and altered physiology.
(4) Organism responses, such as mortality, impaired reproductive and developmental function, and development of diseases, such as cancer.

(5) Population responses, such as alterations in the structure of a population and potential extinction of species within regional and global environments.

In defining the key aspects of an AOP, macro-molecular interactions (1) are considered the initiating event, which is called "anchor 1", and the organ and population responses (4 and 5) are considered adverse outcomes at the organism or population level, collectively called "anchor 2". The aspects of connecting the initiating events to outcomes can take various forms, depending on the chemical itself and the amount of information available, including *in vivo*, *in vitro*, and computational sources. These various linkages also help to define key assays and technologies to enhance information collection and usage for each individual AOP. Ankley and colleagues[6] provide five case studies that illustrate these points. Events occurring in the information flow between the molecular initiating event and adverse outcome are called intermediate events.[28] When an intermediate event represents a biological event that is necessary for an adverse outcome to occur and is quantitatively measurable, it is considered to be a key event. In a systems toxicology approach, key event relationships between adjacent molecular initiating events, key events and adverse outcomes help define alternative approaches to assess environmental hazards. In a signaling pathway, this could be upstream or downstream events that help define more suitable assays or test systems and provide a faster quantitative evaluation of potential adverse outcomes. One of the challenges in the AOP process is to define or estimate the exposure of the xenobiotic in the relevant species under consideration for risk assessment.

1.4.2 Expanding Exposure Concepts

As discussed earlier, measurement of exposure, toxicokinetics, systemically and importantly, at the critical site of action (anchor 1) is an essential piece of information. Unlike pharmaceutical compounds, it is highly unlikely that there would ever be controlled human toxicokinetics data for industrial and environmental chemicals.[29] Extrapolating toxicokinetic models from *in vitro* data has been used with pharmaceutical compounds, and this is termed *in vitro* to *in vivo* extrapolation methodology. These models have been used with some success for environmental and industrial chemicals using high-throughput toxicokinetics (HTTK) models. Several examples of HTTK methodology have been published, including lecture series by scientists from the United States EPA, the National Center for Computational Toxicology, and the Hamner Institutes for Health Sciences.[29,30] The primary methodology is termed "reverse dosimetry", which uses concentrations that produce bioactivity in *in vitro* assays to estimate doses (mg kg^{-1} per day) sufficient to produce steady-state plasma concentrations (C_{ss}) in μM. These approaches assume 100% bioavailability and a linear relationship between C_{ss} and dose. Another approach, called probabilistic reverse dosimetry approaches for estimating exposure distributions (PROcEED)[30] uses

biomarkers of exposure, such as blood and urine levels from biomonitoring studies, to model the most likely exposure concentrations—or intake doses or dose levels—experienced by the study participants that would result in the concentrations measured. These modeling procedures are considered a work in progress (as of early 2016); however, they represent a critical piece of the puzzle in chemical-associated environmental risk assessment. Certain drivers of the *in vivo* toxicokinetics process may be inaccurately estimated by *in vitro* assays such as extra-hepatic metabolism (particularly when using *in vitro* hepatic cell lines), secretion in bile and enterohepatic recirculation, absorption and bioavailability, and active transport in several tissues including renal and hepatic. In addition, a process to include metabolites into the high-throughput screening process for all chemicals will be a necessary part of the functional HTTK process.[29]

1.4.3 Exposome

The exposome is currently defined as the totality of all human environmental exposures (exogenous and endogenous) from conception to death.[31] The National Institute of Environmental Health Sciences[32] has developed a broad definition of environmental exposures, which includes chemical exposures, diet, physical activity, stress, pre-existing disease, and the use of substances that could lead to addictive consequences. The concepts of measuring all exposure events over time is certainly difficult, particularly considering the dynamic aspects of exposures leading to adverse outcomes; however, much effort is being given to establishing biomarkers related to the exposome on both a population and individual basis. These biomarkers are being evaluated for refining exposure assessments in risk assessments; providing correlations leading to exposure–disease associations, particularly in data from epidemiological studies; the potential identification of susceptible individuals or groups; using human data rather than extrapolations from animal data; and potentially identifying interventions in reducing certain exposures and/or treating the adverse outcome.[32] One of the major efforts to define and understand environmental exposures is the published biomonitoring studies, National Health and Nutrition Examination Survey, from the Centers for Disease Control and Prevention National Center for Health Statistics. The *Fourth National Report on Human Exposure to Environmental Chemicals* with updated tables[33] provides national (USA) biomonitoring data (serum and urinary levels) on 265 chemicals from subsets of the population. The website contains details of data sources and data analysis, interpretation of report data, and chemical and toxicological information. Bell and Edwards[34] have described a workflow, a frequent itemset mining approach, to identify relationships between chemicals and health biomarkers and disease. Currently, the most complete information source for toxicology information and exposure identification, including the exposome, is the Toxin and Toxin-Target Database[31] (T3DB; www.t3db.ca). The details of T3DB are discussed in Chapter 2.

1.5 Systems and Network Pharmacology

Systems pharmacology is defined as a translational science that aims to examine all the biological activities in the body related to internal exposure of a drug or drug candidate and the resultant drug responses and pharmacological activities.[35] Systems pharmacology uses both experimental approaches and computational analyses to examine and understand drug action across multiple levels including molecular, cellular, tissue, and whole organisms[36] with consideration to the presence of several interacting pathways.[37] The field has grown and developed rapidly because of the emergence of omics technologies and network analysis capabilities, and the increased number of computer scientists, engineers, and mathematicians involved in addressing and solving complex biological problems.[35] In an NIH white paper by the Quantitative Systems Pharmacology (QSP) workshop group in 2011, QSP was defined as providing an integrated approach to determining and understanding mechanisms of action of drugs and drug candidates in preclinical models (*in vitro* and *in vivo*) and in patients eventually receiving the drugs.[38] The stated goals were to create a knowledge base to facilitate the change of complex cellular networks in pre-determined ways with mono and/or combination therapies; maximize therapeutic benefit by altering the pathophysiology of the disease being treated; and minimize toxicity.[38] Given that the mammalian signaling and regulatory pathways are complex, drug–target interactions can potentially lead to adverse effects due to the propagation of signal flow to distal effectors (off-targets) in multiple cells and tissues.[39] However, using complex pharmacological and toxicological network analyses, both positive and negative effects can be predicted. Zhao and Iyengar[39] have identified key questions that highlight the importance of identifying and pursuing a systems pharmacology approach in drug research as a starting point: (1) what are characteristics of specific diseases where drugs modulating a single target may not provide therapeutic efficacy; (2) how do adverse events arise from intra- and intercellular networking; (3) how does the genomic status of an individual relate to potential drug efficacy particularly when poly-pharmacy (combination) is anticipated; (4) how do combinations of targets and/or signaling nodes in complex diseases predict efficacious outcomes with drug combinations; and (5) can detailed usage of the interactome and genetic status of an individual predict therapeutic efficacy or toxicity? Practically, systems pharmacology allows the application of model-based thinking during target selection and target validation before a lead compound is selected for development.[40] QSP models can incorporate details of single and multiple drug plasma concentrations, systems biology models, pertinent regulatory networks and motifs of upstream and downstream loops including feedback and feedforward processes, and individual genomic and epigenetic characteristics important for individualized patient therapies.[41] Visser and colleagues[42] describe the use of QSP models and the creation of a flexible tool kit at Merck, which has enhanced key drug discovery and development decisions. The tool kit includes PK/PD models, disease

models, comparator models, model-based meta-analysis approaches, clinical trial design simulations, criteria for quantitative decision-making, and overall performance metrics. As an example, these approaches have been used to quantify anticancer drug synergy in resistant cells and predicting effective drug combinations. Models have also been effective in predicting and understanding positive off-target activities that could require early risk–benefit considerations. These activities include endocrine disruptors, peroxisome proliferator-activated-receptor-agonists, 5-HT_{2B} serotonin receptor agonists, and ligand-gated ion channel protein agonists.[43] An example of a tool for simulation and evaluation of QSP models, is the MatVPC, which incorporates visual predictive checks as a diagnostic to evaluate the structural and stochastic parts of a QSP model.[44] Biliouris and colleagues[44] illustrate the use with three models: (1) a three-compartment pharmacokinetics model with oral and intravenous bolus dosing; (2) a two-compartment pharmacokinetics model with multidose intravenous infusion; and (3) a pharmacodynamics model describing the time-course of body weight. Zhang and colleagues[41] describe a Sobol sensitivity analysis that determines how much of the variability in QSP models relates to each input parameter including the interactions of multiple parameters as they relate to the overall model output variability. This is a highly important aspect of QSP model building, refinement, and use, as it identifies the important and influential parameters that drive model output and, therefore, the inherent uncertainty of model predictions.

1.5.1 Secondary Pharmacology and Off-Target Effects

Secondary pharmacology has been described as off-target pharmacology where a drug interacts with other targets as well as the intended target, and multi-target drug research where drugs can interact effectively with multiple targets increasing the therapeutic efficacy in certain diseases.[36,45–49] These effects can provide both beneficial and adverse outcomes, and in some cases these drug qualities define several adverse effects seen with drugs in development and those marketed. Liu and colleagues[49] proposed a drug surveillance network for adverse drug reaction prediction through the integration of chemical compound signatures; biological targets including proteins, transporters, and enzymes, along with pathways; and phenotypic properties. Wang and colleagues[45] report on a protein pharmacology interaction network database, PhIN, where users can generate interacting target networks within and across human biological pathways by defining shared chemical compounds or scaffolds using a defined activity cutoff. The database also defines interactions between human–virus and virus–virus pathways. The database contains ~1 350 000 compounds; ~9400 targets with more than 12 400 000 activity measurements (as of March 2015). This type of database provides information and evidence-based predictions of chemical structures that interact with multiple targets, which would be useful in multi-target drug design and side effect predictions.

1.5.2 Prediction of Potential Adverse Effects

A predictive pharmaco-safety network has been proposed by Cami and colleagues,[50] where known drug–safety relationship networks are combined with adverse event information and detailed biological network information on several drugs as a means to predict likely but prospectively unknown adverse events. In this approach more directed surveillance programs can be instituted for drugs under development and those marketed. A multiple evidence fusion method for both approved and novel molecules was developed by Cao and colleagues[51]. In this approach the authors assumed that drug behavior at different biological levels would provide predictive information on adverse effects, and that semantic relationships between adverse reactions would aid in predicting new or unknown adverse reactions for certain drugs. They also found that drug–adverse-effect networks would allow the inference of unknown associations. These evaluations used similarity measures with drug and adverse event pairs. The authors concluded that these methods are inherently beneficial especially in drug discovery during target selection, drug repositioning, and multi-target inquiry and development. In addition, the methods provide a better focus for large-scale clinical trials, and more focused post-marketing drug surveillance.[50–52]

1.6 Conclusions

Systems biology approaches as applied to the fields of toxicology and pharmacology have increased our abilities to both visualize and understand complex chemical–biological interactions at the molecular, organ, susceptible individual, and species levels. Applying quantitative mechanistic models into a network-based analysis has not only improved our knowledge base on both chemically-induced pharmacological and toxicological effects, but also has allowed new approaches to emerge that rely less on *in vivo* animal testing. The ever growing abundance of databases and tools have caused the practitioners of toxicology, in particular, to step forward out of the proverbial animal toxicity box and approach solutions to problems in new ways. This has improved the understanding of adverse events, hazards, and risk assessment in all related fields: research and development of therapeutics; environmental, workplace, and household chemical exposures; and the design of safer chemicals in green chemistry endeavours.

References

1. N. Plant, An Introduction to Systems Toxicology, *Toxicol. Res.*, 2015, **4**, 9–22.
2. T. Hartung, E. van Vliet, J. Jaworska, L. Bonilla and N. Skinner, Food for thought...Systems Toxicology, *ALTEX*, 2012, **29**, 119.
3. J. Bai and D. Abernethy, Systems Pharmacology to Predict Drug Toxicity: Integration across Levels of Biological Organization, *Annu. Rev. Pharmacol. Toxicol.*, 2013, **53**, 451–473.

4. K. Kongsbak, N. Hadrup, K. Audouze and A. Vinggaard, Applicability of Computational Systems Biology in Toxicology, *Basic Clin. Pharmacol. Toxicol.*, 2014, **115**, 45–49.
5. S. Sturla, A. Boobis, R. FitzGerald, J. Hoeng, R. Kavlock, K. Schirmer, M. Whelan, M. Wilks and M. Peitsch, Systems Toxicology: From Basic Research to Risk Assessment, *Chem. Res. Toxicol.*, 2014, **27**, 314–329.
6. G. Ankley, R. Bennett, R. Erickson, D. Hoff, M. Hornung, R. Johnson, D. Mount, J. Nichols, C. Russom, P. Schmieder, J. Serano, J. Tietge and D. Villeneuve, Adverse Outcome Pathways: A Conceptual Framework to Support Ecotoxicology Research and Risk Assessment, *Environ. Toxicol. Chem.*, 2010, **29**(3), 730–741.
7. C. Willett, J. Rae, K. Goyak, B. Landesmann, G. Minsavage and C. Westmoreland, Pathway-Based Toxicity: History, Current Approaches and Liver Fibrosis and Steatosis as Prototypes, *ALTEX*, 2014, **31**, 407–421.
8. D. Johnson, Estimating human cancer risk from rodent carcinogenicity studies: the changing paradigm for pharmaceuticals, *J. Drug Metab. Toxicol.*, 2012, **3**, e114, DOI: 10.4172/2157-7609.1000e114.
9. *Tox21*, NRC, 2007, http://dels.nas.edu/Report/Toxicity-Testing-Twenty-first/11970.
10. D. Villeneuve, D. Crump, N. Garcia-Reyero, M. Hecker, T. Hutchinson, C. LaLone, B. Landesmann, T. Lettieri, S. Munn, M. Nepelska, M. Ottinger, L. Vergauwen and M. Whelan, Adverse Outcome Pathway (AOP) Development I: Strategies and Principles, *Toxicol. Sci.*, 2014, **142**(2), 312–320.
11. D. Villeneuve, D. Crump, N. Garcia-Reyero, M. Hecker, T. Hutchinson, C. LaLone, B. Landesmann, T. Lettieri, S. Munn, M. Nepelska, M. Ottinger, L. Vergauwen and M. Whelan, Adverse Outcome Pathway (AOP) Development II: Best Practices, *Toxicol. Sci.*, 2014, **142**(2), 321–330.
12. G. Langley, C. Austin, A. Balapure, L. Birnbaum, J. Bucher, J. Fentem, S. Fitzpatrick, J. Fowle, R. Kavlock, H. Kitano, B. Lidbury, A. Muotri, S.-Q. Peng, D. Sakharov, T. Seidle, T. Trez, A. Tonevitsky, A. van de Stolpe, M. Whelan and C. Willett, Lessons from Toxicology: Developing a 21st Century Paradigm for Medical Research, *Environ. Health Perspect.*, 2015, **123**(11), A268–A272.
13. T. Monticello, Drug Development and Nonclinical to Clinical Translational Databases: Past and Current Efforts, *Toxicol. Pathol.*, 2015, **43**, 57–61.
14. C. Keenan, S. Elmore, S. Francke-Carroll, R. Kemp, S. Peddada, J. Pletcher, M. Rinke, S. Schmidt, I. Taylor and D. Wolf, Best practices for the use of historical control data of proliferative rodent lesions, *Toxicol. Pathol.*, 2009, **37**, 679–693.
15. M. Sundqvist, A. Lundahl, M. Någård, U. Bredberg and P. Gennemark, Quantifying and communicating uncertainty in preclinical to clinical human dose-prediction, *CPT: Pharmacometrics Syst. Pharmacol.*, 2015, **4**, 243–254.
16. R. Garcia-Serna, D. Vidal, N. Remez and J. Mestres, Large-scale Predictive Drug Safety: From Structural Alerts to Biological Mechanisms, *Chem. Res. Toxicol.*, 2015, **28**, 1875–1887.

17. A. Jacobs, Prediction of 2-year carcinogenicity study results for pharmaceutical products. How are we doing? *Toxicol. Sci.*, 2005, **88**, 18–23.
18. J. Bowes, A. Brown, J. Hamon, W. Jarolimek, A. Sridhar, G. Waldon and S. Whitebread, Reducing safety-related drug attrition: the use of *in vitro* pharmacological profiling, *Nat. Rev. Drug Discovery*, 2012, **11**, 909–922.
19. E. Blomme and Y. Will, Toxicology Strategies for Drug Discovery: Present and Future, *Chem. Res. Toxicol.*, 2016, **29**(4), 473–504.
20. S. Ivanov, A. Lagunin, P. Pogodin, D. Filimonov and V. Poroikov, Identification of Drug Targets Related to the Induction of Ventricular Tachyarrhythmia through a Systems Chemical Biology Approach, *Toxicol. Sci.*, 2015, **145**(2), 321–336.
21. B. Verbist, G. Verheyen, L. Vervoort, M. Crabbe, D. Beerens, C. Bosmans, S. Jaensch, S. Osselaer, W. Talloen, I. Van den Wyngaert, G. Van Hecke, D. Wuyts, QSTAR Consortium, F. Van Goethem and H. Göhlmann, Integrating High-Dimensional Transcriptomics and Image Analysis Tools into Early Safety Screening: Proof of Concept for a New Early Drug Development Strategy, *Chem. Res. Toxicol.*, 2015, **28**, 1914–1925.
22. A. Voutchkova, T. Osimitz and P. Anastas, Toward a Comprehensive Molecular Design Framework for Reduced Hazard, *Chem. Rev.*, 2010, **110**, 5845–5882.
23. www.opentox.org/toxicity-prediction.
24. Lhasa Derek and Meteor Suites, http://www.lhasalimited.org.
25. Toxcast, http://www.epa.gov/ncct/toxcast/.
26. S. Wink, S. Hiemstra, S. Huppelschoten, E. Danen, M. Niemeijer, G. Hendriks, H. Vrieling, B. Herpers and B. van de Water, Quantitative High Content Imaging of Cellular Adaptive Stress Response Pathways in Toxicity for Chemical Safety Assessment, *Chem. Res. Toxicol.*, 2014, **27**, 338–355.
27. H. Larson, E. Chan, S. Sudarsanam and D. Johnson, Biomarkers, in *Computational Toxicology, Methods in Molecular Biology*, ed. B. Reisfeld and A. Mayeno, Humana Press, Springer Science, NY, 2013, ch. 11, pp. 253–273.
28. K. Groh, R. Carvalho, J. Chipman, N. Denslow, M. Halder, C. Murphy, D. Roelofs, A. Rolaki, K. Schirmer and K. Watanabe, Development and application of the Adverse Outcome Pathway framework for understanding and predicting chronic toxicity: I. Challenges and research needs in ecotoxicology, *Chemosphere*, 2015, **120**, 764–777.
29. J. Wambaugh, B. Wetmore, R. Pearce, C. Strope, R. Goldsmith, J. Sluka, A. Sedykh, A. Tropsha, S. Bosgra, I. Shah, R. Judson, R. Thomas and R. Setzer, Toxicokinetic Triage for Environmental Chemicals, *Toxicol. Sci.*, 2015, **147**(1), 55–67.
30. C. Grulke, K. Holm, M.-R. Goldsmith and Y.-M. Tan, PROcEED: Probabilistic reverse dosimetry approaches for estimating exposure distributions, *Bioinformation*, 2013, **9**(13), 707–709.
31. D. Wishart, D. Arndt, A. Pon, T. Sajed, A. Guo, Y. Djoumbou, C. Knox, M. Wilson, Y. Liang, J. Grant, Y. Liu, S. Goldansaz and S. Rappaport, T3DB: the toxic exposome database, *Nucleic Acids Res.*, 2015, **43**, D928–D934, DOI: 10.1093/nar/gku1004.

32. T. Adler, K. Sawyer, M. Shelton–Davenport and N. Grossblatt, The Exposome: A Powerful Approach for Evaluating Environmental Exposures and their Influences on Human Disease, *Emerging Science for Environmental Health Decisions Newsletter*, National Academies, 2010, Issue 2, ISSN 2376–1679.
33. NHANES, 2015. http://www.cdc.gov/exposurereport.
34. S. Bell and S. Edwards, Identification and Prioritization of Relationships between Environmental Stressors and Adverse Human Health Impacts, *Environ. Health Perspect.*, 2015, **123**(11), 1193–1199.
35. R. Turner, B. Park and M. Pirmohamed, Parsing interindividual drug variability: an emerging role for systems pharmacology, *Wiley Interdiscip. Rev.: Syst. Biol. Med.*, 2015, **7**(4), 221–241.
36. S. Berger and R. Iyengar, Network Analysis in Systems Pharmacology, *Bioinformatics*, 2009, **25**(19), 2466–2472.
37. S. Berger and R. Iyengar, Role of systems pharmacology in understanding adverse drug reactions, *Wiley Interdiscip. Rev.: Syst. Biol. Med.*, 2011, **3**(2), 129–135.
38. *Quantitative and Systems Pharmacology in the Post-genomic Era: New Approaches to Discovering Drugs and Understanding Therapeutic Mechanisms. An NIH White Paper by the QSP Workshop Group*, 2011, www.nigms.nih.gov/training/documents/systemspharmawpsorger2011.pdf.
39. S. Zhao and R. Iyengar, Systems Pharmacology: Network Analysis to Identify Multiscale Mechanisms of Drug Action, *Annu. Rev. Pharmacol. Toxicol.*, 2012, **52**, 505–521.
40. P. Vicini and P. van der Graaf, Systems Pharmacology for Drug Discovery and Development: Paradigm Shift or Flash in the Pan? *Clin. Pharmacol. Ther.*, 2013, **93**(5), 379–381.
41. X.-Y. Zhang, M. Trame, L. Lesko and S. Schmidt, Sobol Sensitivity Analysis. A Tool to Guide the Development and Evaluation of Systems Pharmacology Models, *CPT: Pharmacometrics Syst. Pharmacol.*, 2015, **4**, 69–79.
42. S. Visser, D. de Alwis, T. Kerbusch, J. Stone and S. Allerheiligen, Implementation of Quantitative and Systems Pharmacology in Big Pharma, *CPT: Pharmacometrics Syst. Pharmacol.*, 2014, **3**, e142, DOI: 10.1038/psp.2014.40, 1-10.
43. T. Papoian, H.-J. Chiu, I. Elayan, G. Jagadeesh, I. Khan, A. Laniyonu, C. Li, M. Saulnier, N. Simpson and B. Yang, Secondary pharmacology data to assess potential off-target activity of new drugs: a regulatory perspective, *Nat. Rev. Drug Discovery*, 2015, **14**(4), 294.
44. K. Biliouris, M. Lavielle and M. Trame, MatVPC: A user-friendly MATLAB-based toll for the simulation and evaluation of Systems Pharmacology models, *CPT: Pharmacometrics Syst. Pharmacol.*, 2015, **4**(9), 547–557.
45. Z. Wang, J. Li, R. Dang, L. Liang and J. Lin, PhIN: A Protein Pharmacology Interaction Network Database, *CPT: Pharmacometrics Syst. Pharmacol.*, 2015, **4**, 160–166.
46. L. Espinoza-Fonseca, The benefits of the multi-target approach in drug design and discovery, *Bioorg. Med. Chem.*, 2006, **14**, 896–897.

47. E. Leung, Z.-W. Cao, Z.-H. Jiang, H. Zhou and L. Liu, Network-based drug discovery by integrating systems biology and computational technologies, *Briefings Bioinf.*, 2013, **14**, 491–505.
48. J. Dudley, E. Schadt, M. Sirota, A. Butte and E. Ashley, Drug discovery in a multidimensional world: systems, patterns, and networks, *J. Cardiovasc. Transl. Res.*, 2010, **3**, 438–447.
49. T. Liu, Y. Lin, X. Ween, R. Jorissen and M. Gilson, BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities, *Nucleic Acids Res.*, 2007, **35**(suppl 1), D198–D201. Published online 2006.
50. A. Cami, A. Arnold, S. Manzi and B. Reis, Predicting Adverse Events using Pharmacological Network Models, *Sci. Transl. Med.*, 2011, **3**(114), 1–10.
51. D.-S. Cao, N. Xiao, Y.-J. Li, W.-B. Zeng, Y.-Z. Liang, A.-P. Lu, Q.-S. Xu and A. Chen, Integrating Multiple Evidence Sources to Predict Adverse Drug Reactions Based on a Systems Pharmacology Model, *CPT: Pharmacometrics Syst. Pharmacol.*, 2015, **4**, 498–506.
52. T. Lorderbaum, M. Nasir, M. Keiser, S. Vilar, G. Hripcsak and N. Tatonetti, Systems Pharmacology Augments Drug Safety Surveillance, *Clin. Pharmacol. Ther.*, 2015, **97**(2), 151–158.

CHAPTER 2

Databases Facilitating Systems Biology Approaches in Toxicology

DALE E. JOHNSON*[a,b] AND ANN M. H. HESLIN[b]

[a]University of Michigan, School of Public Health, Department of Environmental Health Sciences, Ann Arbor, MI 48109-2029, USA; [b]University of California, Berkeley, Department of Nutritional Sciences and Toxicology, Morgan Hall, Berkeley, CA 94720-3104, USA
*E-mail: daleej@umich.edu

2.1 Introduction

Several authors have published reviews of the application of systems biology approaches to toxicology, all of which require the use of multiple data sets and resources to draw novel inferences about mechanisms and modes of action of chemicals on biological targets, networks, and systems. These approaches include those in therapeutics research, particularly in understanding and predicting adverse drug reactions, and sometimes environmental or eco-toxicology approaches where information on the chemical(s) in question is sparse.[1–7] As discussed by Sturla and colleagues[5], the core objective of systems toxicology is to uncover and hopefully elucidate mechanisms that causally link exposure to active substances with chemically induced adverse events and disease. The process requires the collection of quantifiable experimental

Issues in Toxicology No. 31
Computational Systems Pharmacology and Toxicology
Edited by Dale E. Johnson and Rudy J. Richardson

Published by the Royal Society of Chemistry, www.rsc.org

data typically coupled with extensive information gained by processing large sets of data positioned within biological networks and pathways. These data are garnered from accessible databases to allow the reflection of molecular changes in the context of cellular, tissue-level, or physiological changes that are linked to disease phenotypes or adverse events at the organism level.[7] Several authors have published lists of relevant databases for toxicology data, both for therapeutic and environmental research.[1,8–10] Accordingly, systems toxicology relies heavily on computational approaches to manage, analyze, and interpret these data with the ultimate goal to aid in the development of predictive *in silico* models that can be used in risk assessment. Computational systems toxicology has the following major areas of focus.[6]

(1) Analyzing the massive amounts of *in vitro* and *in vivo* data contained in databases generated by multiple methods and correlating structural features of the compounds with levels of exposure and outcome.
(2) Representing the relevant mechanisms leading to an adverse outcome as biological network models that describe the normal state and the causal effect of their perturbations upon exposure to chemical compounds.
(3) Quantifying the dose-dependent and time-resolved perturbations of these biological networks their overall biological impact upon exposure and assessing risk.
(4) Building and validating adequate computational models with predictive power that can be applied to risk assessment.

In the adverse outcome pathway model, the sequence of events that lead to an adverse outcome span multiple levels of biological organization,[8] but always contain a molecular initiating event, which is defined as the initial interaction between a chemical molecule and a biomolecule or biosystem that can be causally linked to an outcome *via* a pathway.[8] In the therapeutics toxicology field, systems pharmacology, an emerging interdisciplinary field combining network and chemical biology, provides important tools to uncover and understand adverse drug reactions and may mitigate the drawbacks of traditional methods. In particular, network analysis allows researchers to integrate heterogeneous data sources and quantify the interactions between biological and chemical entities. Recent work in this area has combined chemical, biological, and large-scale observational health data to predict adverse drug reactions in individual patients and global populations.[11]

As mentioned in Chapter 1, several factors have been involved in the rapid changes seen in the toxicology field, including the understanding of genetic changes, which have enhanced the ability to understand diseases, and chemically-induced toxicities at the molecular level. "Omics" technologies and approaches such as transcriptomics, metabolomics, proteomics, and epigenomics have changed the way mechanisms of toxicity and the perturbation of biological systems that lead to adverse outcomes are viewed.[5] These advances have been coupled with the public availability of large data sets of

information and new modeling approaches that have enhanced the ability to understand toxicological events and effects at multiple biological levels.[6] Since scientific approaches, inquiries, and visions aimed at understanding toxicological events and outcomes have been broadened tremendously, this reinforces the need for new and better ways to assess toxicity and risk.

2.2 Categorized Lists of Databases for Systems Toxicology

The following list includes free on-line data sources and tools placed into categories based on source and content.

2.2.1 TOXNET Databases (Including Those with Direct Links from TOXNET)

TOXicology Data NETwork (TOXNET)[12] is a central website hub with links to several toxicology data files that report on several chemicals, primarily those of toxicological and environmental concern. The breadth of information includes chemical nomenclature and current literature that gives evidence and/or speculation on a chemical's toxicological effects.

- *HSDB*[13] (*Hazardous Substances Data Bank*) Peer-reviewed toxicology data for >5000 hazardous chemicals. Data can be searched by relevance or filter by larger category groupings: human health effects, emergency medical treatment, animal toxicity studies, metabolism/pharmacokinetics, pharmacology, environmental fate/exposure, environmental standards & regulations, chemical/physical properties, chemical safety & handling, occupational exposure, standards, manufacturing/use information, laboratory methods, special references, synonyms and identifiers, and administrative information. Additional features of HSDB include "review status tags" that indicate the level of quality review: peer reviewed, QC reviewed (quality control review that has not yet been officially reviewed), and un-reviewed (statements that do not necessarily need scientific review). A complete list of chemicals in the HSDB is available at https://sis.nlm.nih.gov/enviro/hsdbchemicalslist.html.
- *TOXLINE*[14] 5 million references from specialized journals, government reports, meeting abstracts, and other relevant collections of toxicology information. The collection of information includes biochemical, pharmacological, physiological, and toxicological effects of drugs and other chemicals.
- *ChemIDPlus*[15,16] Dictionary of >400 000 chemicals including names, synonyms, and structures; a chemical searching system generated from more than 100 sources. National Library of Medicine databases serve as its primary source of information; however, it also compiles

information from the Canadian Domestic Substances List, European Inventory of Existing Commercial Chemical Substances (EINECS), Environmental Protection Agency (EPA) Toxic Substances Control Act (TSCA) Chemical Substance Inventory, the SUPERLIST set of regulatory resources, and other internet databases such as EPA Substance Registry System, the Food and Drug Administration (FDA) Drugs@FDA system, International Agency for Research on Cancer (IARC), National Institute of Allergy and Infectious Diseases (NIAID), and the National Institute of Standards and Technology (NIST) Chemistry WebBook. All chemicals are searchable by name, synonym, Chemical Abstracts Service (CAS) registry number, molecular formula, classification code, locator code, structure, and/or physical properties. Two versions of ChemID*plus* exist: ChemID*plus*Lite[15] and ChemIDplusAdvanced.[16] ChemID*plus*Lite provides chemical information searching and links to other resources. Unlike ChemID*plus*Advanced, ChemID*plus*Lite does not require plugins or applets. As such, ChemID*plus*Advanced has more advanced search capabilities. The default structure editor is Marvin for JavaScript (ChemAxon) that allows a user to download a single structure Mol file in ChemIDplus. This applet enables users to conduct advanced chemical structure queries (substructure search, similarity search, exact structure search, flex search, and flexplus search), and filter similar and substructure chemical scaffolds. Structure descriptors include *InChI*™ (International Chemical Identifier), InChIKey, and *SMILES*™ (simplified molecular input line entry system) notations, all of which may be downloaded. InChiKeys link to other search engines to find a structure in other systems. Both Lite and Advanced ChemID*plus* records are updated daily. Although ChemIDplus no longer supports Chime (it's previous free chemical display application), another structure-drawing package, Accelrys Draw No Fee, is now publicly accessible.

- *LactMed*[17] drugs and lactation database lists drugs and other chemicals to which breastfeeding mothers may be exposed. It includes information on the levels of such substances in breastmilk and infant blood, and the possible adverse effects in the nursing infant. Suggested therapeutic alternatives to those drugs are provided, where appropriate. All data are derived from the scientific literature and fully referenced.
- *DART*[18] Developmental and Reproductive Toxicology Database and references. It provides >200 000 journal references covering teratology and other aspects of developmental and reproductive toxicology. DART is created from a search profile using PubMed.
- *TOXMAP*[19] TOXMAP is a geographic information system from the Division of Specialized Information Services of the US National Library of Medicine that uses maps of the United States to show the amount and location of toxic chemicals released into the environment. Users can visually explore data derived from the EPA's Toxics Release Inventory (TRI), which provides information on the releases of toxic chemicals into the environment as reported annually by industrial facilities

around the United States. TOXMAP also contains information from the EPA's superfund program, as well as some non-EPA datasets such as the US Census and National Cancer Institute health data.

- *TRI*[20] Toxics Release Inventory. Annual environmental releases of more than 600 toxic chemicals by US facilities as reported annually to the EPA by US industrial and federal facilities. TRI's data reports, beginning with the 1987 reporting year, contain information about the types and amounts of toxic chemicals that are released each year to the air, water, land and by underground injection, as well as information on the quantities of toxic chemicals sent to other facilities for further waste management In agreement with the Pollution Prevention Act of 1990, source reduction and recycling data are also included in TRI.
- *Household Products Database*[21] Potential health effects of chemicals in >10 000 common household products. Information is also available for some industrial-grade products. Products can be searched by brand name, product type, manufacturer, ingredient/chemical, and by health effects. The record for each product shows the ingredients as reported by the manufacturer. For many products, a link to the manufacturer's material safety data sheet is provided, which includes more information such as handling, disposal, and health effects.
- *Haz-Map*[22] Links jobs and hazardous tasks with occupational diseases and their symptoms, in which causality from chemical and/or biological agents has been established based on current scientific evidence.
- *IRIS*[23] Integrated Risk Information System. IRIS contains data in support of human health risk assessment, including hazard identification and dose–response assessments of more than 550 chemicals (as of mid-2014) that evaluate information on health effects (cancer and non-cancer) resulting from exposure to environmental contaminants. IRIS data are reviewed by EPA scientists several times a year and represents EPA consensus.
- *ITER*[24] International Toxicity Estimates for Risk. Risk information for more than 600 chemicals of environmental concern from authoritative groups worldwide. ITER integrates data from Centers for Disease Control Agency for Toxic Substances and Disease Registry, Health Canada, RIVM, US EPA, IARC, NSF International, and independent parties offering peer-reviewed risk values. It is compiled by Toxicology Excellence for Risk Assessment (TERA) and its records that are updated multiple times a year. The *Risk Information Exchange* (RiskIE; http://www.allianceforrisk.org/RiskIE.htm) is a companion database to ITER. It includes in-progress or recently completed risk assessment projects. RiskIE is a database of notifications about a variety of human health risk assessment projects that are underway or recently completed. Projects listed on RiskIE are both chemical- and nonchemical-specific, and range from many types of risk value development to risk methods document development. RiskIE currently tracks >4000 in-progress or recently completed risk assessment projects conducted by 35 different organizations representing 13 different countries.

- *ALTBIB*[25] Resources on Alternatives to the Use of Live Vertebrates in Biomedical Research and Testing.
- *CCRIS*[26] Chemical Carcinogenesis Research Information System. Carcinogenicity and mutagenicity test results for >8000 chemicals (no longer updated).
- *CPDB*[27] Carcinogenic Potency Database. Standardized analyses of the results of 6540 chronic, long-term animal cancer tests (no longer updated).
- *GENE-TOX*[28] Genetic Toxicology Data Bank. Peer-reviewed genetic toxicology test data for more than 3000 chemicals (no longer updated).

2.2.2 US EPA Chemical Toxicity Databases

Chemical safety research data are made publicly available, including rapid, automated (high-throughput) chemical screening data; aggregated public sources of chemical toxicity data; animal toxicity studies; chemical exposure data and prediction models; and high quality chemical structures and annotations.

- *ToxCast and Tox21 Data*[29] Data on >2000 chemicals evaluated in more than 700 high-throughput assays.
- *ACToR*[30] ACToR (Aggregated Computational Toxicology Resource) is the EPA's online warehouse of all publicly available chemical toxicity data and can be used to find all publicly available data about potential chemical risks to human health and the environment.
- *ToxRefDB*[31] The Toxicity Reference Database (ToxRefDB) contains thousands of animal toxicity testing results, currently on 474 chemicals.
- *DSSTox*[32] DSSTox provides the accurate mapping of bioassay and physicochemical property data associated with chemical substances to their corresponding chemical structures.
- *CHAD*[33] Consolidated Human Activity Database; exposure and time-use studies. Data can be downloaded.
- *CPCat*[34] The Chemical and Product Categories database; catalogs the use of >40 000 chemicals and their presence in different consumer products.
- *iCSS Dashboard*[35] Developing tool expected to be the online portal for all chemical research data and studies.

2.2.3 National Toxicology Program Databases

- *CEBS*[36] The Chemical Effects in Biological Systems database houses data of interest to environmental health scientists. CEBS is a public resource, and has received depositions of data from academic, industrial, and governmental laboratories. CEBS is designed to display data

in the context of biology and study design, and to permit data integration across studies for novel meta-analysis. Users are required to install scripts for using Adobe Flex applications with JAWS.

- *DrugMatrix*[37] Comprehensive results of thousands of highly controlled and standardized toxicological experiments in which rats or primary rat hepatocytes were systematically treated with therapeutic, industrial, and environmental chemicals at both non-toxic and toxic doses.
- *ToxFx*[38] An automated toxicogenomics analysis application utilizing toxicogenomic signatures, specially curated biochemical pathways, and other relevant data to interpret toxicity-related transcriptomic data and present the results as a detailed, customized report.
- *Historical Control Databases*[39–41] Historical control data from NTP toxicity studies. Tumor incidences and growth and survival curves for control animals from NTP's 2 year carcinogenesis studies are summarized by species, sex, route of administration, and vehicle. The historical controls for the genetically modified models are also reported.
- *NICEATM LLMA Database*[42] (NTP Interagency Center for the Evaluation of Alternative Toxicological Methods) Analyses to evaluate the usefulness of the murine local lymph node assay (LLNA) to identify potential skin sensitizers.

2.2.4 Additional Toxicity Databases

- *T3DB*[43] The Toxin and Toxin Target Database is a bioinformatics resource that as its future name, Toxic Exposome Database, implies, is specifically designed to capture information about the toxic exposome. The focus of the T3DB is providing mechanisms of toxicity and target proteins for each toxin interactively linked in both directions. It is also fully searchable and supports extensive text, sequence, chemical structure, and relational query searches. The user can also hyperlink information into other databases without re-entering chemical information. The database currently houses 3673 toxins described by 41 733 synonyms, including pollutants, pesticides, drugs, and food toxins, which are linked to 2087 corresponding toxin target records. Altogether there are 42 471 toxin and toxin target associations. Each toxin record (ToxCard) contains more than 90 data fields and holds information such as chemical properties and descriptors, toxicity values, molecular and cellular interactions, and medical information. This information has been extracted from >18 143 sources which include other databases, government documents, books, and scientific literature. It is both modeled after and closely linked to the Human Metabolome Database and DrugBank.
- *FAERS*[44] Adverse Effects Reporting system of post-market safety surveillance for all approved drug and therapeutic biological products.
- *SIDER*[45](Side Effect Resource) Information on marketed drugs and their recorded adverse drug reactions.

- *VAERS*[46] (Vaccine Adverse Event Reporting System).
- *JECDB*[47] (Japan Existing Chemical Data Base) Safety examination of existing chemicals and safety programs in Japan.
- *Offsides*[48] Finds different associations from adverse events reported during clinical trials before drug approval. The Offsides database is a resource of 438 801 off-label—those effects not listed on the FDA's official drug label—side effects for 1332 drugs and 10 097 adverse events.
- *Twosides*[49] A resource of polypharmacy side effects for pairs of drugs. This database contains 868 221 significant associations between 59 220 pairs of drugs and 1301 adverse events. These associations are limited to those that cannot be clearly attributed to either drug alone.
- *DITOP*[50] A comprehensive database providing Drug-Induced Toxicity Related Protein information. The related toxicities include overdose toxicity, idiosyncratic toxicity, drug–drug interactions, and genetic toxicity.

2.2.5 Chemical–Gene–Protein Databases

- *CTD*[51] Comparative Toxicogenomics Database. Provides access to scientific data describing relationships between chemicals, genes, and human diseases. The database contains curated data that describes cross-species interactions for chemical–gene, chemical–protein, and gene–disease. KEGG (Kyoto Encyclopedia of Genes and Genomes) and Reactome pathway data describe known molecular interaction and reaction networks. These data are integrated with chemicals, genes, and diseases in CTD to provide insights into molecular networks that may be affected by chemicals, and possible mechanisms underlying environmental diseases. CTD has a hierarchical arrangement of interactions that characterize physical, regulatory, and biochemical interactions. This vocabulary comprises 70 terms, including *actions* (*e.g.* “binds to”, “imports”), *operators* that describe the degree of a chemical's effect (*e.g.* “increases”), and *qualifiers* that specify the form of the gene or chemical involved in an interaction (*e.g.* “protein” or “chemical metabolite”, respectively). The chemical category integrates a chemical subset of the Medical Subject Headings (MeSH®), the hierarchical vocabulary from the US National Library of Medicine. The information about chemicals includes chemical structures, curated interacting genes and proteins, curated and inferred disease relationships, and enriched pathways and functional annotations. CTD contains curated and inferred chemical–disease and gene–disease associations. Inferred associations are established *via* CTD-curated chemical–gene interactions and inference scores are calculated for all inferred relationships. In gene–gene interactions, CTD represents gene–gene interactions from BioGRID (see later) that consist of genetic and protein interactions curated from primary literature for all major model organisms by BioGRID curators. These interactions are available for each gene and reference, and for the inference networks underlying each chemical–disease association.

In addition, the user can generate pathways for custom collections of genes using the set analyzer tool. Several other databases and tools that associate chemicals–genes–diseases use CTD as the primary source of information.

- *ChemDIS*[52] A chemical–disease inference system based on chemical–protein interactions.
- *STITCH*[53] (Search Tool for Interacting Chemicals) A resource to explore known and predicted interactions of chemicals and proteins. Chemicals are linked to other chemicals and proteins by evidence derived from experiments, databases, and the literature.
- *String*[54] Protein–protein interactive networks.
- *Chemprot*[55] A publicly available compilation of chemical–protein–disease annotation resources that enables the study of systems pharmacology for a small molecule across multiple layers of complexity from molecular to clinical levels.
- *Human Protein Atlas*[56] An interactive database and visualization tool for the human proteome based on antibody methods and transcriptomics analysis across all major tissues and organs of the human body.
- *BioGrid*[57] (Biological General Repository for Interaction Datasets) A searchable interaction repository with data compiled through comprehensive curation efforts. Includes protein and genetic interactions and post-translational modifications from major model organisms. All data can be freely downloaded.

2.2.6 Pathway-Network Databases

- *KEGG*[58] KEGG is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies. The KEGG home page has links to several data-oriented entries including: KEGG pathway maps, BRITE functional hierarchies, KEGG modules, ortholog groups, genomes, genes and proteins, small molecules, glycans, biochemical reactions, enzymes, human diseases, drugs, and health information resources. There are also several analytical tools including mapping tools, pathogen checker of antimicrobial resistance genes, sequence similarity search, and chemical similarity search. The KEGG is a primary source of information for most databases that include pathway information.
- *Reactome*[59] Reactome is a curated and peer-reviewed pathway database that provides intuitive bioinformatics tools for visualization, interpretation, and analysis of pathway knowledge. Data mining and analysis tools include a pathway browser, tools that merge information into a portal, species comparisons, cystoscope plug-in, small molecule search, and literature citation searching.

- *Cytoscape*[60] Software platform for visualizing molecular interaction networks and biological pathways. Networks can be integrated with annotations, gene expression profiles, and other data. Additional features include apps, formerly called plug-ins (mostly free) which allow network and molecular profiling, different layouts, and connection with databases. Cytoscape links are frequently available with several other databases.
- *Pathway Commons*[61] Pathway Commons stores and disseminates knowledge about biological pathways: >42 000 pathways and 1 350 000 interactions from 22 data sources.
- *NDEx*[62] The Network Data Exchange provides an open-source framework where scientists and organizations can share, store, manipulate, and publish biological network knowledge.

2.2.7 Chemistry, Structural Alert, and QSAR Databases and Tools

- *PubChem*[63] Provides information on the biological activities of small molecules. PubChem is organized as three linked databases within the National Center for Biotechnology Information's Entrez information retrieval system. These are PubChem Substance, PubChem Compound, and PubChem BioAssay. PubChem also provides a fast chemical structure similarity search tool.
- *ChemSpider*[64] ChemSpider is a free chemical structure database providing fast text and structure search access to over 50 million structures from hundreds of data sources.
- *ChemProp*[65] Several modules including structural alerts for electrophilic reactivity.
- *CORAL*[66]Quantitative structure–property relationships (QSPR)/quantitative structure–activity relationships (QSAR) analysis for several toxicity endpoints.
- *OECD Toolbox*[67] QSAR toolbox for grouping chemicals into categories.
- *TEST*[68] The Toxicity Estimation Software Tool allows users to easily estimate the toxicity of chemicals using QSAR methodologies.
- *Virtual Computational Chemistry Laboratory*[69] On-line cheminformatics tools to calculate chemical properties including A Log P.
- *Danish (Q)SAR Database*[70] This Danish (Q)SAR database is a repository of estimates from over 70 (Q)SAR models for 166 072 chemicals. The (Q)SAR models encompass endpoints for physicochemical properties, fate, eco-toxicity, absorption, metabolism and toxicity.
- *Advaitabio: iPathwayGuide*[71] presents an advanced pathway analysis platform for high-throughput sequencing data.
- *LAZAR*[72] (Lazy Structure–Activity Relationships) Takes a chemical structure as input and provides several toxicity predictions. LAZAR is built on top of OpenTox www.opentox.org/.

- *Toxtree*[73] Toxtree is a full-featured and flexible user-friendly open source application, which is able to estimate toxic hazard by applying a decision-tree approach. Toxtree could be applied to datasets from various compatible file types. User-defined molecular structures are also supported—they could be entered by SMILES, or by using the built-in 2D structure diagram editor. Toxtree currently has the following plug-ins:

  - Cramer rules and Cramer rules with extensions
  - Verhaar scheme and modified verhaar scheme
  - skin irritation prediction
  - eye irritation prediction
  - START biodegradation and persistence
  - Benigni/Bossa rulebase for mutagenicity and carcinogenicity
  - *in vitro* mutagenicity (Ames test) alerts by ISS
  - structure alerts for the *in vivo* micronucleus assay in rodents (ISSMIC)
  - structural alerts for functional group identification (ISSFUNC)
  - structure alerts for identification of Michael acceptors
  - structure alerts for skin sensitization reactivity domains
  - DNA binding alerts
  - Protein binding alerts
  - Kroes thresholds of toxicological concern decision tree
  - SMARTCyp: cytochrome P450-mediated drug metabolism and metabolites prediction

- *CAESAR*[74] QSAR models supporting the REACH legislation including bioconcentration, skin sensitization, carcinogenicity, and developmental toxicity.
- *ToxAlerts*[75] A web-based platform for collecting and storing toxicological structural alerts from literature and for virtual screening of chemical libraries to flag potentially toxic chemicals and compounds that can cause adverse side effects. An alert is uniquely identified by a SMARTS template, a toxicological endpoint, and a publication where the alert was described. Additionally, the system allows storing complementary information such as name, comments, and mechanism of action, as well as other data.
- *Online Chemical Database*[76] The Online Chemical Modeling Environment is a web-based platform that aims to automate and simplify the typical steps required for QSAR modeling. The platform consists of two major subsystems: the database of experimental measurements and the modeling framework. A user-contributed database contains a set of tools for easy input, search and modification of thousands of records. Includes chemical property predictions, ToxAlert screening, and optimization of different properties with MolOptimizer.
- *ToxRead*[77] Software to assist in making reproducible read-across evaluations. The software shows similar chemicals to the input chemical, structural alerts, and other relevant features in common between chemicals.

2.2.8 Drug and Drug Target Databases

- *DrugBank*[78] The DrugBank database is a bioinformatics and cheminformatics resource that contains detailed drug data with correlating comprehensive drug target, including sequence, structure, and pathway information. The database includes >8000 drug entries including approved small molecules, biologics, nutraceuticals, and >6000 experimental drugs. DrugBank is the source of drug information for most databases that incorporate drug information.
- *TTD*[79] (Therapeutic Target Database) A database that provides information about the known and explored therapeutic protein and nucleic acid targets, the targeted disease, pathway information, and corresponding drugs directed at each target. Included are links to all relevant databases where detailed information exists.
- *Chemmapper*[80] An online platform to predict polypharmacy effects and mode of action for small molecules based on 3D similarity computation. ChemMapper collects >350 000 chemical structures with bioactivities and associated target annotations (as well as >3 000 000 non-annotated compounds for virtual screening).
- *Pharmmapper*[81] An updated integrated pharmacophore-matching platform with statistical methods for potential target identification. A pharmacophore database extracted from all targets in TargetBank, DrugBank, BindDB, and PDTD (Tripos mol2 or MDL SDF formats).
- *PDTD*[82] (Potential Drug Target Database) A dual function database of known and potential drug targets focusing on targets with known 3D structures. PDTD contains 1207 entries covering 841 known and potential drug targets with structures from the Protein Data Bank. This is connected to a docking program, *Tarfisdock*[83] that docks a small molecule into protein targets in PDTD (mol2 formats).
- *PK/DB*[84] (Database for pharmacokinetic properties) was designed with the aim of creating robust databases for pharmacokinetic studies and *in silico* absorption, distribution, metabolism, and excretion (ADME) prediction. The database contains high-quality data for structurally diverse compounds associated with known ADME properties, including human oral bioavailability, human intestinal absorption, plasma protein binding, blood–brain barrier, among others. PK/DB manages 1389 compounds incorporating structurally diverse drug-like and lead-like molecules which represent 4141 pharmacokinetic measurements, including five validated models for *in silico* ADME prediction.
- *BindingDB Binding Database*[85] is a database of measured binding affinities, focusing chiefly on the interactions of protein considered to be drug-targets with small, drug-like molecules. The database contains 1 233 342 binding data, for 6352 protein targets and 541 006 small molecules. There are 2907 protein–ligand crystal structures with BindingDB affinity measurements for proteins with 100% sequence identity, and 7392 crystal structures for proteins with 85% sequence identity.