

Application of Modern Toxicology Approaches for Predicting Acute Toxicity for Chemical Defense

DETAILS

130 pages | 8.5 x 11 | PAPERBACK | ISBN 978-0-309-37666-2

AUTHORS

Committee on Predictive-Toxicology Approaches for Military Assessments of Acute Exposures; Board on Environmental Studies and Toxicology; Board on Life Sciences; Division on Earth and Life Studies; The National Academies of Sciences, Engineering, and Medicine

BUY THIS BOOK

FIND RELATED TITLES

Visit the National Academies Press at NAP.edu and login or register to get:

- Access to free PDF downloads of thousands of scientific reports
- 10% off the price of print titles
- Email or social media notifications of new titles related to your interests
- Special offers and discounts



Distribution, posting, or copying of this PDF is strictly prohibited without written permission of the National Academies Press. (Request Permission) Unless otherwise indicated, all materials in this PDF are copyrighted by the National Academy of Sciences.

Application of Modern Toxicology Approaches for Predicting Acute Toxicity for Chemical Defense

Committee on Predictive-Toxicology Approaches for
Military Assessments of Acute Exposures

Committee on Toxicology

Board on Environmental Studies and Toxicology

Board on Life Sciences

Division on Earth and Life Studies

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

THE NATIONAL ACADEMIES PRESS

Washington, DC

www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, NW Washington, DC 20001

This project was supported by Contract W81K04-11-D-0017, TO#8 between the National Academies of Sciences, Engineering, and Medicine and the US Department of Defense. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the view of the organizations or agencies that provided support for this project.

International Standard Book Number-13: 978-0-309-37666-2

International Standard Book Number-10: 0-309-37666-1

Additional copies of this report are available for sale from the National Academies Press, 500 Fifth Street, NW, Keck 360, Washington, DC 20001; (800) 624-6242 or (202) 334-3313; <http://www.nap.edu>.

Copyright 2015 by the National Academy of Sciences. All rights reserved.

Printed in the United States of America

Suggested citation: The National Academies of Sciences, Engineering, and Medicine. 2015. *Application of Modern Toxicology Approaches for Predicting Acute Toxicity for Chemical Defense*. Washington, DC: The National Academies Press.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

The **National Academy of Sciences** was established in 1863 by an Act of Congress, signed by President Lincoln, as a private, nongovernmental institution to advise the nation on issues related to science and technology. Members are elected by their peers for outstanding contributions to research. Dr. Ralph J. Cicerone is president.

The **National Academy of Engineering** was established in 1964 under the charter of the National Academy of Sciences to bring the practices of engineering to advising the nation. Members are elected by their peers for extraordinary contributions to engineering. Dr. C. D. Mote, Jr., is president.

The **National Academy of Medicine** (formerly the Institute of Medicine) was established in 1970 under the charter of the National Academy of Sciences to advise the nation on medical and health issues. Members are elected by their peers for distinguished contributions to medicine and health. Dr. Victor J. Dzau is president.

The three Academies work together as the **National Academies of Sciences, Engineering, and Medicine** to provide independent, objective analysis and advice to the nation and conduct other activities to solve complex problems and inform public policy decisions. The Academies also encourage education and research, recognize outstanding contributions to knowledge, and increase public understanding in matters of science, engineering, and medicine.

Learn more about the National Academies of Sciences, Engineering, and Medicine at **www.national-academies.org**.

**COMMITTEE ON PREDICTIVE-TOXICOLOGY APPROACHES FOR
MILITARY ASSESSMENTS OF ACUTE EXPOSURES**

Members

DAVID C. DORMAN (*Chair*), North Carolina State University, Raleigh, NC
WEIHSUEH A. CHIU, Texas A&M University, College Station, TX
HAIYAN HUANG, University of California, Berkeley, CA
ANDY NONG, Health Canada, Ottawa, ON
GRACE PATLEWICZ, US Environmental Protection Agency, Research
Triangle Park, NC
DAVID REIF, North Carolina State University, Raleigh, NC
JOHN WADE, Battelle, Arlington, VA
KATRINA WATERS, Pacific Northwest National Laboratories, Richland, WA
BARBARA A. WETMORE, The Hamner Institutes for Health Sciences,
Research Triangle Park, NC
YVONNE WILL, Pfizer, Groton, CT

Staff

ELLEN K. MANTUS, Project Director
MARILEE SHELTON-DAVENPORT, Senior Program Officer
KERI STOEVER, Research Associate
NORMAN GROSSBLATT, Senior Editor
MIRSADA KARALIC-LONCAREVIC, Manager, Technical Information Center
RADIAH ROSE-CRAWFORD, Manager, Editorial Projects
IVORY CLARKE, Senior Program Assistant

Sponsor

US DEPARTMENT OF DEFENSE

COMMITTEE ON TOXICOLOGY

Members

GARY P. CARLSON (*Chair*), Purdue University (retired), West Lafayette, IN
LAWRENCE S. BETTS, Eastern Virginia Medical School, Norfolk
DEEPAK K. BHALLA, Wayne State University, Detroit, MI
DEBORAH A. CORY-SLECHTA, University of Rochester School of Medicine and Dentistry,
Rochester, NY
MARY E. DAVIS, West Virginia University, Morgantown
DAVID C. DORMAN, North Carolina State University, Raleigh
MARGARET M. MACDONELL, Argonne National Laboratory, Argonne, IL
IVAN RUSYN, Texas A&M University, College Station
KENNETH R. STILL, Portland State University, Portland, OR
JOYCE S. TSUJI, Exponent, Inc., Bellevue, WA

Staff

SUSAN N.J. MARTEL, Senior Program Officer for Toxicology
ELLEN K. MANTUS, Scholar and Director of Risk Assessment
MIRSADA KARALIC-LONCAREVIC, Manager, Technical Information Center
RADIAH ROSE, Manager, Editorial Projects
TAMARA DAWSON, Program Associate

BOARD ON ENVIRONMENTAL STUDIES AND TOXICOLOGY¹

Members

ROGENE F. HENDERSON (*Chair*), Lovelace Respiratory Research Institute, Albuquerque, NM
PRAVEEN AMAR, Independent Consultant, Lexington, MA
RICHARD A. BECKER, American Chemistry Council, Washington, DC
MICHAEL J. BRADLEY, M.J. Bradley & Associates, Concord, MA
JONATHAN Z. CANNON, University of Virginia, Charlottesville, VA
GAIL CHARNLEY ELLIOTT, HealthRisk Strategies, Washington, DC
DOMINIC M. DiTORO, University of Delaware, Newark, DE
DAVID C. DORMAN, North Carolina State University, Raleigh, NC
CHARLES T. DRISCOLL, JR., Syracuse University, Syracuse, NY
WILLIAM H. FARLAND, Colorado State University, Fort Collins, CO
LYNN R. GOLDMAN, The George Washington University, Washington, DC
LINDA E. GREER, Natural Resources Defense Council, Washington, DC
WILLIAM E. HALPERIN, University of Medicine and Dentistry of New Jersey, Newark, NJ
STEVEN P. HAMBURG, Environmental Defense Fund, New York, NY
ROBERT A. HIATT, University of California, San Francisco, CA
PHILIP K. HOPKE, Clarkson University, Potsdam, NY
SAMEUL KACEW, University of Ottawa, Ontario
H. SCOTT MATTHEWS, Carnegie Mellon University, Pittsburgh, PA
THOMAS E. MCKONE, University of California, Berkeley, CA
TERRY L. MEDLEY, E. I. du Pont de Nemours & Company, Wilmington, DE
JANA MILFORD, University of Colorado at Boulder, Boulder, CA
MARK A. RATNER, Northwestern University, Evanston, IL
JOAN B. ROSE, Michigan State University, East Lansing, MI
GINA M. SOLOMON, California Environmental Protection Agency, Sacramento, CA
PETER S. THORNE, University of Iowa, Iowa City, IA
JOYCE S. TSUJI, Exponent, Inc., Bellevue, WA

Senior Staff

JAMES J. REISA, Senior Director
ELLEN K. MANTUS, Scholar and Director of Risk Assessment
RAYMOND A. WASSEL, Scholar and Director of Environmental Studies
DAVID J. POLICANSKY, Scholar
SUSAN N.J. MARTEL, Senior Program Officer for Toxicology
MIRSADA KARALIC-LONCAREVIC, Manager, Technical Information Center
RADIAH ROSE-CRAWFORD, Manager, Editorial Projects

¹This study was planned, overseen, and supported by the Board on Environmental Studies and Toxicology.

OTHER REPORTS OF THE BOARD ON ENVIRONMENTAL STUDIES AND TOXICOLOGY

Review of California's Risk-Assessment Process for Pesticides (2015)
Sustainability Concepts in Decision Making, Tools and Approaches for the US Environmental Protection Agency (2014)
Rethinking the Components, Coordination, and Management of US Environmental Protection Agency Laboratories (2014)
Review of the Formaldehyde Assessment in the National Toxicology Program 12th Report on Carcinogens (2014)
Review of the Styrene Assessment in the National Toxicology Program 12th Report on Carcinogens (2014)
Review of EPA's Integrated Risk Information System (IRIS) Process (2014)
Review of the Environmental Protection Agency's State-of-the-Science Evaluation of Nonmonotonic Dose-Response Relationships as They Apply to Endocrine Disruptors (2014)
Assessing Risks to Endangered and Threatened Species from Pesticides (2013)
Science for Environmental Protection: The Road Ahead (2012)
Exposure Science in the 21st Century: A Vision and A Strategy (2012)
A Research Strategy for Environmental, Health, and Safety Aspects of Engineered Nanomaterials (2012)
Macondo Well-Deepwater Horizon Blowout: Lessons for Improving Offshore Drilling Safety (2012)
Feasibility of Using Mycoherbicides for Controlling Illicit Drug Crops (2011)
Improving Health in the United States: The Role of Health Impact Assessment (2011)
A Risk-Characterization Framework for Decision-Making at the Food and Drug Administration (2011)
Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde (2011)
Toxicity-Pathway-Based Risk Assessment: Preparing for Paradigm Change (2010)
The Use of Title 42 Authority at the U.S. Environmental Protection Agency (2010)
Review of the Environmental Protection Agency's Draft IRIS Assessment of Tetrachloroethylene (2010)
Hidden Costs of Energy: Unpriced Consequences of Energy Production and Use (2009)
Contaminated Water Supplies at Camp Lejeune—Assessing Potential Health Effects (2009)
Review of the Federal Strategy for Nanotechnology-Related Environmental, Health, and Safety Research (2009)
Science and Decisions: Advancing Risk Assessment (2009)
Phthalates and Cumulative Risk Assessment: The Tasks Ahead (2008)
Estimating Mortality Risk Reduction and Economic Benefits from Controlling Ozone Air Pollution (2008)
Respiratory Diseases Research at NIOSH (2008)
Evaluating Research Efficiency in the U.S. Environmental Protection Agency (2008)
Hydrology, Ecology, and Fishes of the Klamath River Basin (2008)
Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment (2007)
Models in Environmental Regulatory Decision Making (2007)
Toxicity Testing in the Twenty-first Century: A Vision and a Strategy (2007)
Sediment Dredging at Superfund Megsites: Assessing the Effectiveness (2007)
Environmental Impacts of Wind-Energy Projects (2007)
Scientific Review of the Proposed Risk Assessment Bulletin from the Office of Management and Budget (2007)
Assessing the Human Health Risks of Trichloroethylene: Key Scientific Issues (2006)
New Source Review for Stationary Sources of Air Pollution (2006)
Human Biomonitoring for Environmental Chemicals (2006)
Health Risks from Dioxin and Related Compounds: Evaluation of the EPA Reassessment (2006)
Fluoride in Drinking Water: A Scientific Review of EPA's Standards (2006)
State and Federal Standards for Mobile-Source Emissions (2006)
Superfund and Mining Megsites—Lessons from the Coeur d'Alene River Basin (2005)
Health Implications of Perchlorate Ingestion (2005)
Air Quality Management in the United States (2004)
Endangered and Threatened Species of the Platte River (2004)

Atlantic Salmon in Maine (2004)
Endangered and Threatened Fishes in the Klamath River Basin (2004)
Cumulative Environmental Effects of Alaska North Slope Oil and Gas Development (2003)
Estimating the Public Health Benefits of Proposed Air Pollution Regulations (2002)
Biosolids Applied to Land: Advancing Standards and Practices (2002)
The Airliner Cabin Environment and Health of Passengers and Crew (2002)
Arsenic in Drinking Water: 2001 Update (2001)
Evaluating Vehicle Emissions Inspection and Maintenance Programs (2001)
Compensating for Wetland Losses Under the Clean Water Act (2001)
A Risk-Management Strategy for PCB-Contaminated Sediments (2001)
Acute Exposure Guideline Levels for Selected Airborne Chemicals (nineteen volumes, 2000-2015)
Toxicological Effects of Methylmercury (2000)
Strengthening Science at the U.S. Environmental Protection Agency (2000)
Scientific Frontiers in Developmental Toxicology and Risk Assessment (2000)
Ecological Indicators for the Nation (2000)
Waste Incineration and Public Health (2000)
Hormonally Active Agents in the Environment (1999)
Research Priorities for Airborne Particulate Matter (four volumes, 1998-2004)
The National Research Council's Committee on Toxicology: The First 50 Years (1997)
Carcinogens and Anticarcinogens in the Human Diet (1996)
Upstream: Salmon and Society in the Pacific Northwest (1996)
Science and the Endangered Species Act (1995)
Wetlands: Characteristics and Boundaries (1995)
Biologic Markers (five volumes, 1989-1995)
Science and Judgment in Risk Assessment (1994)
Pesticides in the Diets of Infants and Children (1993)
Dolphins and the Tuna Industry (1992)
Science and the National Parks (1992)
Human Exposure Assessment for Airborne Pollutants (1991)
Rethinking the Ozone Problem in Urban and Regional Air Pollution (1991)
Decline of the Sea Turtles (1990)

*Copies of these reports may be ordered from the National Academies Press
(800) 624-6242 or (202) 334-3313
www.nap.edu*

BOARD ON LIFE SCIENCES

Members

JAMES P. COLLINS (*Chair*), Arizona State University, Tempe, AZ
ENRIQUETA C. BOND, Burroughs Wellcome Fund, Marshall, VA
ROGER D. CONE, Vanderbilt University Medical Center, Nashville, TN
JOSEPH R. ECKER, Salk Institute for Biological Studies, LaJolla, CA
SEAN EDDY, HHMI Janelia Farm Research Campus, Ashburn, VA
SARAH C.R. ELGIN, Washington University, St. Louis, MO
DAVID R. FRANZ, Former Cdr USAMRIID, Consultant, Frederick, MD
STEPHEN FRIEND, Sage Bionetworks, Seattle, WA
ELIZABETH HEITMAN, Vanderbilt University Medical Center, Nashville, TN
RICHARD A. JOHNSON, Global Helix LLC, Washington, DC
JUDITH KIMBLE, University of Wisconsin, Madison, WI
MARY E. MAXON, Science Philanthropy Alliance, Palo Alto, CA
KAREN E. NELSON, J. Craig Venter Institute, Rockville, MD
ROBERT M. NEREM, Georgia Institute of Technology, Atlanta, GA
MARY E. POWER, University of California, Berkeley, CA
MARGARET RILEY, University of Massachusetts, Amherst, MA
LANA SKIRBOLL, Sanofi, Washington, DC
JANIS C. WEEKS, University of Oregon, Eugene, OR
MARY WOOLLEY, Research!America, Alexandria, VA

Staff

FRANCES E. SHARPLES, Director
JO L. HUSBANDS, Scholar/Senior Project Director
JAY B. LABOV, Senior Scientist/Program Director for Biology Education
LIDA ANESTIDOU, Senior Program Officer, ILAR
KATHERINE W. BOWMAN, Senior Program Officer
MARILEE K. SHELTON-DAVENPORT, Senior Program Officer
KEEGAN SAWYER, Program Officer
AUDREY THEVENON, Associate Program Officer
BETHELHEM M. MEKASHA, Financial Associate
ANGELA KOLESNIKOVA, Administrative Assistant
VANESSA LESTER, Research Associate
KANOKO MAEDA, Senior Program Assistant
JENNA OGILVIE, Senior Program Assistant

Preface

The US Department of Defense (DOD) is faced with an overwhelming task in evaluating chemicals that could potentially pose a threat to its deployed personnel. There are over 84,000 registered chemicals, and testing them with traditional toxicity-testing methods is not feasible in terms of time or money. In recent years, there has been a concerted effort to develop new approaches to toxicity testing that incorporate advances in systems biology, toxicogenomics, bioinformatics, and computational toxicology. Given the advances, DOD asked the National Research Council (NRC) to determine how DOD could use modern approaches for predicting chemical toxicity in its efforts to prevent debilitating, acute exposures to deployed personnel.

In this report, the Committee on Predictive-Toxicology Approaches for Military Assessments of Acute Exposures provides an overall conceptual approach that DOD could use to develop a predictive-toxicology system. It reviews the current state of computational and high-throughput approaches for predicting acute toxicity and suggests methods for integrating data and predictions. It concludes with lessons learned from current high-throughput screening programs and suggests some initial steps for DOD investment.

This report has been reviewed in draft form by persons chosen for their diverse perspectives and technical expertise in accordance with procedures approved by the NRC Report Review Committee. The purpose of the independent review is to provide candid and critical comments that will assist the institution in making its published report as sound as possible and to ensure that the report meets institutional standards of objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process. We thank the following for their review of this report: Ellen Berg, BioSeek, Inc.; David Clapham, Harvard University; Mark Cronin, Liverpool John Moores University; Yvonne Dragan, DuPont; John Jenner, Defence Science and Technology Laboratory; Charles Santerre, Purdue University; Rusty Thomas, US Environmental Protection Agency; Ken Turteltaub, Lawrence Livermore National Laboratory; Daniel Wilson, The Dow Chemical Company; and Menghang Xia, National Center for Advancing Translational Sciences.

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations, nor did they see the final draft of the report before its release. The review of the report was overseen by the review coordinator, David Eaton, University of Washington, and the review monitor, Mark Cullen, Stanford University. Appointed by the NRC, they were responsible for making certain that an independent examination of the report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of the report rests entirely with the committee and the institution.

The committee gratefully acknowledges the following for their presentations to the committee during open sessions: Alison Director-Myska, Defense Threat Reduction Agency, and Keith Houck, US Environmental Protection Agency.

The committee is grateful for the assistance of the National Research Council staff in preparing this report. Staff members who contributed to the effort are Ellen Mantus, project director; Marilee Shelton-Davenport, senior program officer; Keri Stoeber, research associate; James Reisa,

director of the Board on Environmental Studies and Toxicology; Norman Grossblatt, senior editor; Mirsada Karalic-Loncarevic, manager of the Technical Information Center; Radiah Rose-Crawford, manager of editorial projects; and Ivory Clarke, senior program assistant.

I especially thank the members of the committee for their efforts throughout the development of this report.

David Dorman, *Chair*
Committee on Predictive-Toxicology Approaches
for Military Assessments of Acute Exposures

Contents

SUMMARY	3
1 INTRODUCTION	9
Study Background, 9	
The Committee and Its Task, 10	
The Committee's Approach to Its Task, 10	
Organization of This Report, 11	
References, 12	
2 CONCEPTUAL FRAMEWORK AND PRIORITIZATION STRATEGY	13
Acute Toxicity of Classical Chemical-Warfare Agents, 13	
Predicting Acute Toxicity of Potential Chemical-Warfare Agents, 14	
A Framework and Strategy for Predicting Acute Toxicity of Potential Chemical-Warfare Agents, 15	
Findings and Recommendations, 24	
References, 24	
3 NONTESTING APPROACHES RELEVANT TO PREDICTION OF ACUTE TOXICITY AND POTENCY	29
Initial Chemical Characterization, 29	
Use of Physicochemical Properties to Predict Physical Hazards, Chemical Reactivity, and Pharmacokinetics, 31	
In Silico Approaches for Predicting Toxic Effects, 36	
Ensuring Scientific Confidence in (Q)SAR Models, 43	
Findings and Recommendations, 43	
References, 44	
4 ASSAYS FOR PREDICTING ACUTE TOXICITY	51
In Vitro Assays, 52	
Nonmammalian In Vivo Animal Models, 60	
Emerging Technologies, 63	
Metabolic Considerations, 65	
Assay Considerations for Improving Prediction of Acute Toxicity, 67	
Findings and Recommendations, 70	
References, 71	
5 INTEGRATION AND DECISION-MAKING FOR PREDICTIVE TOXICOLOGY	80
General Approach to Integration and Decision-Making, 80	
Approaches to Integration, 83	
Approaches to Categorization, 90	
Findings, 92	
References, 93	

6	LESSONS LEARNED AND NEXT STEPS	96
	Modern Approaches for the Assessment of Acute Chemical Toxicity, 97	
	Implementation of the Committee’s Conceptual Model for Assessment of Acute Chemical Toxicity, 98	
	Development of a Modern Tiered Approach for Predicting Acute Toxicity: Initial Steps, 99	
	Other Considerations, 106	
	Findings and Recommendations, 107	
	References, 108	

APPENDIXES

A	BIOGRAPHICAL INFORMATION ON THE COMMITTEE ON PREDICTIVE-TOXICOLOGY APPROACHES FOR MILITARY ASSESSMENTS OF ACUTE EXPOSURES	113
B	AVAILABLE DATA OR DATABASES	116

BOXES, FIGURES, AND TABLES

BOXES

1-1	Statement of Task, 11
2-1	Definitions of Relevant Terms, 14
2-2	Tiered Approach to Predicting Toxicity, 23
3-1	Definitions of Selected Nontesting Approaches, 30
3-2	Primary Data Considered During a Preliminary Characterization of a Chemical of Interest, 31
3-3	In Silico Approaches for Predicting Physicochemical Properties, 32
4-1	ACuteTox Testing Strategy, 68
5-1	Simplified Illustration of Integration and Decision-making, 81
5-2	Example of Meta-Analytic Approach That Uses Irreproducible Discovery Rates to Integrate Categorization Decisions, 85
6-1	An Example of an Integrated Testing Strategy for Predicting Drug-Induced Liver Injury, 98
6-2	The Use of Computational Approaches for Evaluating Chemical-Induced Inhibition of Acetylcholinesterase, 101
6-3	The Use of HTS Assays for Identifying Endocrine Disrupting Potential, 102
6-4	The Use of the BALB/3T3 Neutral-Red Uptake Cytotoxicity Assay to Predict Acute Toxicity, 104
6-5	The Use of HTS Assays to Evaluate Inhibition of Acetylcholinesterase, 106

FIGURES

S-1	Conceptual framework and examples of databases, assays, models and tools for predicting acute chemical toxicity, 4
S-2	Prioritization strategy based on a tiered approach for using predictive-toxicology models and tools to evaluate agents for acute toxicity, 5
2-1	Conceptual framework and examples of databases, assays, models and tools for predicting acute chemical toxicity, 20
2-2	Prioritization strategy based on a tiered approach for using predictive-toxicology models and tools to evaluate agents for acute toxicity, 22
3-1	Conceptual framework for the future development of (Q)SARs, 42
3-2	Key elements associated with evaluating the adequacy of a (Q)SAR model and its prediction as adapted from the REACH technical guidance, 44
4-1	Intended target families and subfamilies for the ToxCast program, 53

- 5-1 Illustration of a general approach to integration and decision-making for applying predictive approaches to acute, debilitating toxicity, 82
- 5-2 Approaches for integrating disparate data sets with LD₅₀ as an example, 84
- 5-3 ToxPi model for integration of acute-toxicity potential, 87
- 6-1 Prioritization strategy based on a tiered approach for using predictive-toxicology models and tools to evaluate agents for acute toxicity, 99

TABLES

- 2-1 Biological Processes and Cellular Targets Associated with Acute Toxicity in Humans or Laboratory Animals, 16
- 3-1 Examples of Heuristic Rules to Predict Oral Absorption, 33
- 3-2 Examples of (Q)SARs for Various Chemical Classes, 37
- 3-3 Examples of Models and Tools to Predict Acute Oral Toxicity, 38

Application of Modern Toxicology Approaches for Predicting Acute Toxicity for Chemical Defense

Summary

As part of its mission to provide military forces, the US Department of Defense (DOD) must anticipate, defend, and safeguard its personnel against chemical threats. Many factors can determine whether a chemical agent could pose a threat, and toxicity clearly is one of them. To assess toxicity, DOD has relied primarily on traditional toxicity testing in which adverse biological responses are measured in laboratory animals that are exposed to high doses of a test agent. The traditional approaches, however, are expensive and time-intensive, raise questions about the applicability of results to human populations, raise concerns about animal welfare, and are impractical for evaluating quickly large numbers of chemicals that could be used against deployed forces. In recent years, various agencies and organizations have attempted to incorporate advances in systems biology, toxicogenomics, bioinformatics, and computational toxicology to develop cost-effective approaches for predicting chemical toxicity. Given the recent advances and developments in toxicity-testing methods and approaches, DOD asked the National Research Council (NRC) to determine the feasibility of developing a toxicity-testing program that uses modern approaches to identify acutely toxic agents rapidly that are relevant to DOD.¹ In response to that request, the NRC convened the Committee on Predictive-Toxicology Approaches for Military Assessments of Acute Exposure, which prepared the present report.

CONCEPTUAL FRAMEWORK AND STRATEGY

As requested by DOD, the committee developed an overall conceptual approach that uses modern approaches for predicting acute, debilitating chemical toxicity. Its approach consisted of three components: (1) a *conceptual framework* that links chemical structure, physicochemical properties, biochemical properties, and biological activity to acute toxicity; (2) a suite of databases, assays, models, and tools that are based on modern *in vitro*, nonmammalian *in vivo*, and *in silico* approaches that are applicable for prediction of acute toxicity; and (3) a *tiered prioritization strategy* for using databases, assays, models, and tools to predict acute toxicity in a manner that balances the need for accuracy and timeliness. The committee based its conceptual framework (Figure S-1) on the premise that whole-animal toxicity can be predicted by using information about lower levels of complexity, even down to the level of chemical structure. Specifically, it is hypothesized that chemical structure, physicochemical properties, biochemical properties, or biological activity in isolated cells and tissues or in nonmammalian organisms can be used to predict acute mammalian toxicity.

The prioritization strategy was formulated on the basis of DOD's stated need to understand the relative threat of the growing list of registered chemical substances. Although the committee cannot prescribe exactly how to manage various policy tradeoffs, such as the tolerance for false negatives and the timeframe required for identifying important hazards, it recommends a tiered prioritization strategy (Figure S-2) that applies increasingly complex approaches to place chemicals into three categories: high confidence of low toxicity, high confidence of high toxicity, and uncertain toxicity because of data inadequacy. The first category allows some chemicals to be deselected on the basis of low acute toxicity, and the emphasis on high confidence indicates a low tolerance for false negatives. The second category allows chemicals to be "selected" on the basis of high acute toxicity, and the emphasis on high confidence indicates the need to focus rapidly on chemi-

¹The verbatim statement of task is provided in Chapter 1 of this report.

4 Application of Modern Toxicology Approaches for Predicting Acute Toxicity for Chemical Defense

icals that might pose a risk. The third category represents chemicals that would move to the next tier. Chemicals could be deselected at any stage by considering other factors, such as chemical availability and weaponizability, that could eliminate them from further consideration. As illustrated in the figure and discussed further in the sections that follow, the testing strategy proceeds through a number of tiers that are successively more predictive and resource-intensive, from initial characterization (Tier 0) to nontesting approaches (Tier 1) to high-throughput and medium-throughput assays (Tier 2) and ultimately to traditional animal testing (Tier 3). Progression through the tiers requires intermediate integration steps that consider the diversity of data both within a tier and across tiers. At each tier, DOD will need to develop policies that are relevant to its mission on how to assign chemicals to various categories and to determine the extent of end-point coverage that is adequate for it to make reliable decisions. The committee notes that an end point could be a clinical outcome or a molecular initiating event. If science advances in such a way that adverse outcome pathways of interest to DOD are known, the strategy shown in Figure S-2 could rely on nontesting and biological assay-based approaches that evaluate molecular initiating events or measurable key events in the pathways.

NONTESTING APPROACHES FOR PREDICTING ACUTE TOXICITY

The committee envisions that nontesting approaches will be an important component of its conceptual framework. Nontesting approaches range from grouping chemicals that are structurally similar to developing quantitative structure–activity relationship (QSAR) models. The underlying assumption of nontesting approaches is that chemical properties that determine how a chemical will interact with a defined biological system are inherent in its molecular structure and thus that structurally similar chemicals should have similar biological activity. The starting point in the application of any nontesting approach is to search for and evaluate information on the chemical of interest. That step constitutes Tier 0 in the committee’s proposed strategy (Figure S-2).

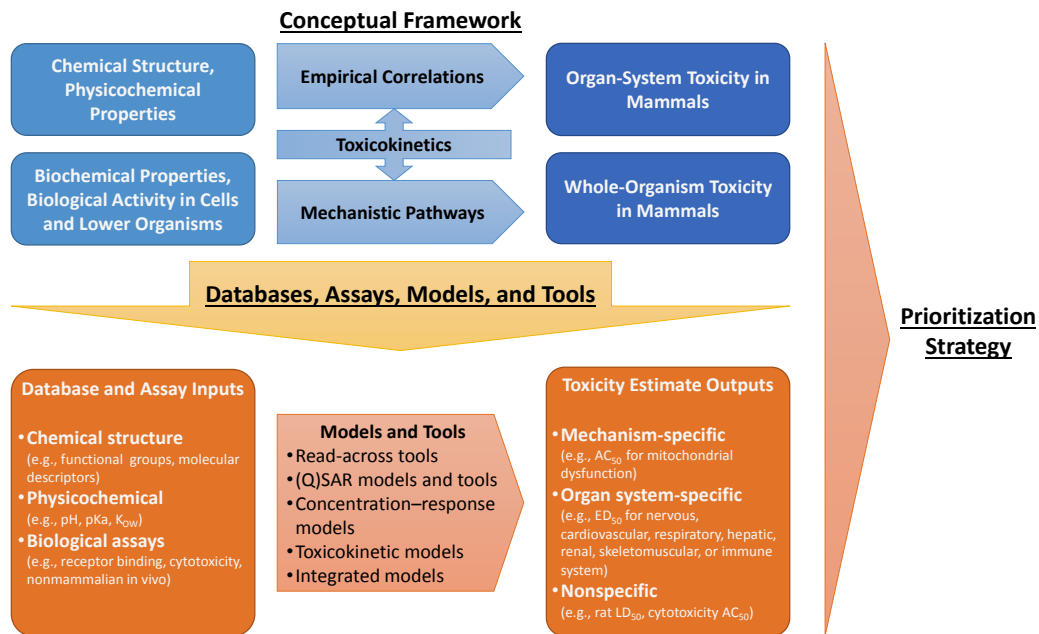


FIGURE S-1 Conceptual framework and examples of databases, assays, models, and tools for predicting acute chemical toxicity.

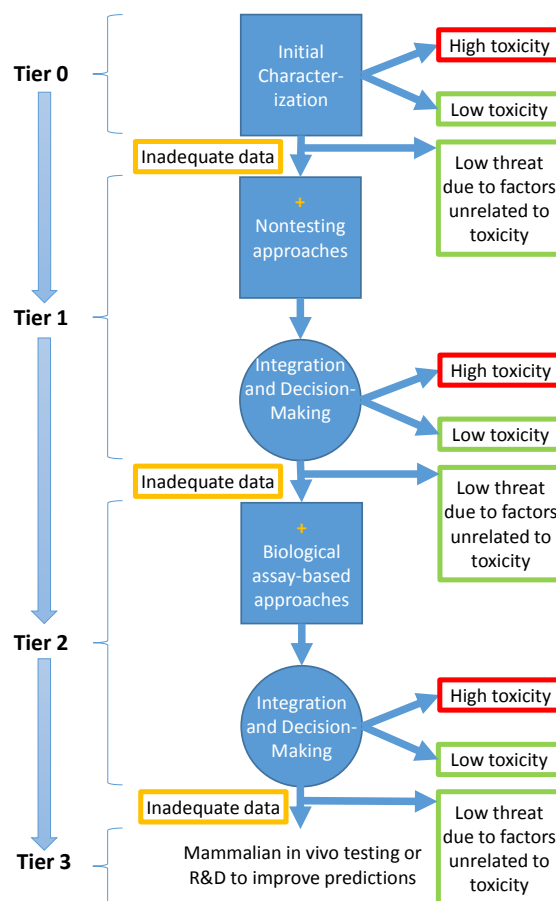


FIGURE S-2 Prioritization strategy based on a tiered approach for using predictive-toxicology models and tools to evaluate agents for toxicity.

As would be expected, information on physical properties, solvation properties, and molecular attributes (physicochemical data) is critical. Physicochemical data can be used to predict a chemical's physical hazard, its reactivity, and its pharmacokinetics, including absorption by different exposure routes, distribution in the body, and likely metabolites. Physicochemical data can be obtained from the literature, derived experimentally, or predicted with various *in silico* techniques. However, many tools that can be used to predict physicochemical properties have limited chemical applicability; that is, they are most applicable for small organic chemicals.

Nontesting approaches have been used to predict acute toxicity. Specifically, a few (Q)SAR models have been developed for predicting *in vivo* acute toxicity.² Most have focused on the prediction of acute rodent oral toxicity, such as estimation of oral LD₅₀ values;³ few attempts have been made to derive models for acute toxicity via other exposure routes, such as inhalation and dermal exposure. Nontesting approaches also have been used to predict toxicity end points, such as neurotoxicity or cytotoxicity. More recent efforts have investigated the integration of *in vitro* assay data with nontesting approaches to strengthen predictions. Key issues with nontesting (and all other) approaches are their relevance and applicability for the broad array of chemicals of interest to

²The committee uses the shorthand notation (Q)SAR to indicate both SAR and QSAR.

³An LD₅₀ value is the dose at which 50% of the population dies.

DOD and the reliability and validity of the data used to develop the models. Furthermore, the exposure routes of interest to DOD are most likely inhalation and dermal exposure, and few nontesting approaches address these exposure routes.

BIOLOGICAL ASSAYS FOR PREDICTING ACUTE TOXICITY

In vitro assays and nonmammalian in vivo assays are important components of the committee's conceptual framework. Numerous screening assays have been developed to measure specific biological activities. The various assay types are described below with some key limitations noted for DOD's purposes.

- *Specific-Protein Assays.* Many enzyme and receptor-binding assays have been developed to examine specific mechanisms of action at the molecular level. Some—such as ones that measure chemical-induced inhibition of acetylcholinesterase activity, altered electron transport in mitochondria, and modulation of ion-channel activity—might be relevant for predicting acute toxicity. Although the protein assays hold some promise, a key limitation is that acute toxicity that is not mediated by chemical action on specific enzymes or receptors will go undetected in these types of assays.

- *Cell-Based Phenotypic Assays.* These assays typically use cultured cells and measure some overall phenotypic output relevant to predicting acute toxicity, such as cellular proliferation, plasma membrane permeability, and adenosine triphosphate content. There is a growing literature on their application as toxicity screens, especially in drug development. Cell-based assays, particularly ones for evaluating cytotoxicity, have demonstrated success in predictive toxicology. A key limitation of cytotoxicity assays is that they do not provide data on some of the most important toxic mechanisms, specifically ones that involve organ-specific or cell-type-specific physiology. Another limitation of many existing cell-based assays is that they rely on immortalized cell lines that have little metabolic capability.

- *Organotypic Models.* Organotypic models more closely mimic the anatomy of organs and have been developed for the skin, eye, lung, liver, and central nervous system. They are especially attractive given their theoretical potential to model metabolism, biodistribution, and biological activity of a chemical in an in vitro system. However, the science of modeling human organs in a culture dish accurately, especially in formats suitable for high-throughput testing, and its application to toxicology are still in their infancy.

- *Nonmammalian in vivo Assays.* In addition to in vitro assays, the committee envisions nonmammalian animal models as a potentially important component of its conceptual framework. Traditional whole-animal assays have been crucial in understanding how chemicals affect metabolism and exhibit pathology at the cell and organ level. However, traditional assays are often expensive, require large amounts of chemicals, and cannot be adapted to even a medium-throughput format. For those and other reasons, alternative animal models have been developed. Ones that are potentially valuable for adapting to high-throughput screening rely on the fruit fly (*Drosophila melanogaster*), a nematode (*Caenorhabditis elegans*), and the zebrafish (*Danio rerio*). One particular advantage of the alternative models is the ability to identify whole-organism or organ-level responses. However, as with all animal models, a key limitation is related to species differences and use of resulting data to extrapolate to human responses. Furthermore, measuring some end points with alternative animal models has lower throughput than many in vitro assays, and little is known about their applicability to the assessment of acute toxicity of chemicals that are relevant to DOD.

In vitro assays, alternative animal models, and other emerging technologies described here and in more detail later in the committee's report hold promise, but some important limitations or considerations should be noted. First, in vitro assays for predicting acute toxicity have focused

primarily on nonmechanistic indicators of toxicity, such as cytotoxicity; they were not developed with a quantitative linkage to any phenotype (acute or chronic). Second, existing assays focus on oral exposure; there has been little consideration of dermal or inhalation exposure. Third, most current *in vitro* assays do not account for important pharmacokinetic characteristics, such as metabolism, that can influence *in vivo* toxicity. Fourth, the nominal chemical concentration used in the assays is not necessarily representative of the concentration at which chemical bioactivity is observed. Fifth, cellular systems commonly use immortalized cancer cell lines, which might fail to detect chemical activity or effects that might occur in normal (nontumor) differentiated cells. Sixth, cells can have different levels of activity or responsiveness, depending on whether they are primary cells, differentiated cells, or immortalized cells and on how many times they have been cultured, so assay reproducibility can be a problem. Seventh, interpreting activity or effective concentrations that result from a high-throughput screening assay can be difficult because activity at high concentrations could represent nonspecific effects and offer little information about specific bioactivity. Conversely, the absence of activity could mean that the tested concentration is below the *in vitro* effective concentration, that the assay does not represent the biological target, or that there are problems with assay reliability. Current efforts in high-throughput screening support the observations noted here, and the committee emphasizes that DOD should use the experience from current high-throughput screening programs to design its screening program to predict acute, debilitating toxicity.

INTEGRATION AND DECISION-MAKING FOR PREDICTIVE TOXICOLOGY

A robust integration and decision-making strategy is needed as part of the committee's suggested tiered prioritization strategy (shown in Figure S-2). As noted, the goal of each tier is to place a chemical into one of three categories: high confidence of high toxicity, high confidence of low toxicity, or inadequate data. That activity will require integrating various data streams and predictions that inform a single acute-toxicity end point ("within-end-point" integration and decision-making) and integrating predictions from several acute-toxicity end points ("cross-end-point" integration and decision-making). The committee's report discusses various methods for integrating data and predictions. Key tasks for DOD will be to define the most informative end points for its purpose (for example, neurotoxicity vs seizures), to set boundaries or toxicity thresholds for what is considered "high" or "low" toxicity for each end point, and to specify the level of confidence needed to make determinations.

One simple approach for integrating multiple end points is to summarize the categorization results for each end point in a "scorecard." Each end point would be evaluated as to whether the chemical exhibited "high toxicity," "low toxicity," or "inadequate data." A chemical would then be assigned to a "high toxicity overall" bin if at least one of the end points scored as "high toxicity," a "low toxicity overall" bin only if *all the end points* scored as "low toxicity," and an "inadequate data overall" bin if neither of the first two conditions is met. That simple approach has the advantage of retaining the end-point-specific information to inform future data generation. It is also consistent with a low tolerance for false negatives in that each end point serves as sufficient evidence to assign a chemical to a "high toxicity overall" bin.

It is possible to use more complex recombination approaches that would not depend strictly on a simple decision rule related to the categories for each end point. For example, one approach would be to provide a summary measure that consisted of a weighted sum of individual toxicity end points. Even if each individual end point is rated as "inadequate data," it is conceivable that the presence of multiple end points close to their corresponding toxicity thresholds would permit a chemical to be categorized as "high" or "low" on the basis of the summary measure. Setting up appropriate decision rules would be a key policy question for DOD if it chose to go forward with implementing the committee's suggested approach for predicting acute, debilitating toxicity.

LESSONS LEARNED AND NEXT STEPS

Several large-scale initiatives have been evaluating *in vitro* testing methods for their ability to predict human toxicity, and the committee considered them as it debated the feasibility of a predictive testing program for DOD. The US Environmental Protection Agency (EPA) ToxCast program and the European ACuteTox program demonstrate that *in vitro* assays have some value for predicting acute toxicity and provide evidence that an *in vitro* screening approach is feasible for evaluating the relative threat of a chemical as an acute hazard. However, most of the assays developed and validated for high-throughput screening programs were not developed specifically for acute-toxicity testing and so might be of little use for identifying chemicals that have the potential to cause acute, debilitating injuries in deployed military personnel. Lessons learned from those programs, however, could provide a great deal of guidance to DOD in its designing a system that uses high-throughput screening and predictive models to evaluate acute toxicity.

On the basis of its review, the committee notes several initial steps that DOD could take to implement the tiered prioritization strategy. First, an investment by DOD in computational and high-throughput screening could yield benefits in characterizing the toxicity of chemicals on which there are few or no toxicity data. Computational methods for predicting acute toxicity are seeing steady growth, and high-throughput screening might prove useful in excluding chemicals that have low toxic potential and in identifying toxic chemicals of greater concern for further testing. Second, there are data to suggest that DOD could use simple cytotoxicity assays to identify chemicals that have low acute-toxicity potential and focus its attention on chemicals that are more toxic. Additional investment would be required to determine whether the assays are relevant for identifying highly toxic chemicals that could be used against deployed troops. Third, the development of targeted mechanistically based assays could provide DOD with a useful resource for understanding and predicting potential toxicity of chemicals; specifically, having explicit knowledge of the mechanisms of action that lead to acute systemic toxicity would be valuable in the design and validation of integrated prediction methods. Completing the steps described here might require DOD to use a variety of reference chemicals, including chemicals of concern, to benchmark the results. Moreover, completing these steps will be facilitated by selecting well-characterized chemicals that can be used to evaluate the predictiveness of DOD's *in vitro* assays and approaches against *in vivo* experimental results.

The committee anticipates that in the next 3–10 years any tiered testing approach will not be able to replace fully the need for targeted mammalian *in vivo* studies to confirm the toxicity of a chemical of interest. Indeed, the state of the science suggests that development of a predictive acute-toxicity program will require extensive DOD investment in computational modeling approaches, assay development, methods for extrapolation of *in vitro* results to *in vivo* conditions, and data-integration methods. To begin the investment, the committee recommends that DOD initiate pilot studies that evaluate chemical classes of highest concern with well-characterized reference chemicals. The pilot studies would allow DOD to develop the novel assays and tools needed to predict acute chemical toxicity efficiently and accurately and to evaluate the rate of false negatives and false positives. The pilot studies could also examine how generalizable the results of various assays and tools are from one chemical class⁴ to another. That research would allow DOD to begin to address the size of the chemical space needed to make predictions about unknown chemicals. The committee emphasizes that DOD could benefit from leveraging its efforts with other federal activities, such as EPA's ToxCast program. Such collaboration would allow DOD to complete pilot studies more rapidly and maximize the return on its investment.

⁴In this context, chemical class is used broadly to include structurally related chemicals, chemicals that have different mechanisms of action, and chemicals that have different toxic end points, such as hepatotoxicity and neurotoxicity.

1

Introduction

The mission of the Department of Defense (DOD) is “to provide the military forces needed to deter war and to protect the security of our country” (DOD 2014). In support of that mission, DOD must protect the health and capabilities of its personnel—many of whom are deployed overseas—by anticipating and safeguarding against chemical and biological threats. Although many factors, such as availability and dissemination potential, need to be considered in evaluating a potential threat, chemical toxicity is critical in determining whether an agent could pose a threat if used by an adversary. Given the numbers of registered chemicals and new chemicals registered each year, evaluating chemical toxicity is especially daunting, particularly in terms of time and money, if one uses traditional toxicity-testing methods. In light of recent advances in toxicity-testing methods and approaches, DOD would like to determine the feasibility of developing a high-throughput predictive system that could rapidly identify acutely toxic agents and threat potentials. Accordingly, DOD asked the National Research Council (NRC) to determine how DOD could use modern approaches for predicting chemical toxicity in its efforts to prevent debilitating acute exposures of deployed personnel. In response to that request, NRC convened the Committee on Predictive-Toxicology Approaches for Military Assessments of Acute Exposures, which prepared this report.

STUDY BACKGROUND

Toxicity testing reached a turning point in 2007 with the release of the NRC report *Toxicity Testing in the 21st Century: A Vision and a Strategy*. The report set forth a vision for transforming traditional toxicity testing by incorporating advances in systems biology, epigenetics, toxicogenomics, bioinformatics, and computational toxicology. The new system that was described in the report would be based primarily on in vitro methods that can be used to evaluate changes in biological processes with cells, cell lines, or cellular components, preferably of human origin. The motivation for the new system was to accomplish four important goals: “(1) to provide broad coverage of chemicals, chemical mixtures, outcomes, and life stages, (2) to reduce the cost and time of testing, (3) to use fewer animals and cause minimal suffering in the animals used, and (4) to develop a more robust scientific basis for assessing health effects of environmental agents” (NRC 2007).

On release of the NRC report, several federal agencies embraced the proposed vision. A collaboration that has been informally referred to as Tox21 was formed between the National Toxicology Program of the National Institute of Environmental Health Sciences, the National Center for Computational Toxicology of the US Environmental Protection Agency (EPA), and the Chemical Genomics Center¹ of the National Institutes of Health; the US Food and Drug Administration joined the collaboration later. The goal of the collaboration has been to advance the

¹The Chemical Genomics Center is now part of the National Center for Advancing Translational Sciences.

vision proposed in the NRC report. EPA launched ToxCast as a separate activity with the goal of developing cost-effective approaches that use high-throughput technologies to predict chemical toxicity. The European Registration, Evaluation, Authorization, and Restriction of Chemicals (REACH) regulation encourages companies and other organizations to develop alternative methods that would substitute for traditional methods and has ultimately led to various research initiatives. All those programs and efforts have led to development of new methods and assays for predicting toxicity.

To protect the armed forces and their ability to serve, DOD's Defense Threat Reduction Agency conducts and sponsors scientific research to predict which chemicals might be used by adversaries as weapons and the toxicity that could occur if such agents were used. To understand potential toxicity at various doses, DOD has largely used *in vivo* toxicity testing in laboratory animals. However, that approach is time-consuming and expensive, must consider species differences in response, and does not enable DOD to keep up with the pace of new chemical registration. To address the challenge of elucidating the toxicity of more chemicals than can be practically tested in whole-animal assays and to address concerns raised with animal testing, DOD asked the NRC to consider the question of whether the new predictive-toxicology approaches being developed could be used to expedite its evaluation of potential chemical hazard. Specifically, are the new assays and approaches relevant to DOD's interest in acute toxicity? If not, is there research that would enable DOD to use predictive-toxicology approaches to identify acute chemical threats?

THE COMMITTEE AND ITS TASK

The committee that was convened as a result of DOD's request included experts in toxicology, computational methods, high-throughput approaches, -omics, physiologically based pharmacokinetic modeling, statistics, model validation, and emergency preparedness (see Appendix A for the committee's biographical information). As noted, the committee was asked to consider the new predictive-toxicology approaches that have been developed in other fields and to determine whether they could be used to meet DOD's needs. The committee's verbatim statement of task is provided in Box 1-1.

THE COMMITTEE'S APPROACH TO ITS TASK

To address its task, the committee held four meetings. In an open session during the first meeting, the committee heard presentations from the sponsor on its activities. On the basis of those discussions and the statement of task, the committee focused its attention on approaches that were considered to be most relevant for predicting acute debilitating² or life-threatening³ effects and on the organ systems that were deemed most likely to be affected. The organ systems of highest concern to DOD included the cardiovascular, respiratory, hepatic, renal, skeletomuscular, immune, and nervous systems, including special senses (vision and hearing). Each organ system was considered by the committee in its deliberations (see Chapter 2, Table 2-1 for further discussion). During the course of its review, the committee sought representative examples that could illustrate nontesting and assay-based approaches to assess acute chemical toxicity; the examples are provided throughout this report. On the basis of its task, the committee excluded from consideration traditional toxicity-testing assays (*in vivo* rodent assays).

²Acute debilitating effects are defined as ones that cause major irreversible morbidity, such as blindness, loss of limb function, paralysis, and severe hypoxia.

³A life-threatening effect is a disease or condition that makes the likelihood of death high unless the exposure is interrupted.

BOX 1-1 Statement of Task

An ad hoc committee under the auspices of the National Research Council will consider how the Department of Defense (DOD) could use modern approaches for predicting chemical toxicity in its efforts to prevent debilitating acute exposures to deployed personnel.

DOD needs to understand the relative threat of the increasingly long list of registered chemical substances, particularly in terms of potential acute hazard. To help DOD achieve its goal to protect its deployed personnel, this study will consider modern approaches for predicting toxicity and suggest an overall conceptual approach for using such information to evaluate acute hazards. The committee will consider the information provided by predictive-toxicology approaches that is increasingly being generated and used in the environmental health and pharmaceutical sectors to enhance or replace information from traditional, empirical testing of chemical safety in animals. The committee will focus on the assays and approaches that are being developed by the United States and European agencies (for example, for the Registration, Evaluation, Authorization, and Restriction of Chemicals (REACH) program, the EPA ToxCast effort, and the NIH/EPA/FDA Tox21 program); these might include computational modeling, structure-activity relationship analysis, analysis of physicochemical characteristics, read-across techniques, and high-throughput screening and other *in vitro* assays. Specifically, the committee will discuss the ability of these approaches to predict acute toxicity at levels relevant to DOD concerns.

In Phase 1 of this study, the committee will comment on the robustness and the relevance of the current approaches to meet DOD's needs. If the approaches being developed by other agencies do not address DOD's concerns about acute toxicity, the committee will broadly describe areas of research that could fill the gaps within the next 5 or 10 years. A second phase of the study, undertaken at the sponsor's request, will provide more detailed recommendations for a research roadmap.

In response to its task to “predict acute toxicity at levels relevant to DOD concerns,” the committee focused its approach on hazard identification, specifically identifying target organ systems and developing toxicity estimates, such as potency estimates. An approach for predicting acute toxicity that involved converting toxicity estimates to human exposure estimates, as has been taken with some chemical-warfare agents (Mioduszewski et al. 2002), was considered beyond the committee's charge. Furthermore, the committee interpreted DOD's stated interest in understanding the relative threat of chemicals that could be used by an adversary against deployed US military personnel to mean prioritizing chemicals in terms of their potential to cause acute toxicity. Thus, the committee was not focused on predicting human clinical signs or identifying at-risk populations. And, the committee did not set bounds for its proposed strategy because it recognized the need for DOD to develop policies to set toxicity thresholds relevant to its mission.

During the open session of its first meeting, the committee received a presentation from EPA on the ToxCast program. The committee considered the efforts of that program that were relevant to predicting acute toxicity and, more broadly, the technical approaches that might inform development of a DOD acute-toxicity program. A detailed review of the ToxCast program and its associated assays and methods was considered beyond the scope of the present report.

ORGANIZATION OF THIS REPORT

The committee's report is organized into six chapters and two appendixes. Chapter 2 describes a conceptual framework and components that would be needed to build an approach based on modern predictive-toxicology methods. Chapter 3 describes the use of nontesting approaches, including quantitative structure–activity relationships, to predict acute chemical toxicity. Chapter 4 provides a brief review of medium-throughput and high-throughput assays that

12 Application of Modern Toxicology Approaches for Predicting Acute Toxicity for Chemical Defense

can be used to predict acute mammalian toxicity. Chapter 5 addresses integration of the biological and chemical data into toxicity predictions. Chapter 6 presents important lessons learned from previous predictive acute-toxicity efforts and the committee's overall conclusions. The committee also identifies several steps that DOD could begin to take toward developing high-throughput assays and computational approaches to identify chemicals that have the potential to induce life-threatening acute toxicity in deployed personnel. Appendix A contains biographical information on the committee, and Appendix B discusses available toxicity data or databases that one could use to find toxicity data.

REFERENCES

- DOD (US Department of Defense). 2014. About the Department of Defense. [online]. Available: <http://www.defense.gov/about/>[accessed March 12, 2015].
- Mioduszewski, R., J. Manthei, R. Way, D. Burnett, B. Gaviola, W. Muse, S. Thomson, D. Sommerville, and R. Crosier. 2002. Interaction of exposure concentration and duration in determining acute toxic effects of sarin vapor in rats. *Toxicol. Sci.* 66(2):176-184.
- NRC (National Research Council). 2007. *Toxicity Testing in the 21st Century: A Vision and a Strategy*. Washington, DC: National Academies Press.

2

Conceptual Framework and Prioritization Strategy

As discussed in Chapter 1, the committee was asked to consider modern approaches for predicting acute, debilitating chemical toxicity and to suggest an overall conceptual approach that uses emerging science to evaluate acute hazards to deployed military personnel. This chapter first discusses current and future needs for toxicity evaluations of chemical-warfare agents, recognizing the increasing number and types of chemicals that are potentially available to adversaries. It then describes the conceptual framework and strategy developed by the committee for systematically applying modern approaches to the prediction of acute toxicity. The overall approach, which is illustrated in Figures 2-1 and 2-2, consists of three components (relevant terms are defined in Box 2-1):

- A *conceptual framework* that links chemical structure, physicochemical properties, biochemical properties, and biological activity to acute toxicity.
- A suite of *databases, assays, models, and tools* that are based on modern in vitro, non-mammalian in vivo, and in silico approaches applicable to predicting acute toxicity.
- A *tiered prioritization strategy* for using databases, assays, models, and tools to predict acute toxicity in a manner that balances the need for accuracy and timeliness.

Later chapters in this report provide details of the types of databases, assays, models, and tools that are available for evaluating acute toxicity, their integration, and next steps that are needed to begin implementing the committee's framework and strategy.

ACUTE TOXICITY OF CLASSICAL CHEMICAL-WARFARE AGENTS

Historically, most chemical-warfare agents have belonged to the following chemical classes: nerve agents (such as sarin and soman), blister or vesicant agents (such as phosgene oxime and sulfur mustards), blood agents (such as cyanide), and pulmonary agents (such as chlorine and phosgene) (DHHS 2014). Those agents have been well studied, and a detailed mechanistic understanding that is based on human data is available for some. For example, the organophosphorus (OP) nerve agents are potent inhibitors of acetylcholinesterase and result in acute cholinergic effects that occur minutes or hours after exposure. Knowledge of their mechanisms of toxicity can be useful in the development of therapeutic countermeasures (Sharma et al. 2015) and in the development of in vitro tests. For example, in vitro methods have been developed for the evaluation of cholinesterase inhibition by nerve-gas agents (Worek et al. 2007). Data on some early, sensitive responses to such agents can support development of acute exposure limits for the general public. For example, a number of studies indicate that pupil constriction (miosis) is the most sensitive acute response to human exposure to OP nerve agents (such as sarin), and such end points have been used as part of the basis of acute exposure limits (NRC 2005).

BOX 2-1 Definitions of Relevant Terms

An *assay* is a laboratory system designed to measure a physical, chemical, or biological end point.

A *model* is a quantitative or qualitative representation of a hypothesis that attempts to explain how different observations are related to one another. In the context of this report, the hypothesis typically concerns how physical, chemical, or biological data (“inputs”) can be used to predict biological outcomes of a given exposure (“outputs”) in a whole animal or human qualitatively or quantitatively.

A *tool* is an application of a model or set of models, such as in a software package, designed to be routinely used in an applied setting as opposed to a research or development setting.

In vitro approaches include high-throughput screening, other *in vitro* assays, and more complex systems, such as organotypic cell cultures.

Nonmammalian in vivo approaches include fish, amphibian, nematode, and insect models.

In silico approaches include computational modeling, structure–activity relationship analysis, analysis of physicochemical characteristics, and read-across techniques (see Chapter 3).

In vivo testing approaches have been developed and applied to assess the toxicity of chemical-warfare agents. The vast majority of available toxicity information has come from traditional toxicity studies in which adverse biological responses were measured in laboratory animals that were exposed to high doses of a test agent. The acute-toxicity data are often used to provide estimates of the amount of an agent that would be required to kill 50% of a population of test animals, such as a lethal dose 50% (LD₅₀) or a lethal concentration 50% (LC₅₀). In addition, pharmacokinetic studies that were designed to identify species differences in chemical absorption, distribution, metabolism, and excretion (Tenberken et al. 2010; Benson et al. 2011a,b) and specialized pharmacokinetic models (such as ones that use a human or porcine skin flap) that were developed to evaluate absorption and toxicity of some chemical-warfare agents (Riviere et al. 1995; Monteiro-Riviere and Inman 1997; Vallet et al. 2008) have been undergoing incremental refinement since their inception. A limitation of the *in vivo* studies, however, is that they tend to be low-throughput, require consideration of species differences in response, and often provide little insight into a chemical’s mechanism of action.

PREDICTING ACUTE TOXICITY OF POTENTIAL CHEMICAL-WARFARE AGENTS

Only a few chemicals have been formally classified as chemical-warfare agents. However, the list of chemicals that could potentially be used by an adversary against deployed US personnel is large and continues to grow as more chemicals enter the marketplace. Therefore, Department of Defense (DOD) efforts to evaluate potential chemical-warfare agents need to consider a wide array of chemicals beyond traditional chemical-warfare agents, including toxins of biological origin (such as trichothecenes, saxitoxin, and tetrodotoxin), industrial chemicals (such as ammonia), pesticides (such as sodium monofluoroacetate), and pharmaceutical agents (such as cocaine and amphetamine) (Holstege et al. 2007). The ability of an adversary to use those or other chemicals will depend on their or their precursors’ availability and weaponizability and on other factors that were deemed beyond the scope of the committee’s work but that might be important in deciding which agents to evaluate for acute toxicity.

To determine the best way to assess the growing list of registered chemical substances, the committee considered the adverse effects of highly toxic agents, including those of classical chemical-warfare agents, and identified the following organ systems to be of greatest importance for evaluating acute, debilitating hazards: cardiovascular, respiratory, hepatic, renal, skeletomuscular, immune, and nervous systems, including special senses (vision and hearing). Sufficient perturbation in those organ systems can lead to a progression in the severity of effects that can result in incapacitation or death of the whole organism.

Given ethical considerations, additional acute human-toxicity data are unlikely to be available except in cases of accidental release or deliberate attack for which exposure estimates are typically highly uncertain or unknown. And, available traditional toxicity-testing data provide little information about acute, debilitating toxicity. For example, information about chronic, reproductive, or developmental hazards—although important for chemical risk assessment in occupational or environmental settings—is of secondary concern in a military environment where acute, debilitating hazards are of immediate importance. As with other toxicity-testing programs, DOD recognizes that it would be prohibitively expensive and time-consuming to test all potential agents with traditional whole-animal toxicity-testing approaches even if such testing were limited to evaluations of acute toxicity. Moreover, traditional *in vivo* testing, particularly for acute toxicity, often does not provide information on the cellular or biological mechanisms of toxicity or in some cases even identify the target organ system.

Although some of the more modern, biological assay-based approaches have been used to elucidate mechanisms of action of many of the classical chemical-warfare agents described above, they have not been used to identify potential chemical-warfare agents. Nonetheless, the fact that some high-throughput screening data on chemical-warfare agents already exist suggests the feasibility of using such approaches to evaluate agents and provides important “reference” data with which results on other agents can be compared. The modern predictive approaches can also inform decisions as to whether additional mammalian *in vivo* testing of an agent is needed and might be able to provide information about the cellular and biological mechanistic events associated with acute toxicity and indicate whether additional testing should focus on a specific organ system or biological target.

A FRAMEWORK AND STRATEGY FOR PREDICTING ACUTE TOXICITY OF POTENTIAL CHEMICAL-WARFARE AGENTS

Conceptual Framework

A predictive-toxicology program to assess acute toxicity ideally will build on knowledge about the cellular targets and mechanisms of action that are related to acute human toxicity. Acute toxicity depends on fewer biological and chemical pathways than those envisioned by NRC (2007) for a general toxicity evaluation. It could be more straightforward, although still challenging, to predict the potential for acute toxicity than the potential for toxicity in the general public in a variety of organ systems, life stages, populations, and exposure timeframes. Specifically, clinical toxicologists have recognized several cellular or biological targets that are often associated with the acute lethal or debilitating effects of chemicals. Table 2-1 provides an overview of those cellular targets and relevant examples and lists some chemicals that affect the targets. It should be noted that there is not necessarily a one-to-one correspondence between mechanistic targets and organ-system targets because multiple mechanisms could affect a single organ system, a single mechanism could affect multiple organ systems, and debilitation or death could occur from multiorgan failure.

TABLE 2-1 Biological Processes and Cellular Targets Associated with Acute Toxicity in Humans or Laboratory Animals^{a, b}

Biological Process or Cellular Target	Example	Chemical or Biological Agent	Example Target Organ System	Examples of in vitro Assay Approaches ^c
Change in neurotransmitter function				
Altered axonal transport	Disruption of microtubule function	Vinca alkaloids β , β' -iminodipropionitrile	Nervous	Tubulin polymerization assessed with flow cytometry (Morrison and Hergenrother 2012)
Altered impulse conduction by axonal membrane	Blocking of Na ⁺ ion channel	Tetrodotoxin	Nervous	Cell-based assays of the membrane potential that use fluorescent dye (Hill et al. 2014)
Reduced precursor availability or neurotransmitter synthesis and storage	Inhibition of acetylcholine uptake into synaptic vesicle	Vesamicol Reserpine (dopamine)	Nervous	PC12 cell-based microelectrode assay (Cui et al. 2006; Chen et al. 2008)
Altered neurotransmitter release	Blocking of release of acetylcholine at neuromuscular junction	Botulinum toxin	Nervous	PC12 cell-based system for in vitro measurements of neurotransmitter release events (Yakushenko et al. 2013)
	Presynaptic release of acetylcholine and other neurotransmitters	α -latrotoxin		
Altered neurotransmitter binding at receptor sites	Neurotransmitter agonists	Opioids, benzodiazepines, nicotine, anatoxin-a, kainic acid	Nervous	Review of selected methods to assess receptor binding (Dunlop et al. 2007); use of stably transfected HEK cells expressing human D2, D3, or D4 dopamine receptors as a screening tool (Vangveravong et al. 2006; Xiao et al. 2014)
	Neurotransmitter antagonists	Curare, α -bungarotoxin, 3-quinuclidinyl benzilate		
Impaired neurotransmitter inactivation mechanisms	Acetylcholinesterase inhibition Altered dopamine transporter Altered serotonin reuptake Altered dopamine reuptake	Nerve gas agents Cocaine Fluoxetine Amphetamine	Nervous	Zebrafish-based (Jin et al. 2013) and enzyme-based (Wille et al. 2010) assays for acetylcholinesterase inhibitors
Altered ion flow				
Altered electrical conduction of heart or cardiomyocyte contractility	Sodium-potassium ATPase blockers	Digoxin	Cardiovascular	Assessment of altered cardiomyocyte contraction (Himmel 2013; Poinon et al. 2013, 2015; Sirenko et al. 2013; Scott et al. 2014) and electrophysiology (Lopez-Izquierdo et al. 2014); organotypic zebrafish heart model (Pieperhoff et al. 2014)
Altered ion pump (Na ⁺ , Ca ⁺⁺ , K ⁺) activity	Inhibit K ⁺ channel function Inhibit Na ⁺ channel function	Dendrotoxin, 4-aminopyridine Tetrodotoxin, saxitoxin	Cardiovascular	Comparison of in vitro potency of saxitoxin in cultured neurons with in vivo results (Jellett et al. 1992; Vale et al. 2008).

Increased permeability of cellular membranes				
Pore formation	Na ⁺ /H ⁺ antiporter	Ionophores	Cardiovascular	Assessment of cell permeability and other end points in multiple strains of mouse embryonic fibroblasts (Suzuki et al. 2014)
Ion-channel interactions	Transient receptor potential cation channel, subfamily A, member 1 (TRPA1) activation	Sulfur mustard Acrolein	Respiratory	Role of TRPA1 as a chemosensor (Büch et al. 2013; Stenger et al. in press)
Chemical reactivity	Acylation of proteins and lipids (pulmonary edema)	Phosgene	Respiratory	Human epithelial lung cells as a system to investigate pulmonary edema (Wijte et al. 2011)
Altered bioenergetics				
Mitochondrial dysfunction	Multiple mechanisms	Various	Multiple	Various HTS of mitochondrial dysfunction (Jensen and Rekling 2010; Sakamuru et al. 2012; Vongs et al. 2011; Attene-Ramos et al. 2013, 2015; Sirenko et al. 2014b; Wills et al. 2013)
Reduced ATP production	Inhibition of oxidative phosphorylation	Fluoroacetate, cyanide , chlordecone, bromethalin	Nervous, cardiovascular, multiple	Monitoring of ATP production or cell concentrations (Steinhoff et al. 2015)
Activation of apoptotic pathways	Multiple	Cisplatin, doxorubicin	Multiple	Cell-imaging methods for cultured cardiomyocytes (Mioulane et al. 2012)
Altered oxygen transport				
Competitive binding to hemoglobin	Carboxyhemoglobin production	Carbon monoxide	Multiple	In vitro assessment of carbon monoxide and cyanide binding to hemoglobin using human blood (Thoren et al. 2013)
Irritant or cytotoxic effects	Pulmonary edema	Phosgene, chlorine , methylisocyanate	Respiratory	Microfluidic system that mimics alveolar-capillary interface of human lung (Huh et al. 2012)
Oxidative stress or ROS formation				
Lipid peroxidation	Hepatic injury	Acetaminophen, carbon tetrachloride	Hepatic	Lipid peroxidation cell-based and cell-free assays (Kelesidis et al. 2014)
ROS formation	Renal injury	Aminoglycosides	Renal	HTS assays to measure ROS formation (Adams et al. 2013; Prasad et al. 2013; Zielonka et al. 2014)
Altered prostaglandin synthesis	Vascular dysfunction	NSAIDs	Cardiovascular	HTS assay for prostaglandin E synthase activity (Andersson et al. 2012)

(Continued)

TABLE 2-1 Continued

Biological Process or Cellular Target	Example	Chemical or Biological Agent	Example Target Organ System	Examples of in vitro Assay Approaches ^c
Damage to DNA and subcellular systems				
Genetic damage	Multiple	Multiple	Multiple	HTS assays to measure genetic damage (Gutzkow et al. 2013; Li et al. 2013; Wasalathanthri et al. 2013; Watson et al. 2014; Bandi et al. 2014; Falk et al. 2014; van der Linden et al. 2014)
DNA or protein adduct formation	DNA alkylation	Aflatoxins, cisplatinin (kidney), sulfur mustard	Multiple	Medium-throughput methods for quantification of sulfur mustard adducts to proteins (Andacht et al. 2014; Pantazides et al. 2015)
Altered protein synthesis	Inactivation of ribosomes	Ricin	Multiple	Assay for the measurement of adenine released from ribosomes or small stem-loop RNAs by ricin toxin A-chain catalysis (Sturm and Schramm 2009)
Disruption of cytoskeleton	Actin or cytoskeleton disassembly	Phalloidin, microcystin	Multiple	Assessment of cytoskeleton integrity in a hepatocyte model (Sirenko et al. 2014a)
Immune-mediated effects				
Immunogenic interactions with cell macromolecules	Alteration of mammalian immune system function	Endotoxin, anthrax exotoxins	Immune	Reviews of endotoxin and anthrax toxins (Thorn 2001; Liu et al. 2013).
Autoimmunity	Autoimmune hepatitis and necrotizing myositis	Statins	Multiple	Reviews of statins and myositis (Jones et al. 2014) and hepatotoxicity (deLemos et al. 2014)

^aThe lists of chemicals and biological targets shown here are not intended to be complete; rather, this table shows a variety of plausible biological targets and responses that need to be considered in evaluating chemicals for acute toxicity that could debilitate or kill deployed troops.

^b**Boldface:** Listed in Chemical Weapons Convention or is a suspected chemical agent of concern.

^cThere are relatively few applications of these methods to the prediction of acute toxicity; thus, the information is provided for illustrative purposes only to demonstrate the types of approaches used to date (see Chapter 4 for additional information).

Abbreviations: ATP, adenosine triphosphate; DNA, deoxyribonucleic acid; HTS, high-throughput screening; NSAID, nonsteroidal anti-inflammatory drug; PC, pheochromocytoma; RNA, ribonucleic acid; ROS, reactive oxygen species.

The relatively detailed knowledge of the multiple mechanisms by which chemicals can cause acute toxicity supports the basic premise of predictive toxicology that whole-animal toxicity can be predicted on the basis of information on lower levels of complexity down to the level of chemical structure. That premise forms the basis of the conceptual framework developed by the committee, illustrated in Figure 2-1. Specifically, it is hypothesized that chemical structure, physicochemical properties, biochemical properties, or biological activity in isolated cells and tissues or in non-mammalian organisms can predict acute mammalian toxicity. The predictions can arise through observations of empirical or statistical correlations or through knowledge of the relevant mechanistic pathways, either of which could potentially be coupled with toxicokinetic information.

Databases, Assays, Models, and Tools

Evaluating the potential for acute toxicity by using the conceptual framework of predictive toxicology requires a suite of databases, assays, models, and tools to cover the relevant physical, chemical, biological, and toxicological space. In general, “input” information on chemical structure, physicochemical properties, biochemical properties, and biological activity that is used to make predictions will be obtained from relevant databases or assays. Chemical-structure data might range from chemical-grouping data (for example, reaction chemistry domains, such as Michael acceptors) to quantitative descriptors of chemical structure (for example, topological descriptors and semiempirical quantum chemical descriptors). Physicochemical-property data include quantities measured in physical or chemical assays, such as boiling point, pH, pKa, and K_{ow} .¹ Biochemical measures are usually measures of specific molecular interactions (such as DNA binding and receptor activation) with biological molecules, such as nucleic acids, proteins (including enzymes and receptors), and lipids. Finally, biological activity might include both specific measures of function (such as acetylcholinesterase inhibition) and nonspecific measures of toxicity (such as cytotoxicity from *in vitro* assays and LC_{50} estimates obtained from assays that use *Drosophila*). Databases and assays for chemical structures and physicochemical properties are discussed in Chapter 3 and Appendix B, and assays for biochemical properties and biological activity in Chapter 4.

In this same context, the prediction “outputs” consist of estimates of end points related to acute toxicity. The end points might be related to particular mechanisms known to cause acute toxicity (see, for example, Table 2-1), end points related to specific organ system targets (noted above), or nonspecific end points, such as death and cytotoxicity. Data on those end points for chemicals of known toxicity (such as classical chemical-warfare agents) can serve as “training” and “test” data for building models or tools to predict the same end points for chemicals on which such data are lacking. Finally, the specific form of the outputs might be qualitative (such as active or inactive), semiquantitative (such as a ranking), or quantitative (such as a numerical estimate of dose). However, as discussed further below, quantitative estimates are likely to be of greatest use for military applications.

A variety of models and tools might be used to provide toxicity estimates. Models and tools might be qualitative (such as decision trees) or quantitative (such as statistical regression) and might include statistically based (or machine-learning-based) models, biologically based models, or a mixture of the two. No model or tool is universally applicable, so it is important that a model’s or tool’s domain of applicability is characterized in terms of the chemical space in which it is predictive and the relevant toxicological end points that are covered. Moreover, toxicokinetic models might need to be integrated into the predictions to address absorption, distribution, metabolism, and excretion relevant to acute toxicity. Finally, models and tools differ with respect to the uncertainty or confidence in their predicted outputs.

¹ K_{ow} is the octanol-water partition coefficient.

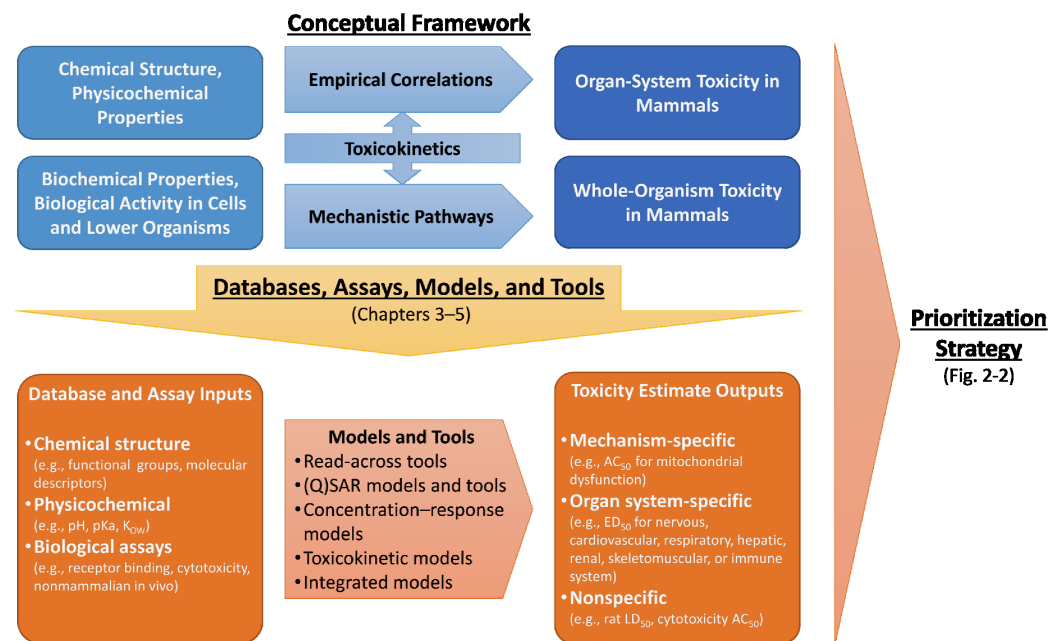


FIGURE 2-1 Conceptual framework and examples of databases, assays, models, and tools for predicting acute chemical toxicity.

As described further in Chapters 3-5, there are many available databases, assays, models, and tools that could be used to predict acute toxicity. Because they vary in their required level of effort, their relevance to acute toxicity, their domain of applicability, the extent to which they address toxicokinetics, and the uncertainty or confidence in their predictions, the committee developed an overall strategy for using them to evaluate acute toxicity. The committee's strategy is described next.

Prioritization Strategy for Evaluating Acute Toxicity

Effective implementation of predictive models and tools depends on first identifying the ultimate (and acceptable) use of the predictive outputs. The committee's task states that DOD needs to understand "the relative threat of the increasingly long list of registered chemical substances, particularly in terms of potential acute hazard." The committee interprets that statement to mean that the goal of the predictive-toxicology approach is to *prioritize* substances in the sense of identifying those of greater and less concern for acute toxicity. Three key issues must be considered in developing a strategy for prioritization: the need for quantitative measures of potency and their uncertainties, the need to minimize false negatives, and the need to screen a large number of chemicals rapidly.

The first key issue is that prioritization with respect to toxicity inherently requires a quantitative measure of potency and a characterization of uncertainty. Ideally, potency should be defined in absolute units, such as an acute oral LD_{50} in milligrams per kilogram per day. Relative potency measures might be informative if they include reference chemicals that have known toxicity and, in that case, could be converted to absolute potency measures if toxicokinetic information is also available to make any necessary adjustments. Qualitative outputs, such as binary categorizations of "active" or "inactive," might be useful as an additional output to target testing for specific end points but are not useful by themselves. Furthermore, in the absence of human data, there will always be inaccuracies in predicting human toxicity, so it is important to characterize the uncertainty

or confidence associated with any predicted potency value. Because a decision-maker might have defined tolerance for errors (such as for false negatives and false positives²), the degree of uncertainty or confidence in a prediction can influence the decision that is made about a particular substance. Therefore, an estimated confidence interval is essential to any prioritization strategy.

The topic of uncertainty leads to the second key issue: given that this task is meant to prevent death and debilitating injuries of US military personnel, it is expected that there will be a low tolerance for false negatives. A likely consequence of reducing the number of false negatives is that a higher percentage of chemicals will be retained for assessment with more accurate but more resource-intensive approaches. The overall time needed to complete the review for the whole chemical space would increase accordingly. However, the timeframe to complete an assessment of thousands of chemicals could be unacceptably long, and a chemical could be successfully weaponized in that timeframe before a decision has been made.

The timeframe raises the third key issue: the prioritization strategy needs to be able to screen chemicals in a manner that allows rapid identification of the ones that pose the greatest risk. A rapid-screening scenario could be acceptable if follow-up screening is conducted to ensure that all potential chemical threats are eventually identified. It is critical that such an approach incorporate a short timeline that progresses efficiently through a multitiered approach to allow timely reconsideration of chemicals that are not originally classified as posing the greatest risk. Lessons learned from the first round of screening could then be leveraged effectively in the reassessment and enable a more informed review and follow-up validation of the initial approach that can also be rapidly implemented. The risk of using this approach lies in a time lag that could result in weaponization of a chemical that was originally not deemed to pose a great threat.

The policy tradeoff of balancing a low tolerance for false negatives with a need to identify important hazards rapidly is beyond the scope of the committee's charge. However, as a general approach, the committee found that the policy tradeoff could be managed through a tiered prioritization approach as illustrated in Figure 2-2. Specifically, the committee's proposed prioritization strategy proceeds through a number of tiers that apply successively more predictive and resource-intensive approaches than the previous ones. At each tier, a chemical is placed into one of three general categories:

(a) *High confidence of low toxicity.* These chemicals would be deselected for further study and are considered to have a low relative acute toxicity. The requirement that the determination be made with high confidence addresses the low tolerance for false negatives.

(b) *High confidence of high toxicity.* These chemicals would be selected and considered to have a high relative acute toxicity. The requirement that the determination be made with high confidence focuses attention quickly on chemicals that might pose a high risk.

(c) *Uncertain toxicity due to inadequate data.* The remaining chemicals would be candidates for moving to the next tier of evaluation for acute toxicity.³ The uncertainties might stem from available predictions of high uncertainty or low confidence or from inadequate coverage of end points deemed important for evaluating acute toxicity. Depending on resource constraints, it might be reasonable to assess the chemicals by using additional factors unrelated to toxicity, such as weaponizability. Thus, some chemicals might be further deselected for further study because they pose a low threat owing to factors unrelated to toxicity (discussion of such factors is beyond the committee's charge). If additional evaluation of toxicity is determined to be needed, the chemical would be moved to the next tier of hazard evaluation to reduce uncertainty concerning the potential for acute, debilitating toxicity. Uncertainty might also be reduced through additional research into and development of approaches to improve acute-toxicity prediction, that is, by

²In this context, a "false negative" occurs when a chemical is identified as having low toxicity when it actually has high toxicity, and "false positive" occurs when a chemical is identified as having high toxicity when it actually has low toxicity.

³Some chemicals in categories (a) and (b) might be carried to a higher tier for validation purposes.

decreasing the number of chemicals in category (c) and increasing the ability to discriminate between categories (a) and (b).

The categorization can be based on a single end point (possibly based on multiple approaches) or on multiple end points. The committee notes that an end point could be a clinical outcome or a molecular initiating event (see Figure 3-1). If science advances in such a way that adverse-outcome pathways of interest to DOD are known, the strategy shown in Figure 2-2 could rely on nontesting and biological assay-based approaches that evaluate molecular initiating events or measurable key events in the pathways.

The committee broadly grouped the available approaches to predicting acute toxicity into four tiers, beginning with an initial chemical characterization (Tier 0), proceeding to nontesting approaches (Tier 1), then to biological assay-based approaches (Tier 2), which includes non-mammalian animal species, and ultimately to traditional whole-animal toxicity testing (Tier 3). The tiers are described further in Box 2-2.

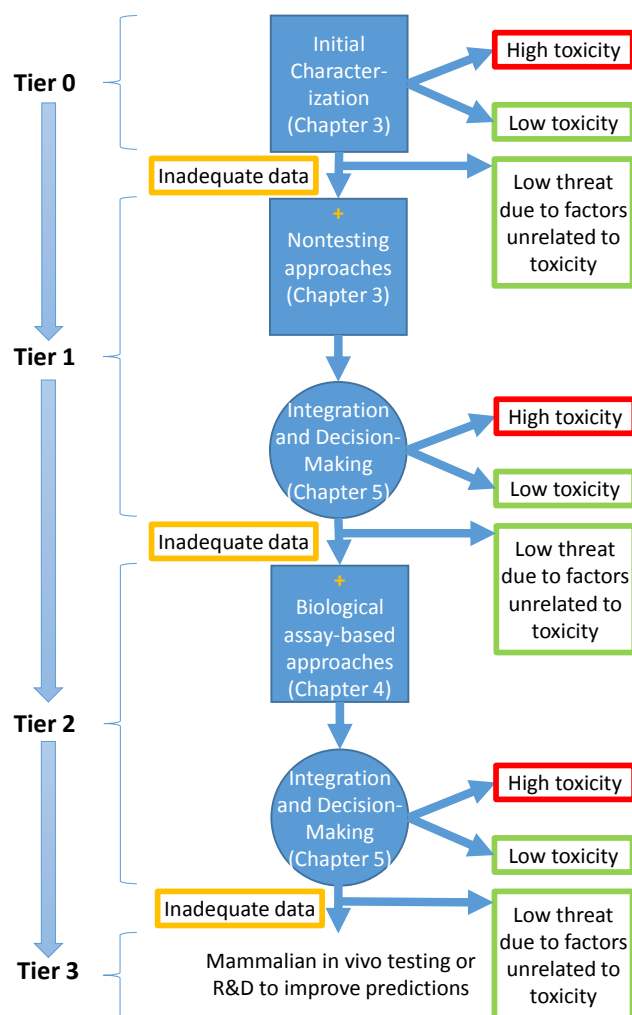


FIGURE 2-2 Prioritization strategy based on a tiered approach for using predictive-toxicology models and tools to evaluate agents for acute toxicity. The strategy can be applied to a single end point (such as lethality, neurotoxicity, and cytotoxicity) and to multiple end points.

BOX 2-2 Tiered Approach to Predicting Toxicity

A **tiered approach to predicting toxicity** consists of successively more predictive and resource-intensive approaches to evaluating toxicity (see Figure 2-2). DOD might deselect a chemical at any tier on the basis of factors unrelated to toxicity, such as availability or weaponizability.

Tier 0 would be an initial chemical characterization of toxicity and physicochemical properties based on existing data. In addition to characterizing acute toxicity, traditional toxicity data can be used to build and test predictive-toxicology models in Tiers 1 and 2, and physicochemical data might be important for understanding potential exposure routes, bioavailability, target-tissue distribution, and potential physical hazards or chemical reactivity associated with an agent. Chapter 3 and Appendix B discuss the availability, accessibility, and sources of acute-toxicity data and other data useful for initial chemical characterization.

Tier 1 uses models and tools that make predictions based on chemical structure and physicochemical properties. Such models and tools, discussed in Chapter 3, are termed nontesting approaches because they do not involve any additional toxicity testing and data generation. Such approaches include the use of structure–activity relationships, quantitative structure–activity relationships, and read-across. As discussed further in Chapter 3, the available approaches and tools differ in their potential applicability to prediction of acute toxicity, their chemical domain of applicability, and their predictive power and degree of uncertainty. There are a number of gaps in chemical space, biological space, and predictivity; for many chemicals or end points, predictions based on nontesting approaches will often be highly uncertain.

Tier 2 is the conduct of biological assays to generate data to reduce uncertainty in the toxicity evaluation. Biological assays in this tier include specific protein assays, cell-based phenotypic assays, organotypic models, and nonmammalian *in vivo* animal models. Toxicity predictions based on such data, discussed in Chapter 4, are termed biological assay-based. Ideally, this biological testing focuses on specific biological targets that are based on information from previous tiers. However, it is also likely to include nonspecific toxicity end points, such as cytotoxicity. As with nontesting approaches, available biological assay-based approaches and tools differ in their potential applicability to prediction of acute toxicity, their chemical domain of applicability, and their predictive power and degree of uncertainty. There are a number of gaps in chemical space, biological space, and predictivity; for many chemicals or end points (although one hopes fewer than in Tier 1), predictions based on biological assay-based approaches will be highly uncertain.

Tier 3 is the conduct of mammalian *in vivo* testing. These traditional approaches are not part of the committee's task. However, the committee notes that as in Tier 2, ideally this toxicity testing will focus on specific biological targets that are based on information from previous tiers. There could also be specific gaps or limitations identified in earlier tiers that could be addressed with additional research or development of new models and tools.

A key step in each tier is integration and decision-making (described in Chapter 5). Even within a tier, such as nontesting approaches (Tier 1), there might be diverse outputs and predictions from different models or tools that need to be synthesized. For example, a simple integration approach could be in the form of a scorecard that counts “positive” and “negative” results from available nontesting approaches; a more sophisticated integration approach might aggregate different predictions. In addition, Tier 2 integration should consider the previous results of nontesting approaches with the newly generated biological assay data. Absorption, distribution, metabolism, and excretion (ADME) considerations can be integrated to provide relevant information, such as chemical bioavailability or distribution to target organs.

The committee envisions that a decision as to whether a chemical is categorized as having high toxicity, low toxicity, or inadequate data could be made for each end point that is relevant to acute toxicity (see examples in Figure 2-1). As discussed previously, such decisions would be based on quantitative toxicity estimates for each toxicity end point and associated levels of confidence or confidence intervals. Defining the specific “thresholds” for assigning a chemical to each category will require expert judgment on the part of DOD. However, reference chemicals with known high and low toxicities could help to inform those boundaries. Overall, chemicals would also be assigned to categories for multiple individual end points that reflect different types of acute toxicity although, as noted in Chapters 3 and 4, there are many gaps in coverage of end points related to acute toxicity at all tiers. Therefore, noting the gaps as part of the prioritization strategy provides guidance on how to target testing in later tiers. And, it is up to DOD to determine the extent of coverage of end points that is adequate for it to make sufficiently reliable decisions at each tier.

FINDINGS AND RECOMMENDATIONS

- **Finding:** There are multiple mechanisms by which chemicals can elicit acute, debilitating toxicity, and these mechanisms provide support for a predictive-toxicology conceptual framework that predicts system, tissue, or organism toxicity on the basis of chemical structure, physicochemical properties, biochemical properties, or biological activity in isolated cells, tissues, and lower organisms.

- **Finding:** Such a conceptual framework that includes databases, assays, models, and tools that are applicable to prediction of acute toxicity could be used to evaluate a large number of chemicals for acute-toxicity potential more rapidly than traditional, mammalian in vivo studies.

- **Finding:** In prioritizing chemicals in terms of their potential to cause acute toxicity, DOD will need to balance a relatively low tolerance for false negatives with a need to evaluate a large number of chemicals rapidly. Regardless of how DOD decides to balance those objectives, they can be managed through a tiered prioritization strategy that applies successively more predictive and resource-intensive approaches as needed.

- **Recommendation:** The committee recommends a prioritization strategy that broadly groups approaches to prediction of acute toxicity into four tiers, beginning with an initial chemical characterization (Tier 0), moving to nontesting approaches (Tier 1), then to biological assay-based approaches (Tier 2), and finally to traditional mammalian in vivo testing (Tier 3). Progression through the tiers will require intermediate integration steps that consider the diverse data within a tier and among tiers. The prioritization strategy can be applied to single or multiple end points.

- **Recommendation:** As part of the prioritization strategy, the committee recommends placing chemicals into one of three general categories at each tier: “high confidence of high toxicity,” “high confidence of low toxicity,” and “inadequate data to evaluate toxicity confidently.” Chemicals placed in the last category, “inadequate data,” are moved to the next tier for additional, more resource-intensive evaluation. Quantitative estimates of how potent the chemicals might be and of the confidence or uncertainty in each estimate will be needed to place chemicals into categories. DOD will need to use expert judgment to define specifically how chemicals are to be assigned to the different categories.

REFERENCES

- Adams, D.J., Z.V. Boskovic, J.R. Theriault, A.J. Wang, A.M. Stern, B.K. Wagner, A.F. Shamji, and S.L. Schreiber. 2013. Discovery of small-molecule enhancers of reactive oxygen species that are nontoxic or cause genotype-selective cell death. *ACS Chem Biol.* 8(5):923-929.

- Andacht, T.M., B.G. Pantazides, B.S. Crow, A. Fidler, D. Noort, J.D. Thomas, T.A. Blake, and R.C. Johnson. 2014. An enhanced throughput method for quantification of sulfur mustard adducts to human serum albumin via isotope dilution tandem mass spectrometry. *J. Anal. Toxicol.* 38(1):8-15.
- Andersson, S., M. Norman, R. Olsson, R. Smith, G. Liu, and J. Nord. 2012. High-precision, room temperature screening assay for inhibitors of microsomal prostaglandin E synthase-1. *J. Biomol. Screen.* 17(10):1372-1378.
- Attene-Ramos, M.S., R. Huang, S. Sakamuru, K.L. Witt, G.C. Beeson, L. Shou, R.G. Schnellmann, C.C. Beeson, R.R. Tice, C.P. Austin, and M. Xia. 2013. Systematic study of mitochondrial toxicity of environmental chemicals using quantitative high throughput screening. *Chem. Res. Toxicol.* 26(9):1323-1332.
- Attene-Ramos, M.S., R. Huang, S. Michael, K.L. Witt, A. Richard, R.R. Tice, A. Simeonov, C.P. Austin, and M. Xia. 2015. Profiling of the Tox21 chemical collection for mitochondrial function to identify compounds that acutely decrease mitochondrial membrane potential. *Environ. Health Perspect.* 123(1):49-56.
- Bandi, S., P. Viswanathan, and S. Gupta. 2014. Evaluation of cytotoxicity and DNA damage response with analysis of intracellular ATM signaling pathways. *Assay Drug Dev. Technol.* 12(5):272-281.
- Benson, J.M., B.M. Tibbetts, W.M. Weber, and G.R. Grotendorst. 2011a. Uptake, tissue distribution, and excretion of ¹⁴C-sulfur mustard vapor following inhalation in F344 rats and cutaneous exposure in hairless guinea pigs. *J. Toxicol. Environ. Health A* 74(13):875-885.
- Benson, J.M., A.P. Gomez, M.L. Wolf, B.M. Tibbetts, and T.H. March. 2011b. The acute toxicity, tissue distribution, and histopathology of inhaled ricin in Sprague Dawley rats and BALB/c mice. *Inhal. Toxicol.* 23(5):247-256.
- Büch, T.R., E.A. Schäfer, M.T. Demmel, I. Boekhoff, H. Thiermann, T. Gudermann, D. Steinritz, and A. Schmidt. 2013. Functional expression of the transient receptor potential channel TRPA1, a sensor for toxic lung inhalants, in pulmonary epithelial cells. *Chem. Biol. Interact.* 206(3):462-471.
- Chen, Y., C. Guo, L. Lim, S. Cheong, Q. Zhang, K. Tang, and J. Reboud. 2008. Compact microelectrode array system: Tool for in situ monitoring of drug effects on neurotransmitter release from neural cells. *Anal. Chem.* 80(4):1133-1140.
- Cui, H.F., J.S. Ye, Y. Chen, S.C. Chong, and F.S. Sheu. 2006. Microelectrode array biochip: Tool for in vitro drug screening based on the detection of a drug effect on dopamine release from PC12 cells. *Anal. Chem.* 78(18):6347-6355.
- deLemos, A.S., D.M. Foureau, C. Jacobs, W. Ahrens, M.W. Russo, and H.L. Bonkovsky. 2014. Drug-induced liver injury with autoimmune features. *Semin. Liver Dis.* 34(2):194-204.
- DHHS (US Department of Health & Human Services). 2014. Chemical Warfare. Classes of Chemical Agents. Specialized Information System [online]. Available: http://sis.nlm.nih.gov/enviro/chemical_warfare.html#a1 [accessed March 13, 2015].
- Dunlop, J., R. Roncarati, B. Jow, H. Bothmann, T. Lock, D. Kowal, M. Bowlby, and G.C. Terstappen. 2007. In vitro screening strategies for nicotinic receptor ligands. *Biochem. Pharmacol.* 74(8):1172-1181.
- Falk, M., M. Hausmann, E. Lukášová, A. Biswas, G. Hildenbrand, M. Davidková, E. Krasavin, Z. Kleibl, I. Falková, L. Ježková, L. Štefančíková, J. Ševčík, M. Hofer, A. Bačíková, P. Matula, A. Boreyko, J. Vachelová, A. Michaelidesová, and S. Kozubek. 2014. Determining Omics spatiotemporal dimensions using exciting new nanoscopy techniques to assess complex cell responses to DNA damage: Part A--radiomics. *Crit. Rev. Eukaryot. Gene Expr.* 24(3):205-223.
- Gutzkow, K.B., T.M. Langleite, S. Meier, A. Graupner, A.R. Collins, and G. Brunborg. 2013. High-throughput comet assay using 96 minigels. *Mutagenesis* 28(3):333-340.
- Hill, A.J., N.A. Jones, I. Smith, C.L. Hill, C.M. Williams, G.J. Stephens, and B.J. Whalley. 2014. Voltage-gated sodium (NaV) channel blockade by plant cannabinoids does not confer anticonvulsant effects per se. *Neurosci. Lett.* 566:269-274.
- Himmel, H.M. 2013. Drug-induced functional cardiotoxicity screening in stem cell-derived human and mouse cardiomyocytes: Effects of reference compounds. *J. Pharmacol. Toxicol. Methods* 68(1):97-111.
- Holstege, C.P., L.K. Bechtel, T.H. Reilly, B.P. Wispelwey, and S.G. Dobmeier. 2007. Unusual but potential agents of terrorists. *Emerg. Med. Clin. North Am.* 25(2):549-566.
- Huh, D., D.C. Leslie, B.D. Matthews, J.P. Fraser, S. Jurek, G.A. Hamilton, K.S. Thorneloe, M.A. McAlexander, and D.E. Ingber. 2012. A human disease model of drug toxicity-induced pulmonary edema in a lung-on-a-chip microdevice. *Sci. Transl. Med.* 4(159):159ra147.
- Jellett, J.F., L.J. Marks, J.E. Stewart, M.L. Dorey, W. Watson-Wright, and J.F. Lawrence. 1992. Paralytic shellfish poison (saxitoxin family) bioassays: Automated endpoint determination and standardization

- of the in vitro tissue culture bioassay, and comparison with the standard mouse bioassay. *Toxicol.* 30(10):1143-1156.
- Jensen, K.H., and C. Rekling. 2010. Development of a no-wash assay for mitochondrial membrane potential using the styryl dye DASPEI. *J. Biomol. Screen.* 15(9):1071-1081.
- Jin, S., K.S. Sarkar, Y.N. Jin, Y. Liu, D. Kokel, T.J. Van Ham, L.D. Roberts, R.E. Gerszten, C.A. Macrae, and R.T. Peterson. 2013. An in vivo zebrafish screen identifies organophosphate antidotes with diverse mechanisms of action. *J. Biomol. Screen.* 18(1):108-115.
- Jones, J.D., H.L. Kirsch, R.L. Wortmann, and M.H. Pillinger. 2014. The causes of drug-induced muscle toxicity. *Curr. Opin. Rheumatol.* 26(6):697-703.
- Kelesidis, T., C.K. Roberts, D. Huynh, O. Martinez-Maza, J.S. Currier, S.T. Reddy, and O.O. Yang. 2014. A high throughput biochemical fluorometric method for measuring lipid peroxidation in HDL. *PLoS One* 9(11):e111716.
- Li, Y., X. Feng, W. Du, Y. Li, and B.F. Liu. 2013. Ultrahigh-throughput approach for analyzing single-cell genomic damage with an agarose-based microfluidic comet array. *Anal. Chem.* 85(8):4066-4073.
- Liu, S., Y. Zhang, M. Moayeri, J. Liu, D. Crown, R.J. Fattah, A.N. Wein, Z.X. Yu, T. Finkel, and S.H. Leppla. 2013. Key tissue targets responsible for anthrax-toxin-induced lethality. *Nature* 501(7465):63-68.
- Lopez-Izquierdo, A., M. Warren, M. Riedel, S. Cho, S. Lai, R.L. Lux, K.W. Spitzer, I.J. Benjamin, M. Tristani-Firouzi, and C.J. Jou. 2014. A near-infrared fluorescent voltage-sensitive dye allows for moderate-throughput electrophysiological analyses of human induced pluripotent stem cell-derived cardiomyocytes. *Am. J. Physiol. Heart Circ. Physiol.* 307(9):H1370-H1377.
- Mioulane, M., G. Foldes, N.N. Ali, M.D. Schneider, and S.E. Harding. 2012. Development of high content imaging methods for cell death detection in human pluripotent stem cell-derived cardiomyocytes. *J. Cardiovasc. Transl. Res.* 5(5):593-604.
- Monteiro-Riviere, N.A., and A.O. Inman. 1997. Ultrastructural characterization of sulfur mustard-induced vesication in isolated perfused porcine skin. *Microsc. Res. Tech.* 37(3):229-241.
- Morrison, K.C., and P.J. Hergenrother. 2012. Whole cell microtubule analysis by flow cytometry. *Anal. Biochem.* 420(1):26-32.
- NRC (National Research Council). 2005. *Impact of Revised Airborne Exposure Limits on Non-Stockpile Chemical Materiel Program Activities*. Washington, DC: National Academies Press.
- NRC (National Research Council). 2007. *Toxicity Testing in the 21st Century: A Vision and a Strategy*. Washington, DC: National Academies Press.
- Pantazides, B.G., B.S. Crow, J.W. Garton, J.A. Quiñones-González, T.A. Blake, J.D. Thomas, and R.C. Johnson. 2015. Simplified method for quantifying sulfur mustard adducts to blood proteins by Ultra-high Pressure Liquid Chromatography-Isotope Dilution Tandem Mass Spectrometry. *Chem. Res. Toxicol.* 28(2):256-261.
- Pieperhoff, S., K.S. Wilson, J. Baily, K. de Mora, S. Maqsood, S. Vass, J. Taylor, J. Del-Pozo, C.A. MacRae, J.J. Mullins, and M.A. Denvir. 2014. Heart on a plate: Histological and functional assessment of isolated adult zebrafish hearts maintained in culture. *PLoS One* 9(5):e96771.
- Pointon, A., N. Abi-Gerges, M.J. Cross, and J.E. Sidaway. 2013. Phenotypic profiling of structural cardiotoxins in vitro reveals dependency on multiple mechanisms of toxicity. *Toxicol. Sci.* 132(2):317-326.
- Pointon, A., A.R. Harmer, I.L. Dale, N. Abi-Gerges, J. Bowes, C. Pollard, and H. Garside. 2015. Assessment of cardiomyocyte contraction in human-induced pluripotent stem cell-derived cardiomyocytes. *Toxicol. Sci.* 144(2):227-237.
- Prasad, R.Y., J.K. McGee, M.G. Killius, D.A. Suarez, C.F. Blackman, D.M. DeMarini, and S.O. Simmons. 2013. Investigating oxidative stress and inflammatory responses elicited by silver nanoparticles using high-throughput reporter genes in HepG2 cells: effect of size, surface coating, and intracellular uptake. *Toxicol. In Vitro* 27(6):2013-2021.
- Riviere, J.E., J.D. Brooks, P.L. Williams, and N.A. Monteiro-Riviere. 1995. Toxicokinetics of topical sulfur mustard penetration, disposition, and vascular toxicity in isolated perfused porcine skin. *Toxicol. Appl. Pharmacol.* 135(1):25-34.
- Sakamuru, S., X. Li, M.S. Attene-Ramos, R. Huang, J. Lu, L. Shou, M. Shen, R.R. Tice, C.P. Austin, and M. Xia. 2012. Application of a homogenous membrane potential assay to assess mitochondrial function. *Physiol. Genomics* 44(9):495-503.
- Scott, C.W., X. Zhang, N. Abi-Gerges, S.D. Lamore, Y.A. Abassi, and M.F. Peters. 2014. An impedance-based cellular assay using human iPSC-derived cardiomyocytes to quantify modulators of cardiac contractility. *Toxicol. Sci.* 142(2):331-338.

- Sharma, R., B. Gupta, N. Singh, J.R. Acharya, K. Musilek, K. Kuca, and K.K. Ghosh. 2015. Development and structural modifications of cholinesterase reactivators against chemical warfare agents in last decade: A review. *Mini Rev. Med. Chem.* 15(1):58-72.
- Sirenko, O., E.F. Cromwell, C. Crittenden, J.A. Wignall, F.A. Wright, and I. Rusyn. 2013. Assessment of beating parameters in human induced pluripotent stem cells enables quantitative in vitro screening for cardiotoxicity. *Toxicol. Appl. Pharmacol.* 273(3):500-507.
- Sirenko, O., J. Hesley, I. Rusyn, and E.F. Cromwell. 2014a. High-content assays for hepatotoxicity using induced pluripotent stem cell-derived cells. *Assay Drug Dev. Technol.* 12(1):43-54.
- Sirenko, O., J. Hesley, I. Rusyn, and E.F. Cromwell. 2014b. High-content high-throughput assays for characterizing the viability and morphology of human iPSC-derived neuronal cultures. *Assay Drug Dev. Technol.* 12(9-10):536-547.
- Steinhoff, R.F., M. Ivarsson, T. Habicher, T.K. Villiger, J. Boertz, J. Krismer, S.R. Fagerer, M. Soos, M. Morbidelli, M. Pabst, and R. Zenobi. 2015. High-throughput nucleoside phosphate monitoring in mammalian cell fed-batch cultivation using quantitative matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Biotechnol J.* 10(1):190-198.
- Stenger, B., F. Zehfuß, H. Mückter, A. Schmidt, F. Balszuweit, E. Schäfer, T. Büch, T. Gudermann, H. Thiermann, and D. Steinritz. In press. Activation of the chemosensing transient receptor potential channel A1 (TRPA1) by alkylating agents. *Arch. Toxicol.*
- Sturm, M.B., and V.L. Schramm. 2009. Detecting ricin: Sensitive luminescent assay for ricin A-chain ribosome depurination kinetics. *Anal. Chem.* 81(8):2847-2853.
- Suzuki, O.T., A. Frick, B.B. Parks, O.J. Trask Jr., N. Butz, B. Steffy, E. Chan, D.K. Scoville, E. Healy, C. Benton, P.E. McQuaid, R.S. Thomas, and T. Wiltshire. 2014. A cellular genetics approach identifies gene-drug interactions and pinpoints drug toxicity pathway nodes. *Front. Genet.* 5:272.
- Tenberken, O., J. Mikler, I. Hill, K. Weatherby, H. Thiermann, F. Worek, and G. Reiter. 2010. Toxicokinetics of tabun enantiomers in anaesthetized swine after intravenous tabun administration. *Toxicol. Lett.* 198(2):177-181.
- Thoren, T.M., K.S. Thompson, P.S. Cardona, A.K. Chaturvedi, and D.V. Canfield. 2013. In vitro absorption of atmospheric carbon monoxide and hydrogen cyanide in undisturbed pooled blood. *J. Anal. Toxicol.* 37(4):203-207.
- Thorn, J. 2001. The inflammatory response in humans after inhalation of bacterial endotoxin: A review. *Inflamm. Res.* 50(5):254-261.
- Vale, C., A. Alfonso, M.R. Vieytes, X.M. Romaris, F. Arévalo, A.M. Botana, and L.M. Botana. 2008. In vitro and in vivo evaluation of paralytic shellfish poisoning toxin potency and the influence of the pH of extraction. *Anal. Chem.* 80(5):1770-1776.
- Vallet, V., C. Cruz, J. Licausi, A. Bazire, G. Lallement, and I. Boudry. 2008. Percutaneous penetration and distribution of VX using in vitro pig or human excised skin validation of demeton-S-methyl as adequate simulant for VX skin permeation investigations. *Toxicology* 246(1):73-82.
- van der Linden, S.C., A.R. von Bergh, B.M. van Vught-Lussenburg, L.R. Jonker, M. Teunis, C.A. Krul, and B. van der Burg. 2014. Development of a panel of high-throughput reporter-gene assays to detect genotoxicity and oxidative stress. *Mutat. Res. Genet. Toxicol. Environ. Mutagen.* 760:23-32.
- Vangveravong, S., E. McElveen, M. Taylor, J. Xu, Z. Tu, R.R. Luedtke, and R.H. Mach. 2006. Synthesis and characterization of selective dopamine D2 receptor antagonists. *Bioorg. Med. Chem.* 14(3):815-825.
- Vongs, A., K.J. Solly, L. Kiss, D.J. Macneil, and C.I. Rosenblum. 2011. A miniaturized homogenous assay of mitochondrial membrane potential. *Assay Drug Dev. Technol.* 9(4):373-381.
- Wasalathanthri, D.P., S. Malla, I. Bist, C.K. Tang, R.C. Faria, and J.F. Rusling. 2013. High-throughput metabolic genotoxicity screening with a fluidic microwell chip and electrochemiluminescence. *Lab. Chip.* 13(23):4554-4562.
- Watson, C., J. Ge, J. Cohen, G. Pyrgiotakis, B.P. Engelward, and P. Demokritou. 2014. High-throughput screening platform for engineered nanoparticle-mediated genotoxicity using CometChip technology. *ACS Nano.* 8(3):2118-2133.
- Wijte, D., M.J. Alblas, D. Noort, J.P. Langenberg, and H.P. van Helden. 2011. Toxic effects following phosgene exposure of human epithelial lung cells in vitro using a CULTEX® system. *Toxicol. In Vitro* 25(8):2080-2087.
- Wille, T., H. Thiermann, and F. Worek. 2010. Development of a high-throughput screening for nerve agent detoxifying materials using a fully-automated robot-assisted biological assay. *Toxicol. In Vitro* 24(3):1026-1031.

28 *Application of Modern Toxicology Approaches for Predicting Acute Toxicity for Chemical Defense*

- Wills, L.P., G.C. Beeson, R.E. Trager, C.C. Lindsey, C.C. Beeson, Y.K. Peterson, and R.G. Schnellmann. 2013. High-throughput respirometric assay identifies predictive toxicophore of mitochondrial injury. *Toxicol. Appl. Pharmacol.* 272(2):490-502.
- Worek, F., P. Eyer, N. Aurbek, L. Szinicz, and H. Thiermann. 2007. Recent advances in evaluation of oxime efficacy in nerve agent poisoning by in vitro analysis. *Toxicol. Appl. Pharmacol.* 219(2-3):226-234.
- Xiao, J., R.B. Free, E. Barnaeva, J.L. Conroy, T. Doyle, B. Miller, M. Bryant-Genevier, M.K. Taylor, X. Hu, A.E. Dulcey, N. Southall, M. Ferrer, S. Titus, W. Zheng, D.R. Sibley, and J.J. Marugan. 2014. Discovery, optimization, and characterization of novel D2 dopamine receptor selective antagonists. *J. Med. Chem.* 57(8):3450-3463.
- Yakushenko, A., E. Kästelhön, and B. Wolfrum. 2013. Parallel on-chip analysis of single vesicle neurotransmitter release. *Anal. Chem.* 85(11):5483-5490.
- Zielonka, J., G. Cheng, M. Zielonka, T. Ganesh, A. Sun, J. Joseph, R. Michalski, W.J. O'Brien, J.D. Lambeth, and B. Kalyanaraman. 2014. High-throughput assays for superoxide and hydrogen peroxide: Design of a screening workflow to identify inhibitors of NADPH oxidases. *J. Biol. Chem.* 289(23):16176-16189.

3

Nontesting Approaches Relevant to Prediction of Acute Toxicity and Potency

The term *nontesting approaches* was coined during the development of the European Union Registration, Evaluation, Authorization, and Restriction of Chemicals (REACH) regulation (EU 2006; ECHA 2008) to include the search and retrieval of existing data, the identification of structural alerts¹ to indicate activity, the grouping of chemicals for read-across, and the development and application of quantitative structure-activity models. In practice, nontesting approaches are used to accomplish various tasks. For example, predictions based on structure-activity relationships (SARs) and quantitative structure-activity relationships (QSARs) are used to fill specific data gaps in lieu of experimental testing, to support findings or conclusions in integrated chemical assessments, and to substantiate predictions of various properties for structurally related chemicals. A key assumption that underpins nontesting approaches is that the property of a chemical with respect to how it will interact with a defined biological system is inherent in its molecular structure; thus, similar chemicals should have similar biological activities (the similarity principle) (Raunio 2011). More detailed information about nontesting approaches can be found in Cronin and Madden (2010).

This chapter discusses nontesting approaches in the context of the conceptual framework described in Chapter 2. Box 3-1 provides definitions for some of the terms used in the chapter. The chapter discusses the use of available data to characterize chemicals of interest and *in silico* approaches to predict physical hazards, chemical reactivity, pharmacokinetic properties, and acute toxicity. It also provides selected examples to demonstrate how computational tools could be used in predictive toxicology.

INITIAL CHEMICAL CHARACTERIZATION

The starting point in the application of any nontesting approach for predicting acute toxicity is a preliminary search and evaluation of available data on the chemical of interest. The effort often begins with database queries, literature searches, and other approaches for finding information about the chemical's structure, physicochemical properties, and acute toxicity (see Box 3-2). The committee notes that the forthcoming REACH regulation requires *in vivo* acute oral toxicity information for chemicals manufactured or imported into Europe at greater than 1 metric ton per year (ECHA 2012). It is anticipated that a large volume of *in vivo* oral data will be potentially disseminated publically after the REACH May 2018 deadline.

¹A structural alert is a chemical structure that has been linked to toxicity or a specific toxicity end point.

BOX 3-1 Definitions of Selected Nontesting Approaches

A *structure–activity relationship (SAR)* is a qualitative association between a chemical (sub)structure (such as a functional group) and the potential of a chemical that contains the (sub)structure to exhibit a particular biological effect.

A *quantitative structure–activity relationship (QSAR)* is “a mathematical relationship between a quantifiable aspect of chemical structure and a chemical property or reactivity or a well defined biological activity, such as toxicity” (EPA 2012). QSARs can be derived to predict quantitative or qualitative end points.

A *quantitative structure–property relationship (QSPR)* is a special case of QSAR in which a physicochemical property is modelled as the response variable.

An *expert system* is a software tool that specifically encodes compilations of SARs, QSARs, or both to enable rational predictions of toxicity to be made on the basis of structure alone. Expert systems are typically categorized as statistical (for example, TOPKAT and Accelrys Inc), knowledge-based (for example, such SAR-based approaches as Derek Nexus and LHASA Ltd), or hybrid (for example, TIMES-SS).

Category approach, analogue approach, and read-across: *Category* and *analogue approaches* are techniques for grouping chemicals; *read-across* is a technique for filling data gaps in category and analogue approaches (ECHA 2008; OECD 2014). Read-across can be qualitative or quantitative and uses existing “information on the property of a substance (source chemical)...to make a prediction of the same property for another substance (target chemical) that is considered similar” with respect to the end point of interest (Worth 2008). Analogue approaches are used for grouping a small number of chemicals when there are no apparent trends in properties.

A *chemical category* is “a group of chemicals whose physicochemical and human health...or environmental toxicological properties...or environmental-fate properties are likely to be similar or follow a regular pattern as a result of structural similarity” (OECD 2007a, 2014). Chemical similarity could be based on a variety of properties, including the presence of a common functional group (such as an aldehyde), common constituents or chemical classes, similar carbon range numbers, or common precursors or breakdown products.

In silico approaches include computational modeling, SAR analysis, analysis of physicochemical characteristics, and read-across techniques.

The available information can help in identifying a chemical’s potential for direct physical hazards, most relevant routes of exposure, likely bioavailability, and potential for inducing (human) toxicity (NRC 2014). It should also be considered before designing or initiating new in vitro or in vivo experimental studies, in interpreting existing empirical data, or in selecting appropriate (Q)SAR models.²

As described below, physicochemical properties of interest in predictive toxicology can be nominally categorized into three broad types: physical properties, solvation properties, and molecular attributes (NRC 2014). There are methods for empirically measuring the properties and in silico approaches for estimating their values (see Box 3-3). It can be particularly helpful to complement estimated values with experimental measurements, when that is possible.

²The committee uses the shorthand notation (Q)SAR to indicate both SAR and QSAR.

BOX 3-2 Primary Data Considered During a Preliminary
Characterization of a Chemical of Interest

Chemical structure: Chemical structure is the spatial arrangement of a molecule's constituent atoms. PubChem, DSSTox, and ChemIDplus are examples of searchable chemical-structure databases.

Physicochemical properties: Physicochemical properties contribute to the inherent hazards posed by a chemical, including its ability to interfere with normal biological processes. Physicochemical properties could also define a chemical's physical hazards of interest (such as corrosivity). Physical properties include freezing point, boiling point, melting point, infrared spectrum, electronic characteristics, viscosity, and density. Other properties of relevance include solvation properties, such as phase partitioning and solubility. One of the more important phase-partition coefficients is obtained from a system in which one solvent is water or an aqueous phase and the second is organic and hydrophobic, such as 1-octanol, that is, the octanol-water partition coefficient (K_{ow}). An example of a database source of physicochemical-property data is the National Institute of Standards and Technology Chemistry WebBook; another is the PHYSPROP database, which is integrated into the US Environmental Protection Agency's EPISuite software.

Acute toxicity (for example, rodent LD_{50} or LC_{50} values): These data might be available from primary sources (such as peer-reviewed literature) and secondary sources (such as the *Merck Handbook*). By far the most convenient sources of data are compiled databases that are readily searchable by chemical identifiers, such as chemical name, Chemical Abstracts Service registry number, or chemical structure. An example is the National Library of Medicine TOXNET® database.

See Appendix B for additional information and examples.

- *Physical properties.* Physical properties include such characteristics as freezing point, melting point, boiling point, vapor pressure, and viscosity. Melting point, boiling point, and vapor pressure can be used to predict a chemical's likely physical state, which is pertinent in determining the most relevant route of exposure for any testing or indeed what practical challenges might need to be overcome in in vitro testing scenarios or even what issues to consider in interpreting in vivo results and associated testing protocols.

- *Solvation properties.* Solvation properties describe a chemical's interaction with different phases and its interaction between phases (for example, $\log K_{ow}$ represents the partitioning between octanol and water). Water solubility and $\log K_{ow}$ are particularly helpful in determining the technical feasibility of performing in vitro test protocols given that they typically use aqueous media.

- *Molecular attributes.* Molecular attributes capture properties related to molecular shape and size. Electronic characteristics of molecules, such as frontier orbital energies and polarizability, that are related to reactivity could be considered to constitute a type of molecular attribute. They also play a role in predicting likely bioavailability and toxicity.

**USE OF PHYSICOCHEMICAL PROPERTIES TO PREDICT PHYSICAL HAZARDS,
CHEMICAL REACTIVITY, AND PHARMACOKINETICS**

Physicochemical data can be used to predict a chemical's physical hazard, reactivity, and pharmacokinetics, including absorption by different exposure routes, distribution in the body, and likely metabolites. Approaches that apply knowledge about a chemical's physicochemical properties to predictive toxicology presume that for a chemical to exert a toxic effect, it typically must be bioavailable to such an extent that it (or its metabolite) reaches a biochemical target, where it can exert its toxic effect (Meek et al. 2013).

BOX 3-3 In Silico Approaches for Predicting Physicochemical Properties

In the absence of data on physicochemical properties, reasonable estimates based on chemical structure are feasible with the use of QSAR and QSPR models (Dearden and Worth 2007). A discussion of those methods is beyond the scope of the present activity. However, the National Research Council report *A Framework to Guide Selection of Chemical Alternatives* provides a succinct discussion of published models that can be used to characterize a number of physicochemical properties (NRC 2014). That report discusses methods used to estimate molecular hydrophobicity (or lipophilicity) and other physicochemical end points, such as aqueous solubility, pKa, and the electronic properties of molecules.

A number of software packages and algorithms are available for predicting physicochemical properties, and predictions are often in excellent agreement with experimentally derived values. For example, pKa can be estimated by using Taft and Hansch fragment coefficients, and Perrin et al. (1981) contains an extensive compilation of the fragment values and relevant equations to do so. For convenience, software tools, such as SPARC or those created by ACD Labs, contain algorithms for estimating pKa directly from chemical structure. The user of such tools, however, must have a basic understanding of the inherent advantages and limitations of the various algorithms as they are related to the accuracy of physicochemical-property prediction. In general, the QSAR models available for prediction of the key physicochemical characteristics are best suited for small organic chemicals that typically have one functional group. Other chemicals—such as pesticides, drug-like chemicals, or other pharmacological actives—typically lack published data to derive such QSARs.

Use of Physicochemical Properties to Predict Physical Hazard and Chemical Stability or Reactivity

Assessing the likely irritant or corrosive effects of a chemical would be a helpful component of a tiered evaluation strategy for predicting acute toxicity.³ In the absence of measured irritant or corrosive data, a handful of (Q)SAR approaches are useful in identifying potential irritants or corrosives. The German Federal Institute for Risk Assessment (BfR) rule base (Gerner et al. 2004; Hulzebos et al. 2005) is one example of an expert system that uses physicochemical exclusion rules and structural alert inclusion rules to determine likely skin or eye irritation hazard.⁴ The BfR rule base has been encoded into software tools, including the Organisation for Economic Co-operation and Development (OECD) QSAR Toolbox⁵ and the European Commission Joint Research Centre Toxtree (Rorije and Hulzebos 2005; Tsakovska et al. 2007). Some QSARs published for specific chemical classes have relied on such properties as pKa, dipole moment, logK_{ow}, and molecular weight or volume to estimate likely irritation potential and potency (Barratt 1996). Both pKa and dipole moment have been found to be useful measures for modeling chemical reactivity depending on whether the target substance is an acid, a base, or a neutral organic; and logK_{ow} and molecular weight have served as useful surrogates for modeling partitioning. QSARs also exist within expert

³Skin irritation or corrosion can be investigated in vitro by virtue of assays, such as Corrositex (OECD 2006) for corrosion and EpiDerm™ (OECD 2013a) for irritation. For eye irritation or corrosion, various ex vivo and in vitro assays are available, including the bovine corneal opacity permeability test (OECD 2013b), the isolated chicken-eye test (OECD 2013c), or the EpiOcular™ eye-irritation test method. A tabulation of assays for irritation and corrosion that have been validated by ECVAM or ICCVAM or that exist as test guidelines under OECD are provided on the AltTox.org Web site (AltTox 2014) and are discussed in Chapter 4.

⁴The BfR rule base combines two approaches: exclusion rules that use physicochemical thresholds to identify chemicals that are not skin irritants or corrosive and inclusion rules that use structural alerts to identify chemicals that are potentially irritants or corrosive (Saliner et al. 2007). The rule base assigns a regulatory classification for skin or eye irritation or corrosion.

⁵The committee refers here to the toolbox by its official name rather than OECD (Q)SAR Toolbox, which would be more appropriate because the Toolbox includes both SAR and QSAR approaches.

systems, such as TOPKAT and MCASE, for the prediction of irritation or corrosion. Saliner et al. (2008) reviewed the status of (Q)SAR approaches for irritation and corrosion.

Consideration should also be paid to the inherent stability or electrophilic reactivity of the chemical. A chemical might exert its effects in its parent form or be transformed abiotically or biotically to a metabolite that is a more relevant target for evaluation. Some substances are rapidly hydrolyzed; for example, acid chlorides and acid anhydrides are rapidly hydrolyzed to their corresponding carboxylic acids. Other substances are capable of being oxidized when exposed to air; for example, *p*-hydroquinone is rapidly oxidized to its corresponding benzoquinone, which is highly reactive. The OECD Toolbox contains simulators that help in predicting such transformations. Consideration of how a chemical might be transformed is important in interpreting experimental data, performing new testing, or using the most relevant target for (Q)SAR analyses.

Use of Physicochemical Properties to Predict Chemical Disposition and Metabolism

Pharmacokinetics describes the disposition of a chemical in an organism and considers chemical absorption, distribution, metabolism, and excretion (ADME). Pharmacokinetic properties can play an important role in the assessment of a chemical's effects on or risks to the body. In silico approaches have been developed to predict many ADME processes; the sections below focus largely on approaches that are directly relevant for predicting acute toxicity.

Absorption: Oral

Some physicochemical properties—such as molecular weight, the number of hydrogen-bond donors and acceptors, and $\log K_{ow}$ —have been shown to be predictive of oral absorption. For example, Lipinski's rule of 5 is considered helpful in evaluating the likely absorption, permeability, and toxicity of drug-like substances (Lipinski et al. 2001) and considers the three properties noted to make predictions about chemical behavior. Other examples of heuristic rules are provided in Table 3-1.

In addition to heuristic rules, several QSAR models have been developed to determine intestinal absorption and oral absorption. Iyer et al. (2007) used a membrane-interaction QSAR analysis to build models for human oral intestinal drug absorption. Castillo-Garit et al. (2008) developed a mathematical model that used linear indexes to predict the in vitro permeability of 157 chemicals in a Caco-2 cell model. Their mathematical model had greater than 80% accuracy in predicting how well a drug would be absorbed by Caco-2 cells. Guerra et al. (2010) developed an artificial neural network by using CODES 2D descriptions to predict oral drug absorption. Suenderhauf et al. (2011) used a broad selection of machine learning and statistical methods to derive classification and prediction models for human intestinal absorption. Several recent reviews discuss the status of such QSAR models (Xu and Mager 2011; Silva and Trossini 2014; Wang and Hou 2015).

TABLE 3-1 Examples of Heuristic Rules to Predict Oral Absorption

Rule	Description ^a	Reference
GlaxoSmithKline rule of 4/400	Chemicals with $c\text{LogP} < 4$ and $\text{MW} < 400$ Da have superior drug-like properties compared with chemicals with $c\text{LogP} > 4$ and $\text{MW} > 400$ Da	Gleeson 2008
Pfizer rule of 3/75	Chemicals with $c\text{LogP} > 3$ and total PSA < 75 Å are 2.5 times more likely to have in vivo toxicity than ones with $c\text{LogP} < 3$ and total PSA > 75 Å	Hughes et al. 2008
AstraZeneca	Alkalinity and increased $c\text{LogP}$ are associated with multiple positive responses in various toxicity assays	Leeson and Springthorpe 2007

^a $c\text{LogP}$ is the name of a software program that generates an estimate of $\log K_{ow}$. Abbreviations: MW, molecular weight; PSA, polar surface area.

There are also physiologically based packages that can predict oral absorption rates of drugs, such as GastroPlus and SimCyp (Kuentz et al. 2006; Rostami-Hodjegan and Tucker 2007; Yang et al. 2007; Simulations Plus 2010; Grbic et al. 2011). However, if the goal is to determine likely oral absorption to help to prioritize chemicals, the heuristic rules described above might be adequate for that task.

Absorption: Dermal

LogK_{ow} , water solubility, and molecular weight are also useful inputs for estimating dermal absorption characteristics of a target substance. QSAR models for predicting the dermal permeability coefficient (K_p)—a measure useful for modeling dermal penetration—typically rely on LogK_{ow} and molecular weight as input variables. Potts and Guy (1992) derived such a model that is also encoded in DERMWIN as part of EPA's EpiSuite software. Over the years, the model derived by Potts and Guy has been modified to address limitations, and many variants now exist. Mitragotri et al. (2011) reviewed the status of models for the prediction of skin permeability in terms of their strengths, limitations, and future prospects.

Other researchers have incorporated additional information, such as degree of hydrogen bonding and melting point, to refine skin penetration estimates (Hostýnek 1997; Magnusson et al. 2004; ten Berge 2009; Dancik et al. 2013a). Models by ten Berge (2009) and Dancik et al. (2013a,b) are helpful in evaluating systemic availability as a result of dermal exposure and thus provide a means of extrapolating a dermal acute-toxicity (LD_{50}) value from an oral acute-toxicity (LD_{50}) value. Kasting's model (as discussed in Dancik et al. 2013a,b) explicitly takes into account various components of the skin structure, including the stratum corneum, viable epidermis, and dermis. The model simulates one-dimensional transient passive transport into a skin slab. Some properties are also required as inputs for a simulation, including logK_{ow} , vapor pressure, melting and boiling points, molecular weight, chemical class (alcohol, hydrocarbon, or other organic), and presence of a pharmacophore⁶ as defined by Yamazaki and Kanaoka (2004). The model is available for use from the National Institute for Occupational Safety and Health Web site (NIOSH 2013).

Absorption and Deposition: Inhalation

For nonvolatile chemicals, particle size is an important consideration because it affects deposition in the respiratory tract and influences whether a particle poses an inhalation hazard (ECETOC 2012; Brown et al. 2013). Brown et al. (2013) predicted that about half of all 10- μm particles penetrate into the thorax and that about 20% or less of all 10- μm particles would penetrate to the extrathoracic airways and into the lower respiratory tract.

For volatile substances, physicochemical characteristics—such as vapor pressure, water solubility, and reactivity—are also important for predicting acute toxicity by the inhalation route (Veith and Wallace 2006; Veith et al. 2009).

Metabolism

Several tissues—including the lung, skin, liver, intestine, and kidney—have enzymes that can convert a parent chemical to a metabolite, for example, through oxidation and conjugation processes. Whereas parent chemical metabolism typically results in a more hydrophilic chemical that is more easily excreted, a reactive toxic metabolite is sometimes formed. Not considering that possibility and focusing solely on the parent chemical will therefore be inadequate in characterizing a chemical's potential to elicit acute toxicity accurately.

⁶A pharmacophore is the collection of steric and electrostatic features of different chemicals that are necessary to ensure optimal molecular interactions with a specific biological target (Langer and Wolber 2004).

Predicting chemical metabolism requires tools that can identify the functional groups or structural components of the parent chemical that are vulnerable to metabolism (sites of metabolism) and the structure of possible metabolites. Additional information about enzyme structure and function and about the effect of metabolism on the induction or inhibition of metabolizing enzymes could also be considered, but for the purposes of predicting the potential of chemicals to elicit acute toxicity, this discussion will focus primarily on the first two factors. Determining the site of metabolism allows prediction of overall metabolic stability, such as the rate of activity (V_{\max} , K_m) or clearance rate (Cl_{int}), that is measured either *in vivo* or *in vitro*. Predicting metabolic structures involves listing possible metabolites from reactions that the chemical could undergo (biotransformation).

Metabolism-predictive tools are based on large compilations of databases derived from metabolism information in the literature, for example, Accelrys Metabolite Database, Metabolite, MetaBase, and MetaDrug (Kirchmair et al. 2012). The metabolism information in a database can be used to identify likely metabolic sites on the basis of what is known about the target chemical structure or a similar chemical structure. The databases can also be used to predict possible metabolites by using information that describes enzyme activity, such as binding pocket sites. Most available metabolism-predictive tools consider a single aspect of metabolic reactions—such as the reaction energy barrier, geometrical properties, or pharmacokinetic properties—to predict sites of metabolism or potential metabolites. ADMET Predictor is an example of a commercial product that uses the Accelrys Metabolite Database to predict various metabolic stability values for a series of cytochromes (Simulations Plus 2010). Kirchmair et al. (2012) provide a comprehensive overview of methods for predicting sites of metabolism.

Knowledge-driven approaches, such as expert systems, allow extrapolation of structure to likely metabolites by using advanced reasoning rules and expert-system metabolite ranking, such as MetabolExpert, META, Meteor, and Metaprint2D-React. However, one problem is that they can generate a large number of metabolites from which it is difficult to determine which metabolites are the relevant and stable ones that should be considered. Other metabolism-predictive tools introduce the expert-system features with more refined computational algorithms to support the decision method and therefore limit the number of metabolites that are generated. Indeed, the software program Tissue Metabolism Simulator (TIMES) uses a comprehensive library of biotransformation information and a heuristic algorithm to generate plausible metabolic maps that are relevant to specific end points, such as skin sensitization or genotoxicity (Mekenyan et al. 2012; Patlewicz et al. 2014). Many of the TIMES metabolism simulators have been made freely available in the OECD QSAR Toolbox.

Limitations and Need for Improvement

Many tools for predicting physicochemical properties that are relevant for the evaluation of chemical disposition and distribution factors are available, but they are limited by their training sets.⁷ Such tools are generally most applicable for small organic chemicals—chemicals that have molecular weights of 500 Da or less (that is, not mixtures or polymers).

In vitro and *in silico* predictions of absorption for various routes of exposure are still crude, and current models might have little applicability to the Department of Defense (DOD).⁸ For oral absorption, Lipinski's rule of 5, which is based on experience in the drug-discovery world, might

⁷Training sets are data that are used to develop predictive models or tools.

⁸For example, *in vitro* assays for absorption (such as Caco-2 monolayer crossing) were developed primarily to predict systemic absorption after deliberate oral dosing (Artursson and Karlsson 1991). Thus, they are expected to be much less predictive for exposure routes (dermal and inhalation) that are more relevant to acute battlefield exposure. Likewise, crossing the blood–brain barrier is especially relevant for neurotoxicity, and although there are computational approaches for predicting blood–brain barrier penetration (Gerbtzoff and Seelig 2006; Carpenter et al. 2014), no *in vitro* assay accurately measures this property.

provide a convenient set of heuristics for chemicals of interest to DOD but would need to be evaluated to determine its applicability. The prediction of dermal permeability at the simplest level is illustrated by QSARs that predict $\log K_p$ or $\log J_{max}$ with such inputs as molecular weight and $\log K_{ow}$ as exemplified by Potts and Guy (1992) or Magnusson et al. (2004) (see also Fitzpatrick et al. 2004 and Mitragotri et al. 2011). Although refinements have been made to simulate penetration or systemic bioavailability (Dancik et al. 2013a,b), the underlying characteristics are still based largely on the heterogeneous dataset first compiled by Flynn (1990), which is limited in its coverage of chemicals.

Current metabolism-predictive approaches have several limitations. First, most of the existing metabolism-predictive tools were designed primarily to inform drug development. There are few examples in which such modeling tools have been used to evaluate volatile or lipophilic chemicals (Peyret and Krishnan 2012; Kirman et al. 2015). Thus, the available metabolism training sets will need to be expanded for the chemicals of interest. Second, although metabolism-predictive tools adequately predict transformations of various chemicals, they do a poor job of distinguishing differences in reactivity of closely related structural analogues. In most cases, the tools can only estimate the reactivity of the individual molecular sites. As a result, they have limited use for prioritizing a broad set of structurally related chemicals (Kirchmair et al. 2012). One interim solution that DOD might consider is to evaluate metabolites with known toxic effects and incorporate more metabolically competent test systems into its test battery. Chapter 4 describes each approach in some detail.

IN SILICO APPROACHES FOR PREDICTING TOXIC EFFECTS

In silico models incorporate a variety of physicochemical features that can be used to predict receptor binding, toxicity, and other biological outcomes. Many (Q)SAR models developed for use in toxicology have been built on a longstanding recognition that the physicochemical properties of a chemical, especially lipophilicity, are highly relevant to prediction of acute toxicity. It has been shown that the presence or absence of various physicochemical properties can be used to group chemicals into toxicity categories (Greene and Song 2011). That concept is well established and used in the pharmaceutical industry to reduce attrition in drug discovery, reduce toxicity, and improve the drug-likeness of chemicals (see Table 3-1).⁹ Indeed, several studies have shown how simple measures, such as $\log K_{ow}$ and total polar surface area (TPSA), provide useful indicators of potential toxicity in vivo. Hughes et al. (2008) showed for a dataset of 245 substances that substances that had low $\log K_{ow}$ and high TPSA were about 2.5 times more likely to be “clean” (nontoxic) than to be toxic. Precisely the reverse was true of chemicals that had high $\log K_{ow}$ and low TPSA, properties that increased the likelihood of chemical binding to multiple biological targets that could contribute to toxicity.

Although the chemical domain of concern for DOD goes beyond that of drug-like substances, an understanding of the type of physicochemical properties that can affect adverse outcomes and the range of property values for which the effect is likely to be substantial can offer useful insights for guiding the assessment of acute toxicity of chemicals of interest to DOD. The sections that follow describe in silico approaches that are available or could be developed for predicting acute toxicity that is relevant to DOD’s concerns.

Acute Oral Toxicity

There are a few (Q)SAR models and expert systems for prediction of in vivo acute toxicity. The predictiveness of the models, however, can be variable. For some chemicals, the models

⁹Drug-likeness refers to molecules that contain functional groups or have physical properties similar to those of known drugs (Walters and Murecko 2002).

provide predicted values that deviate by several orders of magnitude from the experimental data. Furthermore, efforts have focused largely on the prediction of acute rodent oral toxicity (see Appendix B for a description of various toxicity data sources). Fewer attempts have been made to derive models of acute toxicity via other routes of exposure, such as dermal or inhalation, although predictions based on extrapolation from acute oral LD₅₀ values have been attempted.

Available (Q)SARs for acute systemic toxicity have been reviewed (Cronin and Dearden 1995; Cronin et al. 2003; Lessigiarska et al. 2005; Tsakovska et al. 2006; Devillers and Devillers 2009; Lapenna et al. 2010). Several QSAR models have identified hydrophobicity and electronic and steric effects as important model parameters. Many literature-based models were developed for a single chemical class, such as alcohols, barbiturates, pyrimidines and their derivatives, and benzene derivatives. Examples are provided in Table 3-2.

In contrast, models based on heterogeneous datasets have typically been incorporated into expert systems. There are, however, examples of models based on heterogeneous data that have not been incorporated into expert systems, and there are examples of models that use a hybrid approach. Table 3-3 provides several examples of various types of models and tools. There has been an evolution in the types of (Q)SAR models developed over the years to predict acute toxicity. Expert systems tended to favor large datasets (global models) that use chemistry-based descriptors to derive estimates of rodent oral toxicity. Hybrid expert systems consider biological activity, such as cytotoxicity information as described in Chapter 4, and chemistry-based descriptors as inputs. More recently, there has been a return to local (Q)SAR models; they are integrated into batteries of (Q)SARs that can predict acute toxicity of diverse chemicals.

Acute Dermal Toxicity

To the committee's knowledge, there are no notable QSARs for the prediction of rodent dermal acute-toxicity values. Dermal LD₅₀ values might be estimated by extrapolating from oral LD₅₀ values in some cases by using toxicokinetic information. A case study of three cosmetic substances was performed by Gajewska et al. (2014) to evaluate such an extrapolation.

Moore et al. (2013) found that the toxicity of chemicals was usually greater by the oral route than the dermal route. They proposed that data on oral acute systemic toxicity could be used in lieu of equivalent dermal testing with little or no concern for underclassification according to the Globally Harmonized System of Classification and Labeling of Chemicals (GHS).¹⁰ For example, dermal testing of a substance that has an oral LD₅₀ of greater than 2,000 mg/kg

TABLE 3-2 Examples of (Q)SARs for Various Chemical Classes

Chemical Class	Description	Reference
Alcohols	Four molecular-structure descriptors and two indicator variables formed the basis of a categorical model that categorized 95 alcohols into ranges of LD ₅₀ values.	Guilian and Naibin 1998
Barbiturates	The number of valence electrons and logK _{ow} were found to be predictive of LD ₅₀ values of a set of 11 barbiturates.	Hansch and Kurup 2003
Pyrimidines and derivatives	The energy of the lowest unoccupied molecular orbital and logK _{ow} represented the descriptors used in a QSAR for pyrimidines and their derivatives.	Cronin et al. 2002
Benzene derivatives	Electronegativity, dipole moment, and the presence of nitrogen-containing groups were most important in predicting the acute oral toxicity of benzene derivatives.	Toropov et al. 2008

¹⁰The GHS is an internationally agreed-on system created by the UN to replace the various classification and labeling standards used in different countries by using consistent criteria for classification and labeling (UNECE 2015).

TABLE 3-3 Examples of Models and Tools for Predicting Acute Oral Toxicity

Model	Description	Outputs	Reference
Models Based on Heterogeneous Datasets That Have Been Incorporated into Expert Systems			
Toxicity Prediction by Komputer-Assisted Technology (TOPKAT)	19 QSAR regression models are based on a number of structural, topological, and electropological indexes and experimental values for 4,000 chemicals in the Registry of Toxic Effects of Chemical Substances (RTECS) database.	Rat LD ₅₀ (oral) Rat LC ₅₀ (inhalation)	TOPKAT, Accelrys (reviewed in Lapenna et al. 2010)
Toxicity Estimation Software Tool (TEST)	Based on chemicals in the RTECS database. Uses a variety of QSAR models (hierarchical method, Food and Drug Administration [FDA] method, single-model method, group-contribution method, nearest-neighbor method, and consensus method) to yield toxicity estimates.	Rat LD ₅₀ (oral)	EPA 2014
ACD/Labs Tox suite	Predictions are based on a combination of expert knowledge of various basal and extracellular effects (such as cholinesterase inhibition, ATP synthesis, and CNS disruption) and QSAR analysis of more than 10,000 chemicals. Predictions are provided with reliability estimations, and chemicals are classified into five toxicity categories.	Rat and mouse LD ₅₀ (routes include oral, intraperitoneal, subcutaneous, and intravenous)	ACD/Labs 2015
ProTox	Prediction is based on the analysis of the similarity of chemicals that have known median LD ₅₀ that are taken from a dataset of 38,000 chemicals and incorporates the identification of toxic fragments.	Rodent LD ₅₀ (oral)	Drwal et al. (2014); ProTox (2015)
Models Based on Heterogeneous Datasets That Have Not Been Incorporated into Expert Systems			
Consensus models	Use rodent in vivo acute oral data from the National Library of Medicine databases as reported in ChemIDplus. Predictions are based on different QSAR statistical techniques, including random forest, FDA MDL-QSAR program's approach to k-nearest neighbor, and hierarchical clustering.	Rodent LD ₅₀ (oral)	Zhu et al. (2009a)
Multiclassification methods	Based on a dataset containing 12,204 diverse chemicals and published acute oral rodent LD ₅₀ s. Model predictions obtained with machine-learning methods, such as support vector machine, C4.5 decision tree, random forest, k-nearest neighbor, naive Bayes algorithms, and MACCS and FP4 fingerprints.	Rat LD ₅₀ (oral)	Li et al. (2014)
Global Hybrid Approaches^a			
Tiered approach	All chemicals separated into two groups: one based on the relationship between the in vitro half-maximal inhibitory concentration (IC ₅₀) and rodent LD ₅₀ and the other contained the remaining chemicals. A two-step QSAR modeling approach was then applied. The derived binary classification QSAR models predicted group membership on the basis of the in vitro–in vivo relationships, and a second QSAR model estimated the LD ₅₀ s for the chemical subsets.	Rodent LD ₅₀ (oral)	Zhu et al. (2009b)

Chemical and biological descriptors	Dataset consisted of 67 chemicals obtained from the literature. Used structural information and in vitro basal cytotoxicity to predict human acute toxicity.	Indirect measures of human toxicity (e.g., LC ₅₀)	Lee et al. (2010)
Inclusion of concentration–response data derived from high-throughput screening	Used quantitative high-throughput screening concentration–response data to complement traditional chemical descriptors in the modeling of acute oral rodent LD ₅₀ s.	Rodent LD ₅₀ (oral)	Sedykh et al. (2011)
Local Hybrid Approaches^b			
OASIS Pipeline	Relies on a baseline model for neutral organic substances supplemented with mechanistic SARs for different reaction–chemistry domains. The approach is underpinned by 3-D QSARs where relevant to predict acute oral-toxicity categories by using the same RTECS dataset as used by Zhu et al. (2009a). Complements the approach outlined by Koleva et al. (2011).		Mekenyan et al. (personal communication, December 2014)
Local lazy method	Based on a dataset of 9,617 chemicals. Uses local structure–toxicity relationships associated with a query substance to develop acute oral LD ₅₀ models.	Rodent LD ₅₀ (oral)	Lu et al. (2014)

^aGlobal hybrid approaches use models derived on the basis of a large heterogeneous dataset that includes chemical and biological information.

^bLocal hybrid approaches aim to derive models that are specific to chemical class or reaction chemistry but will still be applicable to a broad spectrum of chemicals.

would provide no added value for categorizing its hazard. Moore et al. (2013), however, reported that a majority of chemicals that they evaluated would be classified more stringently if oral classifications were applied directly to the dermal route. One approach to address the tendency for overclassification would be to consider whether a chemical is absorbed by the skin.

Acute Inhalation Toxicity

There are only a handful of QSARs for acute inhalation toxicity. One example is that for volatile substances. Veith et al. (2009) derived a baseline narcosis model that related vapor pressure (as logVP in millimeters of mercury) to the 4-hour molar logLC₅₀ in rodents for neutral organic substances. Veith and Wallace (2006) also established relationships for reactive (electrophilic) chemicals in which reactivity, as measured in the glutathione depletion assay (Schultz et al. 2005), was related to the molar logLC₅₀. The underlying basis of their strategy mimics the Adverse Outcome Pathway (AOP) construct as described by Ankley et al. (2010).

The expert system TOPKAT incorporates a global model for the prediction of acute inhalation toxicity. The rat inhalation LC₅₀ module contains five models related to different chemical classes to cover a reasonable breadth of chemical coverage.

Neurotoxicity

Neurotoxicity is another debilitating effect associated with some acute exposures. A few QSARs have been derived for neurotoxicity, but they are quite limited in scope. Cronin (1996) derived a neurotoxicity QSAR for a set of 44 common solvents that depended on logK_{ow} and membrane permeability. A number of modeling approaches to derive QSAR models for organophosphorus pesticides have also been developed (Devillers 2004; Garcia-Domenech et al. 2007).

A handful of SARs exist that might identify structural alerts for neurotoxic potential. Chemical classes associated with neurotoxicity include some organic solvents, organophosphorus chemicals, and carbamates, which can induce chronic toxic encephalopathy, delayed neurotoxicity, and cholinergic effects, respectively. For example, Derek Nexus includes the following structural alerts: organophosphate (for direct and indirect anticholinesterase activity), *N*-methyl or *N,N*-dimethyl carbamate (for direct anticholinesterase activity), and gamma-diketones (for neurotoxicity) (ECHA 2014).

Neurotoxicity is clearly an effect whose mechanisms need to be better elucidated, and models developed accordingly.

Cytotoxicity

Ekwall (1983) suggested that for most chemicals, toxicity was a consequence of nonspecific alterations in cellular function; thus, evaluating the cytotoxic potential of chemicals with cytotoxicity assays could provide an indication of their potential *in vivo* toxicity. There have been many attempts to explore the correlation between *in vivo* acute toxicity and cytotoxicity data and a number of efforts to predict cytotoxicity from chemical structure. For example, as part of the Multicenter Evaluation of *In Vitro* Cytotoxicity program (Ekwall et al. 1998), 50 reference chemicals were tested in 61 cytotoxicity assays in the hope of predicting acute oral LD₅₀s. The coefficients of determination (r^2) were 0.61 for rat LD₅₀s and 0.65 for mouse LD₅₀s. And, Lesigarska et al. (2006) demonstrated how acute toxicity in rats, mice, and humans could be predicted by using QSAR models that incorporated cytotoxicity data, other biological end points, and chemical structural descriptor data.

A host of QSAR models have been derived to predict cytotoxicity of various chemical classes. In many cases, hydrophobicity was a predominant descriptor related to cytotoxicity. For example, McKarns et al. (1997) correlated the loss of membrane integrity in rat liver epithelial cells with hydrophobicity as modeled by logK_{ow} for a series of 11 alcohols. Other QSARs, as

summarized by Tsakovska et al. (2006), have been developed for *p*-substituted benzyl alcohols, phenols, anilines, chlorophenols, and polybrominated diphenyl ethers. Papa et al. (2009) developed QSARs for three toxicological end points: mouse oral LD₅₀ values, inhibition of NADH oxidase (EC₅₀), and effect on mitochondrial membrane potential (EC₅₀). Freidig et al. (2007) found that nonspecific cytotoxicity could help to identify irritant chemicals.

As part of the European Union framework program, ACuteTox, many investigations were performed to explore in vivo–in vitro relationships. Clothier et al. (2013) used Spearman rank-correlation analysis and hierarchical-cluster analysis to identify in vitro testing strategies for predicting acute toxicity. Classification-based and regression-based quantitative structure–toxicity relationship (QSTR) and toxicophore models were developed by Kar and Roy (2013). They used in vitro cytotoxicity data collected from the ACuteTox database.¹¹ Their QSTR models showed that cytotoxicity was influenced by the presence of hydrophobic aliphatic groups, a ring aromatic group, and hydrogen-bond donors. The in silico models derived were considered capable of identifying the essential structural attributes and quantifying the molecular properties that drive in vitro basal cytotoxicity. Prieto et al. (2013) proposed a heuristic testing strategy for identifying potential neurotoxicants that considered octanol–water partition coefficients, the prediction results from the neutral red uptake assay performed in 3T3 cells, and in silico predictions of intestinal absorption and blood–brain barrier passage.

Limitations and Need for Improvement

The greatest focus in the literature has been on deriving QSAR models to predict oral rodent LD₅₀s. There are some models for specific chemical classes, but there has been greater interest in exploring the feasibility of deriving global models. Many of the global models have been data-driven, although some attempts have included consideration of a chemical mechanistic approach akin to that described for acute fish toxicity (Bradbury et al. 1990; Schultz et al. 2006). More recently and in part as stimulated by work within the ACuteTox program, predicting in vivo acute toxicity has considered the use of (Q)SAR approaches in conjunction with in vitro cytotoxicity data. This shift of integrating in vitro and in silico approaches is consistent with the framework of AOPs as one means of incorporating more mechanistic information in testing and assessment approaches for different purposes.

A key issue in all approaches is their relevance and applicability to DOD chemicals. The relevance of the (Q)SAR models cited earlier would need to be probed for the types of chemicals under consideration by DOD to determine the extent to which the existing models are appropriate in light of the applicability domain and the decision context in question. If the substances of interest are entirely or mostly out of the applicability domain, new data might need to be identified or other primary sources exploited to collate and compile more relevant information for the derivation of new models or refinement of existing models. It will also be critical in such an evaluation to consider the robustness of the training set and associated data that are used to develop the (Q)SAR model. The OECD principles for the validation of (Q)SAR provide a convenient framework for assessing the validity and applicability of (Q)SAR models (OECD 2007b).

A second issue is the nonavailability of tools to assess toxicity by nonoral routes of exposure. As discussed above, most tools have been developed to evaluate the oral exposure route.

A third issue is that the existing (Q)SARs are often lacking in mechanistic basis. Future (Q)SARs could conceivably be derived to predict key events as reflected in Figure 3-1. (Q)SARs would be developed to estimate outcomes of initial or intermediate events within a pathway rather than to predict an adverse outcome directly as indicated in Figure 3-1. An example of that approach is the early work of the ACuteTox program in which QSARs were derived to estimate in vitro cytotoxicity rather than in vivo acute rat toxicity.

¹¹See <http://www.acutetox.eu/>.

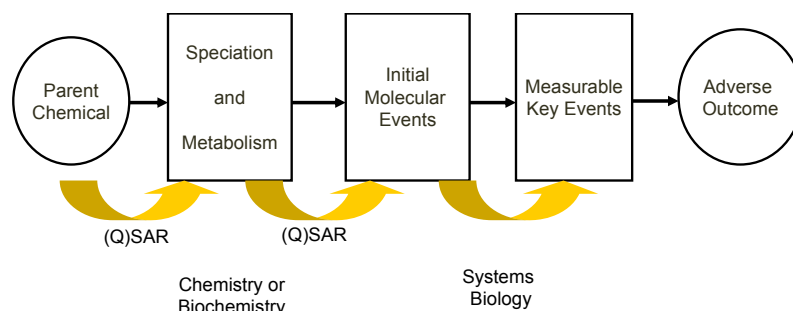


FIGURE 3-1 Conceptual framework for the future development of (Q)SARs.

Another interesting example of exploiting the framework shown in Figure 3-1 focuses on mitochondrial inhibition, a mechanism known to drive acute toxicity of some chemicals. Bhhata-ri et al. (2014) compared data from high-throughput screening assays of mitochondrial toxicity in HepG2 cells (see Attene-Ramos et al. 2015) with *in silico* data on absorption and first-pass metabolism¹² and obtained promising results for predicting acute toxicity in multiple species. They found that mitochondrial inhibition predicted the minimum toxicity in fish and daphnia and that the lower the assay AC_{50} ,¹³ the more likely that the toxicity was driven by mitochondrial toxicity. However, mitochondrial inhibition did not often predict the toxicity of chemicals in rats because of the lack of data on oral bioavailability and first-pass metabolism. However, simulations that used *in silico* models for bioavailability and metabolism did improve toxicity predictions. A similar approach has been put forward by LHASA Ltd as a contribution to the OECD AOP work program (OECD 2011). A set of SARs has been derived from substances that inhibit complexes I, III, IV, and V of the electron transport chain, which characterize molecular initiating events in the pathway that leads to mitochondrial toxicity.

Another route by which models could be derived would involve integrating data from a variety of inputs (such as *in vitro* IC_{50} s or AC_{50} s and rodent LD_{50} s) to predict acute toxicity. Several such examples have been investigated as part of the ACuteTox program as described above. Clothier et al. (2013) used Spearman rank-correlation analysis and hierarchical clustering to help to identify a combination of *in vitro* test systems for predicting *in vivo* acute toxicity. Kinsner-Ovaskainen et al. (2013) used classification and regression-tree analysis of *in vitro* data from the ACuteTox program to predict acute oral-toxicity categories. Kopp-Schneider et al. (2013) likewise investigated various data-mining approaches with the ACuteTox program data.

A second alternative route of developing (Q)SARs would consider the biological pathways involved in acute debilitating toxicity (such as altered oxygen transport, changes in neurotransmitter function, and disruption of cytoskeleton), which could be elucidated in a construct based on mechanistic pathways, and appropriate assays could be mapped to the key biological events. This approach would provide a different basis for development of integrated approaches, including specific (Q)SAR models that address specific key biological events of the AOP. The chemical applicability domain of the assays that characterize each key event could be extracted to inform new SARs that would facilitate profiling of untested substances. This type of approach has been attempted for skin sensitization (Patlewicz et al. 2014). Before such a strategy can be used, an approach to assessing the validity of the assays (Chapter 4), of the prediction models (data integration models) (Chapter 5), and of the pathways (Chapter 2) needs to be established (see Patlewicz et al. 2015).

¹²First-pass metabolism can occur at the site of chemical absorption (for example, in the gastrointestinal tract) or in the liver. In general, first-pass metabolism reduces the amount of a chemical that reaches the systemic circulation.

¹³ AC_{50} is the concentration required to elicit a 50% response in an *in vitro* assay.

ENSURING SCIENTIFIC CONFIDENCE IN (Q)SAR MODELS

Ensuring scientific confidence in a (Q)SAR model relies on an assessment of model validity and model applicability. Both are critical for the appropriate interpretation and use of the predictions derived. The OECD validation principles for (Q)SARs provide a useful construct for evaluating and characterizing a given (Q)SAR model to determine its scientific validity. The five principles describe the need for a defined end point, an unambiguous algorithm, a defined domain of applicability, measures of performance, and a mechanistic interpretation if possible (OECD 2004, 2007b). The applicability domain,¹⁴ which involves extracting an applicability domain on the basis of the training set used to derive the QSAR model, provides a basis for judging the relevance and reliability of a prediction made for a given target substance. There are many ways to extract an applicability domain. Typical approaches include structural, mechanistic, metabolic, and descriptor considerations (Netzeva et al. 2005; Dimitrov et al. 2005). Freely available and commercial tools also exist to determine an applicability domain, namely, AMBIT Discovery and Domain Manager (Nikolova-Jeliazkova and Jaworska 2005; Patlewicz et al. 2011). Sazonovas et al. (2010) presented an alternative approach to extracting an applicability domain for an LD₅₀ model. Each prediction was associated with a reliability index that depended on the target's similarity to the training set and the consistency of experimental results with regard to the baseline model in the local chemical environment.

The applicability domain forms only one facet in judging the relevance and reliability of a prediction. To ensure the reliability of a prediction, one also needs to evaluate how well the models make correct predictions for similar chemicals. Such similar chemicals might be identified on the basis of the same characteristics or descriptors that were used to derive the original QSAR model or on the basis of structural similarity. The predictivity of similar analogues will form a second facet of judging the relevance of a model for the chemical of interest. The framework outlined in the REACH guidance (see Figure 3-2) might be helpful in summarizing the key considerations for the assessment of a (Q)SAR model and its associated prediction.

FINDINGS AND RECOMMENDATIONS

- **Finding:** Multiple databases are available for performing an initial chemical characterization of molecular structure, physicochemical properties, and available acute toxicity data. In addition, a number of in silico models are available for predicting physicochemical properties.
- **Finding:** Physicochemical and structural properties are critical for chemical characterization in that they can help to predict a chemical's potential to pose a physical hazard, its reactivity, and its pharmacokinetic characteristics, such as bioavailability and likely routes of exposure.
- **Finding:** Although a number of tools are available to predict the site of metabolism and likely metabolic products on the basis of chemical structure and physicochemical properties, they are designed largely for pharmaceutical agents. Information about the likely metabolic products will be critical for informing experimental design, including assay selection.
- **Finding:** Several (Q)SAR models that use structural properties or physicochemical properties are available for predicting acute oral LD₅₀s. Few models for predicting inhalation LC₅₀s are available, and none for predicting dermal LD₅₀s was identified.
- **Finding:** A few QSAR models for predicting neurotoxicity and cytotoxicity are available, but not for other end points that are relevant for acute, debilitating toxicity. Current research in (Q)SAR models is focusing on incorporating more biological information, such as integrating in vitro data (for example, on cytotoxicity) and information on specific AOPs.

¹⁴The applicability domain is the array of chemicals for which the (Q)SAR can confidently be applied for purposes of toxicity prediction (Aptula and Roberts 2006).

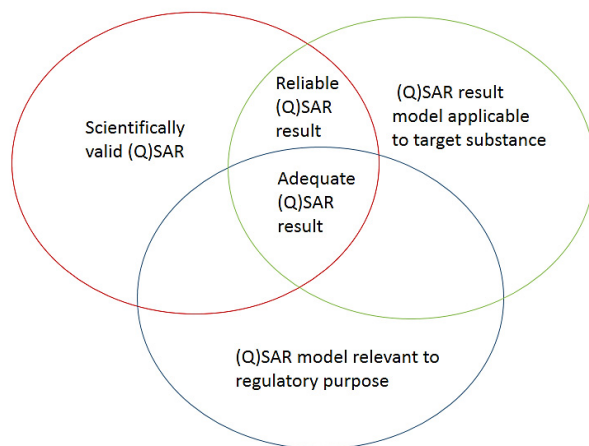


FIGURE 3-2 Key elements associated with evaluating the adequacy of a (Q)SAR model and its prediction as adapted from the REACH technical guidance.

- **Recommendation:** DOD should evaluate the applicability, relevance, and reliability of available (Q)SAR models to meet its needs of assessing a chemical's potential to cause acute, debilitating toxicity. OECD and REACH principles and guidelines for evaluating and characterizing (Q)SAR models might provide useful frameworks for conducting such an evaluation. Furthermore, DOD should evaluate the applicability, relevance, and reliability of models and tools for predicting physicochemical properties and metabolism.

- **Recommendation:** To fill remaining gaps in nontesting approaches, DOD should consider a number of options for further research and development, including extrapolation of oral LD₅₀ to other exposure routes through pharmacokinetic models; development of new (Q)SAR models for acute lethality, focusing particularly on inhalation and dermal exposure; and development of (Q)SAR models augmented with biological information, such as in vitro data and information on targets or mechanisms of acute toxicity.

REFERENCES

- ACD/Labs. 2015. Acute Toxicity Prediction Module. Data Sheet [online]. Available: http://www.acdlabs.com/download/docs/datasheets/datasheet_acute.pdf [accessed March 24, 2015].
- AltTox. 2014. Table of Validated and Accepted Alternative Methods [online]. Available: <http://alttox.org/mapp/table-of-validated-and-accepted-alternative-methods/> [accessed March 18, 2015].
- Ankley, G.T., R.S. Bennett, R.J. Erickson, D.J. Hoff, M.W. Hornung, R.D. Johnson, D.R. Mount, J.W. Nichols, C.L. Russom, P.K. Schmieder, J.A. Serrano, J.E. Tietge, and D.L. Villeneuve. 2010. Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environ. Toxicol. Chem.* 29(3):730-741.
- Aptula, A.O., and D.W. Roberts. 2006. Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: General principles and application to reactive toxicity. *Chem. Res. Toxicol.* 19(8):1097-1105.
- Artursson, P., and J. Karlsson. 1991. Correlation between oral drug absorption in humans and apparent drug permeability coefficients in human intestinal epithelial (Caco-2) cells. *Biochem. Biophys. Res. Commun.* 175(3):880-885.
- Attene-Ramos, M.S., R. Huang, S. Michael, K.L. Witt, A. Richard, R.R. Tice, A. Simeonov, C.P. Austin, and M. Xia. 2015. Profiling of the Tox21 chemical collection for mitochondrial function to identify compounds that acutely decrease mitochondrial membrane potential. *Environ. Health Perspect.* 123(1):49-56.

- Barratt, M.D. 1996. Quantitative structure-activity relationships for skin irritation and corrosivity of neutral and electrophilic organic chemicals. *Toxicol. In Vitro* 10(3):247-256.
- Bhhatari, B., D. Wilson, P.S. Price, M.J. Bartels, S. Chaunhuri, and E.W. Carney. 2014. Predicting Acute Toxicity Using In Vitro ToxCast HTS Mitochondrial Inhibition Assays. Presented at the Second ToxCast Data Summit, September 29-30, 2014, Durham, NC.
- Bradbury, S.P., T.R. Henry, and R.W. Carlson. 1990. Fish acute toxicity syndromes in the development of mechanism specific QSARs. Pp. 295-315 in *Practical Applications of Quantitative Structure Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*, W. Karcher, and J. Devillers, eds. Dordrecht, The Netherlands: Kluwer.
- Brown, J.S., T. Gordon, O. Price, and B. Asgharian. 2013. Thoracic and respirable particle definitions for human health risk assessment. Part. *Fibre Toxicol.* 10:12.
- Carpenter, T.S., D.A. Kirshner, E.Y. Lau, S.E. Wong, J.P. Nilmeier, and F.C. Lightstone. 2014. A method to predict blood-brain barrier permeability of drug-like compounds using molecular dynamics simulations. *Biophys. J.* 107(3):630-641.
- Castillo-Garit, J.A., Y. Marrero-Ponce, F. Torrens, and R. García-Domenech. 2008. Estimation of ADME properties in drug discovery: Predicting Caco-2 cell permeability using atom-based stochastic and non-stochastic linear indices. *J. Pharm. Sci.* 97(5):1946-1976.
- Clothier, R., M.J. Gómez-Lechón, A. Kinsner-Ovaskainen, A. Kopp-Schneider, J.E. O'Connor, P. Prieto, and S. Stanzel. 2013. Comparative analysis of eight cytotoxicity assays evaluated within the ACute-Tox Project. *Toxicol. In Vitro* 27(4):1347-1356.
- Cronin, M.T. 1996. Quantitative structure-activity relationship (QSAR) analysis of the acute sublethal neurotoxicity of solvents. *Toxicol. In Vitro* 10(2):103-110.
- Cronin, M.T., and J.C. Dearden. 1995. QSAR in toxicology. 2. Prediction of acute mammalian toxicity and interspecies correlations. *Quant. Struct. Act. Relat.* 14(2):117-120.
- Cronin, M.T., and J.C. Madden, eds. 2010. *In Silico Toxicology: Principles and Applications*. London: Royal Society of Chemistry.
- Cronin, M.T., J.C. Dearden, J.C. Duff, R. Edwards, N. Manga, A.P. Worth, and A.D. Worgan. 2002. The importance of hydrophobicity and electrophilicity descriptors in mechanistically based QSARs for toxicological endpoints. *SAR QSAR Environ. Res.* 13(1):167-176.
- Cronin, M.T., J.C. Dearden, J.D. Walker, and A.P. Worth. 2003. Quantitative structure activity relationships for human health effects: Commonalities with other endpoints. *Environ. Toxicol. Chem.* 22(8):1829-1843.
- Dancik, Y., M.A. Miller, J. Jaworska, and G.B. Kasting. 2013a. Design and performance of a spreadsheet-based model for estimating bioavailability of chemicals from dermal exposure. *Adv. Drug Deliv. Rev.* 65(2):221-236.
- Dancik, Y., J.A. Troutman, and J. Jaworska. 2013b. A framework incorporating the impact of exposure scenarios and application conditions on risk assessment of chemicals applied to skin. *In Silico Pharmacol.* 1:10; doi: 10.1186/2193-9616-1-10.
- Dearden, J., and A. Worth. 2007. *In Silico Prediction of Physicochemical Properties*. JRC Scientific and Technical Reports EUR 23051[online]. Available: https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/doc/EUR_23051_EN.pdf [accessed March 16, 2015].
- Devillers, J. 2004. Prediction of mammalian toxicity of organophosphorus pesticides from QSTR modeling. *SAR QSAR Environ. Res.* 15(5-6):501-510.
- Devillers, J., and H. Devillers. 2009. Prediction of acute mammalian toxicity from QSARs and interspecies correlation. *SAR QSAR Environ. Res.* 20(5-6):467-500.
- Dimitrov, S., G. Dimitrova, T. Pavlov, N. Dimitrova, G. Patlewicz, J. Niemela, and O. Mekenyan. 2005. A stepwise approach for defining the applicability domain of SAR and QSAR models. *J. Chem. Inf. Model.* 45(4):839-849.
- Drwal, M.N., P. Banerjee, M. Dunkel, M.R. Wettig, and R. Preissner. 2014. ProTox: A web server for the in silico prediction of rodent oral toxicity. *Nucleic Acids Res.* 42(Web Server issue):W53-W58.
- ECETOC (European Centre for Ecotoxicology and Toxicology of Chemicals), 2012. *Category Approaches, Read-across, (Q)SAR*. Technical Report No. 116. Brussels, Belgium: ECETOC.
- ECHA (European Chemicals Agency). 2008. *Guidance on Information Requirements and Chemical Safety Assessment*. Chapter R.6: QSARs and Grouping of Chemicals [online]. Available: http://echa.europa.eu/documents/10162/13632/information_requirements_r6_en.pdf [accessed February 12, 2015].
- ECHA (European Chemicals Agency). 2012. *Guidance on Registration, Version 2.0*. *Guidance for the Implementation of REACH*. May 2012 [online]. Available: http://echa.europa.eu/documents/10162/13632/registration_en.pdf [accessed April 30, 2015].

- ECHA (European Chemicals Agency) 2014. Guidance on Information Requirements and Chemical Safety Assessment, Chapter R.7a. Endpoint Specific Guidance, Version 3.0 [online]. Available: http://echa.europa.eu/documents/10162/13632/information_requirements_r7a_en.pdf.
- Ekwall, B. 1983. Screening of toxic compounds in mammalian cell cultures. *Ann. N.Y. Acad. Sci.* 407:64-77.
- Ekwall, B., F.A. Barile, A. Castano, C. Clemedson, R.H. Clothier, P. Dierickx, B. Ekwall, M. Ferro, G. Fiskesjö, L. Garza-Ocanas, M.J. Gómez-Lechón, M. Gülđen, T. Hall, B. Isomaa, A. Kahru, G. Kerszman, U. Kristen, M. Kunitomo, S. Kärenlampi, L. Lewan, A. Loukianov, T. Ohno, G. Persoone, L. Romert, T.W. Sawyer, H. Segner, R. Shrivastava, A. Stamatii, N. Tanaka, M. Valentino, E. Walum, and F. Zucco. 1998. MEIC evaluation of acute systemic toxicity. Part VI. Prediction of human toxicity by rodent LD50 values and results from 61 in vitro tests. *ATLA* 26(Suppl. 2):617-658.
- EPA (US Environmental Protection Agency). 2012. Glossary of Terms: Methods of Toxicity Testing and Risk Assessment [online]. Available: <http://www.epa.gov/pesticides/science/comptox-glossary.html#q> [accessed May 7, 2015].
- EPA (US Environmental Protection Agency). 2014. Toxicity Estimation Software Tool (TEST) [online]. Available: <http://www.epa.gov/nrmrl/std/qsar/qsar.html#TEST> [accessed March 18, 2015].
- EU (European Union). 2006. Regulation (EC) No. 1907/2006 of the European Parliament and of the Council concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No. 793/93 and Commission Regulation (EC) No. 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. OJ E.U. 396:11-1849 [online]. Available: <http://faolex.fao.org/docs/pdf/eur68317.pdf>.
- Fitzpatrick, D., J. Corish, and B. Hayes. 2004. Modelling skin permeability in risk assessment--the future. *Chemosphere* 55(10):1309-1314.
- Flynn, G.L. 1990. Physicochemical determinants of skin absorption. Pp. 93-127 in *Principles of Route-to-Route Extrapolation for Risk Assessment*, T.R. Gerrity, and C.J. Henry, eds. New York: Elsevier.
- Freidig, A.P., S. Dekkers, M. Verwei, E. Zvinavashe, J.G. Bessems, and J.J. van de Sandt. 2007. Development of a QSAR for worst case estimates of acute toxicity of chemically reactive compounds. *Toxicol. Lett.* 170(3):214-222.
- Gajewska, M., A. Worth, C. Urani, H. Briesen, and K.W. Schramm. 2014. Application of physiologically-based toxicokinetic modelling in oral-to-dermal extrapolation of threshold doses of cosmetic ingredients. *Toxicol. Lett.* 227(3):189-202.
- García-Domenech, R., P. Alarcón-Elbal, G. Bolas, R. Bueno-Marí, F.A. Chordá-Olmos, S.A. Delacour, M.C. Mouriño, A. Vidal, and J. Gálvez. 2007. Prediction of acute toxicity of organophosphorus pesticides using topological indices. *SAR QSAR Environ. Res.* 18(7-8):745-755.
- Gerebtzoff, G., and A. Seelig. 2006. In silico prediction of blood-brain barrier permeation using the calculated molecular cross-sectional area as main parameter. *J. Chem. Inf. Model.* 46(6):2638-2650.
- Gerner, I., M.D. Barratt, S. Zinke, K. Schlegel, and E. Schleder. 2004. Development and prevalidation of a list of structure-activity relationship rules to be used in expert systems for prediction of the skin-sensitizing properties of chemicals. *ATLA* 32(5):487-509.
- Gleeson, M.P. 2008. Generation of a set of simple, interpretable ADMET rules of thumb. *J. Med. Chem.* 51(4):817-834.
- Grbic, S., J. Parojcic, S. Ibric, and Z. Djuric. 2011. In vitro-in vivo correlation for gliclazide immediate-release tablets based on mechanistic absorption simulation. *AAPS PharmSciTech.* 12(1):166-171.
- Green, N., and M. Song. 2011. Predicting in vivo safety characteristics using physiochemical properties and in vitro assays. *Fut. Med. Chem.* 3(12):1503-1511.
- Guerra, A., N.E. Campillo, and J.A. Páez. 2010. Neural computational prediction of oral drug absorption based on CODES 2D descriptors. *Eur. J. Med. Chem.* 45(3):930-940.
- Guilian, W., and B. Naibin. 1998. Structure activity relationships for rat and mouse LD50 of miscellaneous alcohols. *Chemosphere* 36(7):1475-1483.
- Hansch, C., and A. Kurup. 2003. QSAR of chemical polarisability and nerve toxicity, 2. *J. Chem. Inf. Comp. Sci.* 43(5):1647-1651.
- Hostynek, J.J., and P.S. Magee. 1997. Modelling in vivo human skin absorption. *Quant. Struct. Act. Relat.* 16(6):473-479.
- Hughes, J.D., J. Blagg, D.A. Price, S. Bailey, G.A. DeCrescenzo, R.V. Devraj, E. Ellsworth, Y. Fobian, M.E. Gibbs, R.W. Gilles, N. Greene, E. Huang, T. Krieger-Burke, T. Wager, L. Whiteley, and Y. Zhang. 2008. Physiochemical drug properties associated with in vivo toxicological outcomes. *Biorg. Med. Chem. Lett.* 18(17):4872-4875.

- Hulzebos, E., J.D. Walker, I. Gerner, and K. Schlegel. 2005. Use of structural alerts to develop rules for identifying chemical substances with skin irritation or skin corrosion potential. *QSAR Comb. Sci.* 24(3):332-342.
- Iyer, M., Y.J. Tseng, C.L. Senese, J. Liu, and A.J. Hopfinger. 2007. Prediction and mechanistic interpretation of human oral drug absorption using MI-QSAR analysis. *Mol. Pharm.* 4(2):218-231.
- Kar, S., and K. Roy. 2013. First report on predictive chemometric modeling, 3D-toxicophore mapping and in silico screening of in vitro basal cytotoxicity of diverse organic chemicals. *Toxicol. In Vitro.* 27(2):597-608.
- Kinsner-Ovaskainen, A., P. Prieto, S. Stanzel, and A. Kopp-Schneider. 2013. Selection of test methods to be included in a testing strategy to predict acute oral toxicity: An approach based on statistical analysis of data collected in phase I of the ACuteTox project. *Toxicol. In Vitro* 27(4):1377-1394.
- Kirchmair, J., M.J. Williamson, J.D. Tyzack, L. Tan, P.J. Bond, A. Bender, and R.C. Glen. 2012. Computational prediction of metabolism: Sites, products, SAR, P450, enzyme dynamics and mechanism. *J. Chem. Inf. Model.* 52(3):617-648.
- Kirman, C.R., L.L. Aylward, B.A. Wetmore, R.S. Thomas, M. Sochaski, S.S. Ferguson, S.A. Csiszar, and O. Jolliet. 2015. Quantitative property-property relationship for screening-level prediction of intrinsic clearance: A tool for exposure modeling for high-throughput toxicity screening data. *Appl. In Vitro Toxicol.* 1(2):140-146.
- Koleva, Y.K., M.T. Cronin, J.C. Madden, and J.A. Schwöbel. 2011. Modelling acute oral mammalian toxicity. 1. Definition of a quantifiable baseline effect. *Toxicol. In Vitro* 25(7):1281-1293.
- Kopp-Schneider, A., P. Prieto, A. Kinsner-Ovaskainen, and S. Stanzel. 2013. Design of a testing strategy using non-animal based test methods: Lessons learnt from the ACuteTox project. *Toxicol. In Vitro* 27(4):1395-1401.
- Kuentz, M., S. Nick, N. Parrott, and D. Röthlisberger. 2006. A strategy for preclinical formulation development using GastroPlus as pharmacokinetic simulation tool and a statistical screening design applied to a dog study. *Eur. J. Pharm. Sci.* 27(1):91-99.
- Langer, T., and G. Wolber. 2004. Pharmacophore definition and 3D searches. *Drug Discov. Today Technol.* 1(3):203-207.
- Lapenna, S., M. Fuart-Gatnik, and A. Worth. 2010. Review of QSAR Models and Software Tools for Predicting Acute and Chronic Systemic Toxicity. JRC Technical Report EUR 24639 EN. Luxembourg: Office of the European Union.
- Lee, S., K. Park, H.S. Ahn, and D. Kim. 2010. Importance of structural information in predicting human acute toxicity from in vitro cytotoxicity data. *Toxicol. Appl. Pharmacol.* 246(1-2):38-48.
- Leeson, P.D., and B. Springthorpe. 2007. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discov.* 6(11):881-890.
- Lessigiarska, I., A.P. Worth, and T.I. Netzeva. 2005. Comparative Review of QSARs for Acute Toxicity. JRC report EUR 21559 EN. European Commission, Joint Research Centre, Ispra, Italy.
- Lessigiarska, I., A.P. Worth, T.I. Netzeva, J.C. Dearden, and M.T. Cronin. 2006. Quantitative structure-activity-activity and quantitative structure-activity investigations of human and rodent toxicity. *Chemosphere* 65(10):1878-1887.
- Li, X., L. Chen, F. Cheng, Z. Wu, H. Bian, C. Xu, W. Li, G. Liu, X. Shen, and Y. Tang. 2014. In silico prediction of chemical acute oral toxicity using multi-classification methods. *J. Chem. Inf. Model.* 54(4):1061-1069.
- Lipinski, C.A., F. Lombardo, B.W. Dominy, and P.J. Feeney. 2001. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 46(1-3):3-26.
- Lu, J., J. Peng, J. Wang, Q. Shen, Y. Bi, L. Gong, M. Zheng, X. Luo, W. Zhu, H. Jiang, and K. Chen. 2014. Estimation of acute oral toxicity in rat using local lazy learning. *J. Cheminform.* 6:26; doi:10.1186/1758-2946-6-26.
- Magnusson, B.M., Y.G. Anissimov, S.E. Cross, and M.S. Roberts. 2004. Molecular size as the main determinant of solute maximum flux across the skin. *J. Invest. Dermatol.* 122(4):993-999.
- McKarns, S.C., C. Hansch, W.S. Caldwell, W.T. Morgan, S.K. Moore, and D.J. Doolittle. 1997. Correlation between hydrophobicity of short chain aliphatic alcohols and their ability to alter plasma membrane integrity. *Fundam. Appl. Toxicol.* 36(1):62-70.
- Meek, M.E., H.A. Barton, J.G. Bessems, J.C. Lipscomb, and K. Krishnan. 2013. Case study illustrating the WHO IPCS guidance on characterization and application of physiologically based pharmacokinetic models in risk assessment. *Regul. Toxicol. Pharmacol.* 66(1):116-129.

- Mekenyan, O.G., P.I. Petkov, S.V. Kotov, S. Stoeva, V.B. Kamenska, S.D. Dimitrov, M. Honma, M. Hayashi, R. Benigni, E.M. Donner, and G. Patlewicz. 2012. Investigating the relationship between in vitro-in vivo genotoxicity: Derivation of mechanistic QSAR models for in vivo liver genotoxicity and in vivo bone marrow micronucleus formation which encompass metabolism. *Chem. Res. Toxicol.* 25(2):277-296.
- Mitragotri, S., Y.G. Anissimov, A.L. Bunge, H.F. Frasch, R.H. Guy, J. Hadgraft, G.B. Kasting, M.E. Lane, and M.S. Roberts. 2011. Mathematical models of skin permeability: An overview. *Int. J. Pharm.* 418(1):115-129.
- Moore, N.P., D.J. Andrew, D.L. Bjerke, S. Creton, D. Dreher, T. Homes, P. Prieto, T. Seidle, and T.G. Rowan. 2013. Can acute dermal systematic toxicity tests be replaced with oral tests? A comparison of route specific systemic toxicity and hazard classifications under the Globally Harmonised System of Classification and Labelling of Chemicals (GHS). *Regul. Toxicol. Pharmacol.* 66(1):30-37.
- Netzeva, T., A. Worth, T. Aldenberg, R. Benigni, M. Cronin, P. Gramatica, J. Jaworska, S. Kahn, G. Klopman, C. Marchant, G. Myatt, N. Nikolova-Jeliazkova, G. Patlewicz, R. Perkins, D. Roberts, T. Schultz, D. Stanton, J. van de Sandt, W. Tong, G. Veith, and C. Yang. 2005. Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. *ATLA* 33(2):155-173.
- Nikolova-Jeliazkova, N., and J. Jaworska. 2005. An approach to determining applicability domains for QSAR group contribution models: An analysis of SRC KOWWIN. *ATLA* 33(5):461-470.
- NIOSH (National Institute for Occupational Safety and Health). 2013. Skin Exposures and Effects: Finite Dose Skin Permeation Calculator [online]. Available: <http://www.cdc.gov/niosh/topics/skin/finiteSkinPermCalc.html> [accessed March 18, 2015].
- NRC (National Research Council). 2014. Physicochemical properties and chemical fate. Pp. 47-67 in *A Framework to Guide Selection of Chemical Alternatives*. Washington, DC: National Academies Press.
- OECD (Organisation for Economic Co-operation and Development). 2004. OECD Principles for the Validation, For Regulatory Purposes of (Quantitative) Structure-Activity Relationship Models [online]. Available: <http://www.oecd.org/chemicalsafety/risk-assessment/37849783.pdf> [accessed March 17, 2015].
- OECD (Organisation for Economic Co-operation and Development). 2006. OECD Guideline for Testing of Chemicals No. 435: In Vitro Membrane Barrier Test Method for Skin Corrosion. Adopted July 19, 2006 [online]. Available: <http://ntp.niehs.nih.gov/iccvam/suppdocs/feddocs/oecd/oecd435.pdf> [accessed March 18, 2015].
- OECD (Organisation for Economic Co-operation and Development). 2007a. Guidance on Grouping of Chemicals. OECD Environmental Health and Safety Publication Series on Testing and Assessment No. 80. Paris: OECD [online]. Available: [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono\(2007\)28](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2007)28).
- OECD (Organisation for Economic Co-operation and Development). 2007b. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationships [(Q)SAR] Models. OECD Environment Health and Safety Publication Series on Testing and Assessment No. 69. Paris: OECD [online]. Available: [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono\(2007\)2](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2007)2) [accessed March 17, 2015].
- OECD (Organisation for Economic Co-operation and Development). 2011. Report of the Workshop on Using Mechanistic Information in Forming Chemical Categories, December, 8-10, 2010, Crystal City, VA. OECD Environment Health and Safety Publication Series on Testing and Assessment No.138. Paris: OECD [online]. Available: [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=/env/jm/mono\(2011\)8&doclanguage=en](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=/env/jm/mono(2011)8&doclanguage=en) [accessed March 17, 2015].
- OECD (Organisation for Economic Co-operation and Development). 2013a. OECD Guideline for Testing of Chemicals No. 431: In Vitro Skin Corrosion: Reconstructed Human Epidermis (RHE) Test Method. Adopted July 26, 2013 [online]. Available: <https://ntp.niehs.nih.gov/iccvam/suppdocs/feddocs/oecd/oecd-tg431-2013-508.pdf> [accessed March 18, 2015].
- OECD (Organisation for Economic Co-operation and Development). 2013b. OECD Guideline for Testing of Chemicals No. 437: Bovine Corneal Opacity and Permeability Test Method for Identifying i) Chemicals Inducing Serious Eye Damage and ii) Chemicals Not Requiring Classification for Eye Irritation or Serious Eye. Adopted July 26, 2013 [online]. Available: <http://ntp.niehs.nih.gov/iccvam/suppdocs/feddocs/oecd/oecd-tg437-2013-508.pdf> [accessed March 18, 2015].
- OECD (Organisation for Economic Co-operation and Development). 2013c. OECD Guideline for Testing of Chemicals No. 438: Isolated Chicken Eye Test Method for Identifying i) Chemicals Inducing Serious Eye Damage and ii) Chemicals Not Requiring Classification for Eye Irritation or Serious Eye Damage.

- Adopted July 26, 2013 [online]. Available: <http://ntp.niehs.nih.gov/iccvam/suppdocs/feddocks/occd/occd-tg438-2013-508.pdf> [accessed March 18, 2015].
- OECD (Organisation for Economic Co-operation and Development). 2014. Guidance on Grouping of Chemicals, Second Edition. OECD Environment Health and Safety Publication Series on Testing and Assessment No.194. Paris: OECD [online]. Available: [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono\(2014\)4&doclanguage=en](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2014)4&doclanguage=en) [accessed March 17, 2015].
- Papa, E., M. Luini, and P. Gramatica. 2009. Quantitative structure-activity relationship modelling of oral acute toxicity and cytotoxic activity of fragrance materials in rodents. *SAR QSAR Environ. Res.* 20(7-8):767-779.
- Patlewicz, G., M.W. Chen, and C.A. Bellin. 2011. Non-testing approaches under REACH--help or hindrance? Perspectives from a practitioner within industry. *SAR QSAR Environ. Res.* 22(1-2):67-88.
- Patlewicz, G., C. Kuseva, A. Kesoba, I. Popova, T. Zhechev, T. Pavlov, D.W. Roberts, and O. Mekenyan. 2014. Towards AOP application – implementation of an integrated approach to testing and assessment (IATA) into a pipeline tool for skin sensitization. *Regul. Toxicol. Pharmacol.* 69(3): 529-545.
- Patlewicz, G., T. Simon, R. Budinsky, J.C. Rowlands, and R.A. Becker. 2015. Proposing a scientific confidence framework to help support the application of adverse outcome pathways for regulatory purposes. *Regul. Toxicol. Pharmacol.* 71(3):463-477.
- Perrin, D.D., B. Dempsey, and E.P. Serjeant. 1981. *pKa Prediction for Organic Acids and Bases*. London: Chapman and Hall.
- Peyret, T., and K. Krishnan. 2012. Quantitative property-property relationship for screening-level prediction of intrinsic clearance of volatile organic chemicals in rats and its integration within PBPK models to predict inhalation pharmacokinetics in humans. *J. Toxicol.* 2012:286079.
- Potts, R.O., and R.H. Guy. 1992. Predicting skin permeability. *Pharm. Res.* 9(5):663-669.
- Prieto, P., A. Kinsner-Ovaskainen, S. Stanzel, B. Albella, P. Artursson, N. Campillo, R. Cecchelli, L. Cerrato, L. Díaz, E. Di Consiglio, A. Guerra, L. Gombau, G. Herrera, P. Honegger, C. Landry, J.E. O'Connor, J.A. Páez, G. Quintas, R. Svensson, L. Turco, M.G. Zurich, M.J. Zurbano, and A. Kopp-Schneider. 2013. The value of selected in vitro and in silico methods to predict acute oral toxicity in a regulatory context: Results from the European Project ACuteTox. *Toxicol. In Vitro* 27(4):1357-1376.
- ProTox. 2015. Prediction of Rodent Oral Toxicology. Charite University of Medicine, Berlin, Germany [online]. Available: <http://tox.charite.de/tox/> [accessed March 18, 2015].
- Raunio, H. 2011. In silico toxicology – non-testing methods. *Front. Pharmacol.* 2:33.
- Rorije, E., and E.M. Hulzebos. 2005. Evaluation of (Q)SARs for the Prediction of Skin Irritation/ Corrosion Potential. National Institute of Public Health and Environment (RIVM-SEC)/European Chemicals Bureau (ECB), Bilthoven, The Netherlands [online]. Available: https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/information-sources/qsar-document-area/Evaluation_of_Skin_Irritation_QSARs.pdf [accessed March 17, 2015].
- Rostami-Hodjegan, A., and G.T. Tucker. 2007. Simulation and prediction of in vivo drug metabolism in human populations from in vitro data. *Nat. Rev. Drug Discov.* 6(2):140-148.
- Saliner, A.G., I. Tsakovska, M. Pavan, G. Patlewicz, and A.P. Worth. 2007. Evaluation of SARs for the prediction of skin irritation/corrosion potential: Structural inclusion rules in the BfR decision support system. *SAR QSAR Environ. Res.* 18(3-4):331-342.
- Saliner, A.G., G. Patlewicz, and A.P. Worth. 2008. A review of (Q)SAR models for skin and eye irritation and corrosion. *QSAR Comb. Sci.* 27(1):49-59.
- Sazonovas, A., P. Japertas, and R. Diziapetris. 2010. Estimation of reliability of predictions and model applicability domain evaluation in the analysis of acute toxicity (LD50). *SAR QSAR Environ. Res.* 21(1): 127-148.
- Schultz, T.W., J.W. Yarbrough, and E.L. Johnson. 2005. Structure-activity relationships for reactivity of carbonyl-containing compounds with glutathione. *SAR QSAR Environ. Res.* 16(4):313-322.
- Schultz, T.W., R.E. Carlson, M.T. Cronin, J.L. Hermens, R. Johnson, P.J. O'Brien, D.W. Roberts, A. Siraki, K.B. Wallace, and G.D. Veith. 2006. A conceptual framework for predicting the toxicity of reactive chemicals: Modeling soft electrophilicity. *SAR QSAR Environ. Res.* 17(4):413-428.
- Sedykh, A., H. Zhu, H. Tang, L. Zhang, A. Richard, I. Rusyn, and A. Tropsha. 2011. Use of in vitro HTS derived concentration-response data as biological descriptors improves the accuracy of QSAR models of in vivo toxicity. *Environ. Health Perspect.* 119(3):364-370.
- Silva, F.T., and G.H. Trossini. 2014. The survey of the use of QSAR methods to determine intestinal absorption and oral bioavailability during drug design. *Med. Chem.* 10(5):441-448.
- Simulations Plus. 2010. *Manual GastroPlus™*. Simulations Plus, Inc., CA.

- Suenderhauf, C., F. Hammann, A. Maunz, C. Helma, and J. Huwyler. 2011. Combinatorial QSAR modeling of human intestinal absorption. *Mol. Pharm.* 8(1):213-224.
- ten Berge, W. 2009. A simple dermal absorption model: Derivation and application. *Chemosphere* 75(11):1440-1445.
- Toropov, A.A., B.F. Rasulev, and J. Leszczynski. 2008. QSAR Modeling of acute toxicity by balance of correlations. *Bioorg. Med. Chem.* 16(11):5999-6008.
- Tsakovska, I., I. Lessigiarska, T. Netzeva, and A.P. Worth. 2006. Review of (Q)SARs for Mammalian Toxicity. EUR 22846 EN. European Commission, Directorate General, Joint Research Centre, Ispra, Italy [online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.103.6822&rep=rep1&type=pdf> [accessed March 17, 2015].
- Tsakovska, I., A.G. Saliner, T. Netzeva, M. Pavan, and A.P. Worth. 2007. Evaluation of SARs for the prediction of eye irritation/corrosion potential: Structural inclusion rules in the BfR decision support system. *SAR QSAR Environ. Res.* 18(3-4):221-235.
- UNECE (U.N. Economic Commission for Europe). 2015. Globally Harmonized System of Classification and Labelling of Chemicals (GHS) [online]. Available: http://www.unece.org/trans/danger/publi/ghs/ghs_welcome_e.html [accessed March 18, 2015].
- Veith, G.D., and K.B. Wallace. 2006. Inhalation Toxicity in Fish and Rodents. Presented at the McKim Conference, June 27-29, 2006, Duluth, MN.
- Veith, G.D., E.P. Petkova, and K.B. Wallace. 2009. A baseline inhalation toxicity model for narcosis in mammals. *SAR QSAR Environ. Res.* 20(5-6):567-578.
- Walters, W.P., and M.A. Murcko. 2002. Prediction of 'drug-likeness'. *Adv. Drug Deliv. Rev.* 54(3):255-271.
- Wang, J., and T. Hou. 2015. Advances in computationally modeling human oral bioavailability. *Adv. Drug Deliv. Rev.* 86:11-16.
- Worth, A. 2008. QSAR and Chemical Read-Across within the Context of Non-testing Strategies. Presentation at the 1st SETAC Europe Special Science Symposium: Integrated Testing Strategies for REACH, October, 23-24, 2008, Brussels, Belgium [online]. Available: http://reach.setac.eu/embed/presentations_reach/09_WORTH_QSAR_Chem_Read_Across.pdf. [accessed March 18, 2015]
- Xu, C., and D.E. Mager. 2011. Quantitative structure-pharmacokinetic relationships. *Expert Opin. Drug Metab. Toxicol.* 7(1):63-77.
- Yamazaki, K., and M. Kanaoka. 2004. Computational prediction of the plasma protein-binding percent of diverse pharmaceutical compounds. *J. Pharm. Sci.* 93(6):1480-1494.
- Yang, J., M. Jamei, K.R. Yeo, G.I. Tucker, and A. Rostami-Hodjegan. 2007. Prediction of intestinal first-pass drug metabolism. *Curr. Drug Metab.* 8(7):676-684.
- Zhu, H., T.M. Martin, L. Ye, A. Sedykh, D. Young, and A. Tropsha. 2009a. Quantitative structure-activity relationship modelling of rate acute toxicity by oral exposure. *Chem. Res. Toxicol.* 22(12):1913-1921.
- Zhu, H., L. Ye, A. Richard, A. Gobraikh, F.A. Wright, I. Rusyn, and A. Tropsha. 2009b. A novel 2-step hierarchical quantitative structure-activity relationship modelling workflow for predicting acute toxicity of chemicals in rodents. *Environ. Health Perspect.* 117(8):1257-1264.

4

Assays for Predicting Acute Toxicity

The future path of toxicity testing was foreshadowed early in the 2000s with publication of frameworks or roadmaps that called for an increased emphasis on the use of in vitro assays that evaluate key biological pathways and molecular mechanisms linked to human disease (EPA 2003; NTP 2004). High-throughput testing would allow less expensive, rapid screening of large numbers of chemicals to set testing priorities on the basis of predicted adverse health effects. The National Research Council (NRC) report *Toxicity Testing in the 21st Century: A Vision and a Strategy* (NRC 2007) built on the early publications and gave rise to a variety of large-scale initiatives to see how in vitro testing methods can be used to predict human toxicity. To implement the strategy outlined in the NRC report, a collaboration was formed between the National Toxicology Program (NTP), the US Environmental Protection Agency (EPA) National Center for Computational Toxicology, and the National Institutes of Health National Chemical Genomics Center (NCGC)¹ to identify mechanisms of chemically induced biological activity, to set priorities among chemicals for more extensive toxicological evaluation, and to develop more predictive models of human biological response (MOU 2008²; Austin et al. 2008; Kavlock et al. 2009; Krewski et al. 2009). The collaboration is now formally referred to as the Tox21 program.

The EPA-sponsored ToxCast program (Dix et al. 2007; Kavlock et al. 2012), the Tox21 program, and the European ACuteTox program (Clemedson et al. 2007; Clemedson 2008) are specific examples of large-scale initiatives to evaluate in vitro testing methods for their ability to predict human toxicity. Phase I of the ToxCast program evaluated about 300 conventional pesticide active ingredients in a battery of cell-free and cell-based assays (Judson et al. 2010). In phase II, the chemical space was broadened to include chemicals used in consumer products and industrial processes and unmarketed drugs donated by pharmaceutical companies (Kavlock et al. 2012; Sipes et al. 2013). The completed first phase of the Tox21 screening program tested about 2,800 chemicals—including solvents, fire retardants, dyes, preservatives, plasticizers, therapeutic agents, inorganic and organic pollutants, drinking water-disinfection byproducts, and natural products—in 50 assays (Shukla et al. 2010; Attene-Ramos et al. 2013). The studies laid the groundwork for efforts to characterize the ability of cell-free and cell-based assays and data-modeling approaches to predict activity and potency in selected biochemical targets. Most of the assays developed and validated for high-throughput screening (HTS) applications like the ToxCast program provide information about activation of molecular-receptor families or biochemical activities that are of interest to the pharmaceutical industry. In fact, for most assays, there is not a direct linkage between specific cells or tissue and chemical or mechanistic targets associated with acute lethal or debilitating effects outlined in Table 2-1. Thus, the assays might be of less use for identifying chemicals that potentially can cause acute, debilitating injuries in deployed personnel.

¹NCGC is now part of the National Center for Advancing Translational Sciences.

²High Throughput Screening, Toxicity Pathway Profiling, and Biological Interpretation of Findings, Memorandum of Understanding Between NTP, NCGC and EPA, January 2008. Available: http://www.niehs.nih.gov/about/highlights/assets/docs/memorandum_of_understanding_508.pdf.

This chapter reviews currently available tools and their limitations for immediate implementation (in the next 3-10 years) by the US Department of Defense (DOD) to screen chemicals of interest to DOD for acute toxicity.

IN VITRO ASSAYS

Numerous in vitro screening assays have been developed for measuring specific biological activities of chemicals in specific organs or cell types with an eye to elucidating mechanisms of action. For the purposes of hazard assessment, biological activity in an in vitro system can identify a mechanism of action or response that could be extrapolated to an in vivo end point. This section reviews relevant in vitro biochemical assays that could potentially be used to assess acute toxicity.

Specific-Protein Assays

Testing whether a chemical inhibits a particular enzyme or binds to a particular receptor or other biomolecule is the most direct way to test a chemical for a specific mechanism of action at the molecular level. Enzyme and receptor-binding assays tend to be reliable, to exhibit relatively good agreement between different laboratories, and to be well suited for high-throughput formats. The specific protein, protein complex, receptor, or other biomolecule of interest can be provided in the assay as a pure molecule, as a partially purified complex, or as a component of living cells.³

Given their high value for predicting specific molecular effects, protein assays are the most frequent type of assay in the ToxCast program (Figure 4-1). Some of the ToxCast assays have direct relevance to predicting acute toxicity of chemicals that could be used as warfare agents. For example, assays of acetylcholinesterase activity are applicable to cholinesterase-inhibiting nerve agents, and mitochondrial electron-transport assays are applicable to cyanide. Specific-protein assays in the ToxCast and ACuteTox studies that were designed to detect nervous system effects, such as modulation of ion-channel activity, are also relevant for predicting acute toxicity. However, it is important to recognize that, like many of the other assays in ToxCast, most of the specific-protein assays included in ToxCast were designed for needs other than predicting acute toxicity—for example, to predict endocrine-disrupting activity—and might have little value for predicting acute toxicity.

Limitations and Needs for Improvement of Protein Assays

A major limitation is the biological space that is covered by specific-protein assays that are designed to assess the actions of a chemical on specific enzymes or receptors. More complicated or nonspecific mechanisms might also be involved in acute toxicity. For example, vesicant action on the skin is not mediated by the action of a chemical on a specific enzyme or receptor. Toxic chemicals that act through nonspecific mechanisms might not be identified in specific-protein assays or might be identified in multiple assays but with poor correlation between dose and response, which potentially confuses analyses. Another limitation is that assays for particular acute toxic mechanisms (such as activity on particular ion channels) might be missing from the suite of assays or might be unreliable. A final limitation is that some specific-protein assays involve general mechanisms that are common to many cell types and thus might not predict particular organ toxicities themselves.

³Assays that use living cells to evaluate the effects of chemicals on specific proteins could have been considered in the section that follows, but for simplicity, this section reviews all assays whose purpose is to measure effects of a chemical on the biochemical activity of a specific protein or other biomolecule.

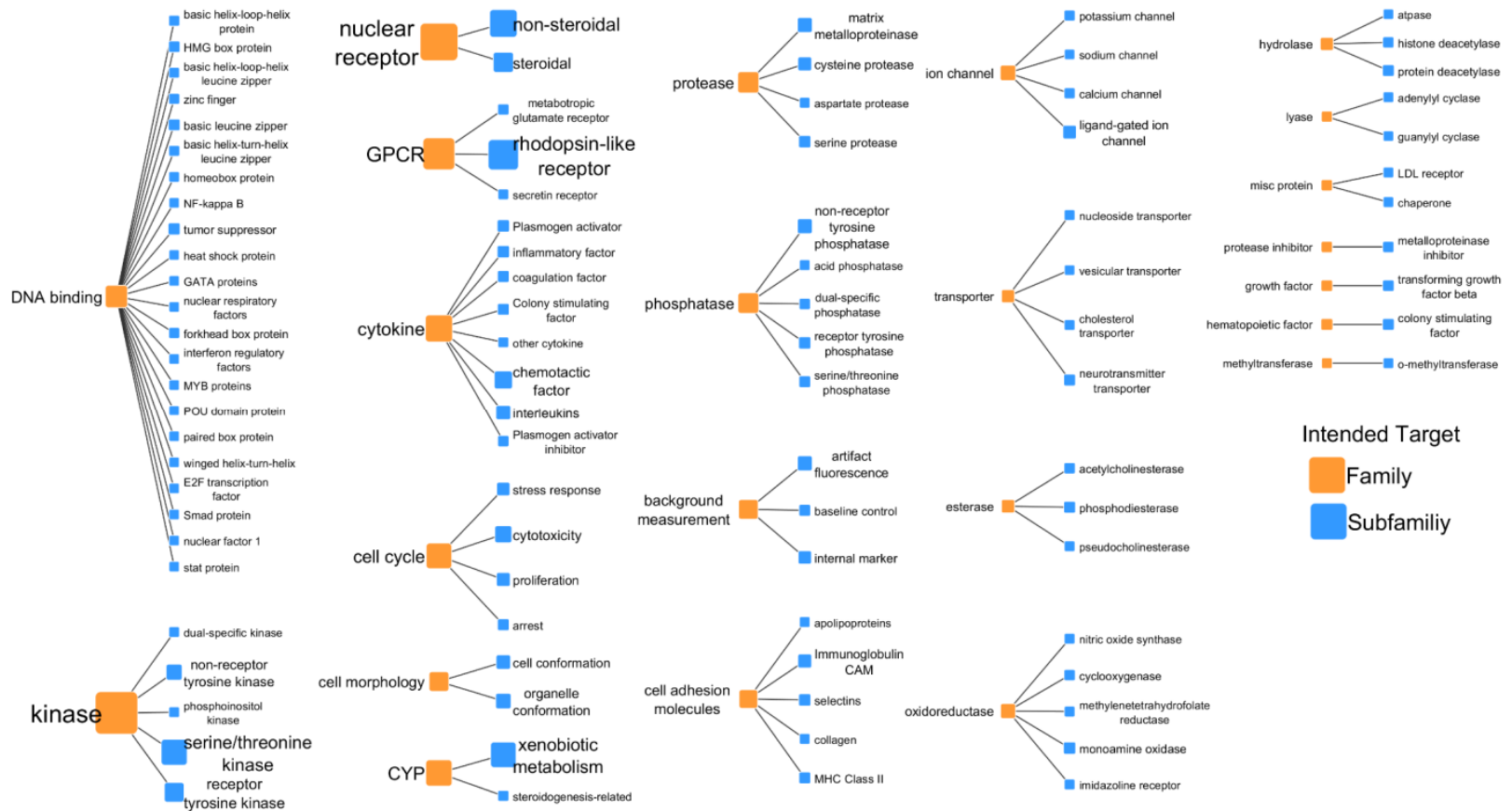


FIGURE 4-1 Intended target families and subfamilies for the ToxCast program. The number of assays for each intended target is represented by the sizes and font sizes of the orange nodes. For each intended target family, the assays are subdivided into assay end points (see EPA 2014a, p.13).

To make specific-protein assays more useful for DOD's needs than they are now, there is a need to identify the subset of existing specific-protein assays that are directly mechanistically and quantitatively relevant to debilitating injuries, to fill gaps in assay types (for example, to develop assays for perturbation of particular ion channels that are not included in current platforms), to develop methods for identifying and classifying chemicals that have potent but non-specific toxic actions, and to evaluate performance and predictive value of specific-protein assays for predicting acute toxicity by using a panel of reference and test chemicals relevant to DOD's interests.

Cell-Based Phenotypic Assays

This section describes high-throughput assays that use cultured cells and measure some overall phenotypic output that is relevant to predicting acute toxicity, such as cell proliferation, plasma membrane permeability, or adenosine triphosphate (ATP) content. There is a large scientific literature on the application of cell-based assays to drug development and a growing literature on their application to toxicology. The ToxCast program includes more than 100 cell-based assays whose purpose is to measure cytotoxicity and other aspects of cellular phenotype. Simple cytotoxicity assays have long been used, with partial success, to predict animal and human toxicity and to estimate starting concentrations for animal toxicity studies.

A key consideration in all *in vitro* cell-based phenotypic assays is the choice of cell type. The general approach is to attempt to mimic specific human cell types; however, few data suggest, for example, that hepatocytes best predict liver injury or that renal tubule cells best detect renal injury (Lin and Will 2012). Assays are conducted with cells that are grown in a single layer (cell monolayer culture), and these conditions inevitably provide only a crude approximation of *in vivo* tissue environments, cell types, and cell-cell interactions. Furthermore, immortalized (often cancer-derived) and other cell lines have provided the mainstay of cell-based assays for decades because they are convenient for obtaining large numbers of cells in a standard state and for enabling cell lines to be readily shared between laboratories. However, the cells are often cultured using different media conditions, and this can affect cell response to chemicals. For example, high glucose concentrations in the growth media might increase the resistance of neural cells to the mitochondrial toxicant 1-methyl-4-phenylpyridinium (Mazzio et al. 2010). Similarly, cytotoxicity results from assays that evaluate mitochondrial disruption can be influenced by the Crabtree effect, in which cancer cells preferentially use glycolysis instead of oxidative phosphorylation for ATP production (Marroquin et al. 2007). Another limitation related to cell type is that many *in vitro* studies use cell lines that do not reflect the variation that is found in the human population. For some chemicals, sensitivity to cytotoxicity varied by as much as a factor of 100 in a single cell type taken from a broad population sample (Abdo et al. 2015).

There is an increasing shift away from cell lines toward cell-based assays that use cell types that are more physiologically relevant, such as animal or human primary cells, and human induced pluripotent stem (iPS) cells that have differentiated into specific cell types (Kraushaar et al. 2012; Godoy et al. 2013). However, unless great care is taken to mimic tissue-relevant physical and chemical environments, those cell types might provide little improvement over cell lines in physiological relevance. That consideration has led to an increasing push toward the use of organotypic model systems (discussed below).

Another key technical consideration is the measurement of assay results (readout). Readouts can be broadly divided into ones that average the response of a number of cells in a tissue culture well and ones that assess individual cell behavior, the latter sometimes called high-content assays (discussed below in the section "Emerging Technologies"). Whole-well readouts are most commonly used and have the advantage of being fast and simple, and it is straightforward to generate statistical metrics of assay performance and chemical activity. High-content readouts have the advantage, in principle, of providing much more information—for example, on

cell-to-cell heterogeneity in response and on specific cytotoxic mechanisms—but scoring and interpreting the additional information is computationally challenging (Wink et al. 2014).

Few studies have evaluated a large number (hundreds) of chemicals for their ability to predict acute toxicity (O'Brien et al. 2006; Xu et al. 2008; Lin and Will 2012; Porceddu et al. 2012). Most studies examine only a handful of chemicals, and studies that evaluate only specific classes of chemicals, such as endocrine disruptors or hepatotoxicants, substantially bias estimates of model performance (Thomas et al. 2012).

Cytotoxicity Assays

Measurements of cell life or death in culture have a long history in toxicology research (Ekwall 1983). Methods for measuring cytotoxicity in cell culture usually involve direct measurement of the fraction of cells that have intact membranes, for example, with neutral red uptake or fluorescent DNA dye uptake; measurement of the metabolism of surviving cells, for example, with reduction of 3-(4,5-dimethyl-thiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT), reduction of Alomar blue, or uridine uptake; or measurement of ATP content, cell number, total DNA content, total protein content, or cell proliferation. Cytotoxicity assays are normally run for a day or more (3 days is common), so viable cells will proliferate and increase the live-cell signal in control wells. Thus, the assays usually combine measurements of acute cell lethality, cell proliferation, or cell metabolism and might represent the simultaneous occurrence of several mechanisms (Huang et al. 2008). The longer an assay is run, the more it tends to factor in effects on cell proliferation unless it is run on nonproliferating cell types. Short-term (1-hour) assays are also used but need to include sensitive measures of cell injury, such as mitochondrial membrane potential or ATP content. Some assays are based on ultraviolet (UV) or fluorescence readouts and warrant caution because some chemicals can artificially interfere with UV- or fluorescence-based assays. In the pharmaceutical industry, luminescent readouts have largely replaced UV- and fluorescence-based assays.

In vitro cytotoxicity tests have been recommended as an adjunct to animal tests to improve initial dose selection and modestly reduce the number of animals used. The registry of cytotoxicity (RC) prediction model has been recommended (NIEHS 2001a,b) for evaluating the predictive accuracy of candidate cytotoxicity assays.⁴ Many RC prediction models evaluate the correlation between in vivo LD₅₀s and in vitro cytotoxicity IC₅₀s.⁵ The BALB/c 3T3 and normal human keratinocyte (NHK) neutral red uptake (NRU) cytotoxicity assays have been used to predict acute toxicity (NIEHS 2001a,b), but a major drawback of those cell-based assays is that they are difficult to automate. Recent advances have led to the development of commercially available “ready-to-go” cell plates that simply need to be defrosted and fed with media. The plates can then be used in the conduct of high-throughput assays.

Recent systematic studies that evaluated the predictiveness of in vitro assays for acute toxicity included several cell-based cytotoxicity assays partly because of their ease of use, reproducibility, and proven (albeit limited) predictive value. For example, of the 53 assays in the European Union (EU) ACuteTox project, seven constitute either a cell lethality assay (for example, NRU) or a metabolism assay (for example, MTT reduction or 2-deoxyglucose and uridine uptake) that effectively measures the number of living cells remaining after treatment. Notably, several cytotoxicity assays—an MTT assay in primary rat hepatocytes; a cytotoxicity panel that measures intracellular Ca²⁺ levels, mitochondrial membrane potential, and plasma membrane potential in HepG2, SH-SY5Y and A.704 cells; and a basal cytotoxicity NRU assay in BALB/3T3 cells—generated data of sufficient quality to be considered in future acute-toxicity testing strategies (Kinsner-Ovaskainen et al. 2013).

⁴Halle (Spielmann et al. 1999; Halle 2003) describes inclusion criteria for data in the RC database.

⁵IC₅₀ is the chemical concentration at which 50% inhibition is achieved.

Gene-Expression and Protein-Secretion Assays

Gene-expression and protein-secretion assays aim to identify a specific toxic mechanism by measuring expression of a specific gene or secretion of a specific protein that has been implicated as a biomarker of a particular toxic mechanism in humans. Gene-expression readouts can use engineered reporter genes, whereby an artificial promoter drives expression of an easily assayed reporter gene (such as luciferase) in response to activation of a toxic pathway or endogenous gene circuits. Engineered reporter genes provide easily standardized, reliable readouts, and a panel of such assays can be used to cover multiple biological pathways.

In recent years, several vendors have offered reporter assays for a variety of biological pathways involved in toxicity, such as inflammation, apoptosis, and endoplasmic reticulum stress. Their main limitation is that an artificial promoter in a generic cell line (such as HEK cells) might not indicate pathway activity in the same way as the natural pathway in a specific human cell type; that is, the assays could fail to capture biological context dependence. Reporter gene assays are of particular value for predicting endocrine disruptors because they work by reporting the expression of genes that are naturally regulated by sex hormones. However, endocrine disruption does not qualify as a likely mechanism of systemic toxicity that would result in an acute, debilitating injury in deployed personnel.

Expression of many endogenous genes can be simultaneously analyzed at the mRNA or protein level to create a “toxicogenomic signature” (discussed below under “Emerging Technologies”). Assays for induction of endogenous genes can be applied to primary cells that synthesize specific proteins in response to chemicals. A major advantage of this approach over artificial promoter constructs is that it measures endogenous gene-expression pathways and is thus more likely to indicate physiologically relevant pathway perturbation. A major disadvantage is that separate readouts must be developed for each protein, and this makes the approach more complex and labor-intensive than measuring the expression of a reporter like luciferase from engineered gene-expression constructs.

Protein-secretion assays are of particular value for measuring the activity of immunotoxins that induce expression and secretion of specific inflammatory cytokines, such as TNF- α , IL1, and IL6 from white blood cells. Cytokine measurements are usually made with an immune assay, typically a capture ELISA assay that uses two antibodies to provide high specificity and sensitivity. The relevance of the cytokine assays for evaluating acute toxicity associated with chemicals of relevance to DOD is limited.

Several recent systematic studies have combined multiple high-throughput assays, including assays for the expression of single endogenous genes in primary cells with mRNA or protein readouts. For example, the ACuteTox project included four cytokine secretion assays performed on primary white blood cells isolated from human blood and four assays that measured synthesis of neuronal and glia proteins, at the mRNA level, from aggregates of primary cells derived from rat brain (Kinsner-Ovaskainen et al. 2013). A meta-analysis found a variety of problems with all the assays, including failure to measure in the correct range for immunoassays and lack of reproducibility of primary cell aggregates. The ToxCast program evaluated the performance of the Biologically Multiplexed Activity Profiling (BioMAP) human primary cell disease models with the ToxCast phase I library (Houck et al. 2009). The chemicals that were tested generated relatively weak signatures compared with the reference pharmacological probes and drugs, and this raised the question of whether the concentrations that were used (up to 40 μ M) were adequate. Furthermore, a follow-up study demonstrated that some chemicals with known biological activity, such as pharmaceuticals, gave false-negative results in the assays (Kleinstreuer et al. 2014). All the problems highlighted here indicate the difficulty of working with complex primary cells in a high-throughput assay format.

Limitations and Needs for Improvement of Cell-Based Phenotypic Assays

Cell-based assays are efficient in assessing chemical mechanisms of action. They have also been useful for predicting some mechanisms of chronic toxicity, such as endocrine disruption, but might have less applicability to acute-toxicity prediction. Furthermore, conventional cell-based cytotoxicity assays typically lack metabolic competence and often miss chemicals that require bioactivation. They also fail to provide data on some of the most important toxic mechanisms, notably ones that involve organ- or cell-type specific physiology. For example, they miss chemicals that perturb synaptic transmission and neurotransmitter metabolism.

Responses seen in cytotoxicity assays can also be influenced by the choice of cell used. In the Tox21 cytotoxicity profiling screen, cytotoxic response patterns varied by cell type and indicated differences in sensitivity and kinetics of the response (Xia et al. 2008). Overall, the human blood-derived cells (Jurkat), neuron-derived cells (SH-SY5Y), and rodent cells (N2a, H-4-II-E and NIH 3T3) were most sensitive to chemical-induced cytotoxicity, and human fibroblastic, endothelial, and skin cells (HUV-EC-C, BJ, and MRC-5) were least sensitive. Another finding from the study was the lack of similarity in the patterns of chemical activity in cells that were derived from the same tissue but from different species, for example, human HepG2 and rat H-4-II-E hepatoma cells.

To improve the utility of cell-based assays for predicting acute toxicity, several aspects of experimental design need to be considered. First, cell-based assays that express the relevant biological processes and toxicologic mechanisms under study need to be selected. In some cases, that might require using primary cell cultures rather than cultured immortalized cells. Second, relevant chemical concentrations and exposure durations need to be used, and appropriate test and reference chemicals need to be identified for evaluating the performance and predictive value of the assays. Ultimately, the readout of a cell-based assay must be mechanistically and quantitatively linked to a chemical's toxicity phenotype in humans if it is to be truly predictive.

Organotypic Models

Many of the limitations of cell-based assays, as noted above, arise from the fact that single cells growing on a dish are inevitably poor models for human tissues. That is the case even if the cells themselves were derived from human or animal tissue because the artificial *in vitro* environment changes their phenotype in hours. The culture-induced changes affect cells' responses to chemicals and the cells' (especially liver cells') ability to biotransform chemicals into more or less toxic species. To address the limitation, researchers have been developing approaches for co-culture of multiple cell types or for cultures of whole organs as slices or cell aggregates. Those approaches are collectively referred to as organotypic models.⁶ The discussion below highlights a few organ systems that are relevant to acute toxicity of chemicals.

Skin Organotypic Models

Skin, as the largest human organ, is important for the absorption of many classical chemical-warfare agents and constitutes the principal barrier to and defense against absorption of toxic lipophilic chemicals. In addition, the skin is the primary site of action of acute blistering agents. The military's historical reliance on chemical-protective boots, suits, and gloves emphasizes the importance placed on dermal protection against chemical exposure. Furthermore, one of the first

⁶EPA defines organotypic culture models as "tissue culture models that mimic *in vivo* tissue architecture through interactions of heterotypic cell types (e.g., epithelium-stroma) and extracellular matrices (ECM). They can be established from isolated cells or from tissue fragments harvested *in vivo*, and will bridge the gap between conventional monolayer cell cultures and whole-animal systems" (EPA 2013).

questions that the “warfighter” asks when told that a potent, toxic chemical can be absorbed through intact skin is, How much and how quickly? Although definitive dermal absorption studies can be conducted on animals by using small numbers of chemicals that are synthesized with radiolabels, various *in vitro* models have been developed by using instrumented flow cells with human or porcine skin explants. Those *in vitro* models are far more amenable to high-throughput screening than are dermal absorption studies with radiolabeled chemicals (Basketter et al. 2012).

Organotypic cultures of skin in formats that are applicable to screening thousands of chemicals are relatively advanced. The field has benefited from the relatively simple anatomical organization of skin, from the fact that proliferating human keratinocytes are relatively easy to grow, from the abundant availability of human tissue from minor surgical procedures, from the need for artificial skin for treating burn patients, and, not least, from huge investment by the cosmetics industry. Especially in the European Union, cosmetics manufacturers have been under pressure to increase safety testing while reducing animal use. In an example that is generally encouraging for toxicity prediction, scientists at L’Oreal, Inc. in France have used the EpiSkinSM model to predict irritant activity (Cotovio et al. 2005, 2008). The model consists of primary human keratinocytes that have been induced to self-organize into a multilayered structure similar to skin by use of bioengineered substrates. MTT reduction and release of the inflammatory cytokine IL-1 α were measured. The model accurately (> 80%) predicted irritant activity of 184 cosmetic ingredients (Cotovio et al. 2008). The study suggests that good—although not perfect—predictions can be made for acute skin irritants by using a sophisticated organotypic culture model. In contrast, numerous *in vitro* and *in vivo* models have been examined in attempts to emulate the blistering (vesication) seen in human skin on exposure to sulfur mustard or lewisite, most with little or no success. Chemical-induced skin blistering might be limited to humans, or it requires epithelial and fibroblast immune cell functions that are not accounted for in current organotypic skin models.

Eye Organotypic Models

Loss of vision would be an incapacitating effect of concern to DOD. Testing for eye irritation has benefited from investment by the cosmetics industry, although current organotypic cornea models are generally less advanced than skin models. Commercial models are available and include the EpiOcular™ OCL-200 tissues from MatTek Corporation (Ashland, MA) and the SkinEthic™ Reconstituted Human Corneal Epithelium developed by a consortium of European cosmetics companies in response to banning of rabbit testing. Systematic evaluation of the systems continues but mostly in the chemical space relevant to cosmetics. Both systems achieved benchmarks for between-laboratory reproducibility and are undergoing tests of predictive value (Alépée et al. 2013; Pfannenbecker et al. 2013). On the basis of the results with skin, a reasonable degree of predictive power is expected.

Lung Epithelium Organotypic Models

The lung is a relevant organ for both chemical absorption and acute toxicity. Multiple 3-D organotypic models have been described, including the commercial EpiAirway™ (MatTek Corporation, Ashland, MA) and MucilAir™ (Epithelix Sàrl, Geneva, Switzerland) systems. The systems use primary cells cultured at an air–liquid interface, and they model airway function better than traditional submerged tissue culture. A recent paper evaluated the models relative to each other and to two conventional submerged tissue–culture systems for their ability to predict rodent lung toxicity of 19 chemicals (Sauer et al. 2015). None of the systems performed well in

predicting lung toxicity in rats, and the 3-D organotypic systems performed no better than conventional tissue-culture models. The findings suggest that organotypic lung epithelium models lag behind skin models in predictive value. The poor predictivity probably reflects the greater complexity of airway epithelium or perhaps differences in investment.

There does not appear to be a particularly good *in vitro* model for lung damage associated with chemicals (for example, phosgene) that are known to increase pulmonary permeability that results in noncardiogenic pulmonary edema. The development of *in vitro* assays to evaluate similar compounds identified as chemicals of interest, for example, on the basis of chemical structure or quantitative structure–activity relationships will require further DOD investment to replace animal inhalation-toxicity studies.

Liver Organotypic Models

The liver is especially relevant to biotransformation of chemicals into more or less toxic metabolites and is an important site of acute toxic action of some chemicals. Hepatocytes are the major biochemical engine of the liver and are responsible for metabolism and excretion of many xenobiotics. Hepatocytes can be cultured in standard 2-D formats, but their phenotype with respect to xenobiotic metabolism and responses changes rapidly under culture conditions. Immortalized cell lines derived from hepatocellular carcinoma (such as HepG2) have often been used as a surrogate for hepatocytes, but their biology is even more distant from hepatocytes *in situ*. Because of the relevance of liver toxicity to drug development, the pharmaceutical industry has made major investments in modeling liver biology in 2-D cell cultures, 2-D co-cultures, 3-D cell cultures, and engineered organotypic systems (Godoy et al. 2013).

Multiple approaches have been taken to build organotypic models of human and rodent liver, and substantial improvements over 2-D hepatocyte culture systems in recapitulating normal liver biology, drug metabolism, and drug responses have been noted (Khetani and Bhatia 2008; Godoy et al. 2013; Messner et al. 2013). Despite improvements, predictive-toxicology studies still tend to focus on 2-D cultures of primary hepatocytes of hepatocellular carcinoma-derived cell lines. For example, the EU-funded LIINTOP project is evaluating multiple 2-D culture models for liver and intestine (Gómez-Lechón et al. 2010), and the EU ACuteTox project included five assays with HepG2 cells (Kinsner-Ovaskainen et al. 2013).

The continued reliance on 2-D cultures and hepatoma-derived cell lines is problematic because testing in more robust *in vitro* models would probably generate more-predictive data. For example, Khetani and Bhatia (2008) showed that a microengineered 2-D culture system in which hepatocytes are co-cultured with stromal cells provided better modeling of gene expression, metabolism, and drug action than conventional 2-D cultures. Another promising system involves co-culture of hepatocytes with Kupffer cells in small spheroids (Messner et al. 2013). The microengineered systems often have lower throughput and are more expensive than conventional 2-D cultures, but given the importance of the liver in toxicology the expense might be worthwhile. Although it remains to be seen how their overall predictivity differs from that of simpler models, the use of complex liver models should improve the recognition of inflammation and other mechanisms of toxicity that are not easily detected in simpler hepatocyte cultures. For example, Khetani et al. (2013) demonstrated that using co-cultured hepatocytes better predicted drug-induced liver injury in a small test set of 45 chemicals.

Godoy et al. (2013) provides an exceptionally comprehensive overview of hepatic models. It is interesting to note, given that the liver is perhaps the most thoroughly characterized of the organotypic model systems, that the authors concluded that “one key message is that despite our enthusiasm for *in vitro* systems, we must never lose sight of the *in vivo* situation. Although hepatocytes have been isolated for decades, the hunt for relevant alternative systems has only just begun.”

Neural Organotypic Models

The brain has multiple neuronal and glial subtypes, complicated neuronal networks that have different types of chemical synapses, important cell–cell interactions, and myelinated axons. That cellular complexity helps to make it an important site of action for many acute toxicants and makes it difficult to identify neurotoxic effects with conventional cytotoxicity assays or other *in vitro* systems. Brain aggregate cultures replicate some organotypic structural and functional features and have been used as a model system for neurotoxicity testing. For example, the EU ACuteTox project included seven assays of cell aggregates derived from rat brain as part of a battery of 50 assays that included many involving 2-D cultures of neurons (Forsby et al. 2009). The authors concluded that “using aggregate cell cultures prepared from embryonic rat brain and a multiparametric endpoint scheme, all chemicals known to be highly toxic in humans also showed high toxicity (significant effects in the lower micromolar range) to extreme toxicity (significant effects at nanomolar concentrations) in aggregate cultures” (Honneger et al. 2009). They also noted that inclusion of data on metabolism and pharmacokinetics of the blood–brain barrier would likely improve predictive value.

Limitations and Needs for Improvement of Organotypic Models

Additional organotypic models have been developed for the heart, kidney, and skeletal muscle. Organotypic models have high potential for predicting acute toxicity and potentially can recapitulate the metabolism and biological activity of a chemical. That said, the science of accurately modeling human organs in a culture dish, especially in formats suitable for high-throughput testing, is still in its infancy. Progress is being made, but much caution is warranted, particularly for acute-toxicity prediction, which has not been the goal of most studies. Organotypic assays are much more complex and expensive than pure protein-based and cell-based assays and less robust than rodent models because they do not integrate multiple physiological systems. Their reliability has often been called into question in systematic studies. Their predictivity is also far from guaranteed; for example, the lack of success in predicting vesicant activity by using organotypic skin cultures is troubling.

For organotypic cultures to be used in DOD screening for acutely toxic chemicals, there is a need to evaluate the potential of organotypic assays for acute-toxicity prediction and to invest further in the basic science of organotypic cultures. The potential usefulness is high, even if current systems are far from ideal.

NONMAMMALIAN IN VIVO ANIMAL MODELS

The use of an *in vivo* approach facilitates the crucial understanding of how chemicals affect complex metabolic targets and pathology at the cell and organ level. Mice, rats, rabbits, and other laboratory mammals have been used extensively to study chemical toxicity. However, *in vivo* assays with those species are often expensive, use large amounts of toxic test chemicals, and are difficult to use in a moderate-throughput to high-throughput manner. Those drawbacks have led toxicologists to develop alternative animal models for chemical testing. Many alternative test organisms share biological processes with rodents and other mammals, including humans. Three test platforms of note that can be adapted to high-throughput screening rely on insect, nematode, and zebrafish models (Giacomotto and Ségalat 2010). They can complement other cell-based *in vitro* test systems, and data from assays that use the alternative animals could be used to set priorities among chemical candidates for future traditional animal testing. Development and application of the nonmammalian models could help DOD to screen for pathway-specific effects and could demonstrate how chemical toxicity varies among species.

This review of alternative animal models is not intended to be exhaustive; rather, the committee has focused on end points that are relevant to acute toxicity and on selected examples that illustrate how these systems could be adapted for high-throughput testing of acute effects.

Fruit Fly Models

The fruit fly (*Drosophila melanogaster*) has been used as a model organism in studies of genetics and developmental biology for over 100 years (Rubin and Lewis 2000). Fruit flies have a fully mapped genome, and many protocols for biochemical and genetic analysis are well established. More than 60% of human genes have functional orthologs in *D. melanogaster* (Bier 2005). A variety of molecular tools, including mutagenesis and RNAi, are available for modifying fruit fly genetics. A dedicated Web-based database (FlyBase) contains information relative to fruit fly genetics and its molecular biology (Drysdale 2008).

Fruit flies have the potential to be used for chemical-toxicity screens (Nichols 2006; Whitworth et al. 2006; Segalat 2007). In particular, fruit flies and other insect models have improved our understanding of the molecular action of pyrethroids, which act on both mammalian and insect sodium channels, and other insecticides (Peterson et al. 2008). Despite their use in neurotoxicology research, there are important limitations (Rand 2010). For example, γ -aminobutyric acid, acetylcholine, and other neurotransmitters often have roles in the insect nervous system that are different from their roles in vertebrate nervous systems (Peterson et al. 2008).

Specialized video-based equipment has been developed to assess flying, chemotaxis, geotactic climbing,⁷ and other behaviors (Sawin-McCormack et al. 1995; Rand 2010; Sokolowski 2001; Podratz et al. 2011; Gregory et al. 2012; Podratz et al. 2013). In addition, eclosion⁸ and adult lethality are simple end points that can be assessed without the need of a microscope. However, the transition from larva to adult fly is complex and occurs by mechanisms distinct from those seen in mammals; this draws into question the utility of LD₅₀s obtained for this developmental stage (Rand 2010). There are other limitations of the use of fruit flies. For example, chemical administration to the fly embryo must overcome the barriers presented by the hydrophobic vitelline membrane (Limbourg and Zalokar 1973; Rand 2010). In addition, fly toxicokinetics of xenobiotics can differ substantially from that in mammals, including humans.

Nematode Models

The most widely used nematode model for biomedical research is likely *Caenorhabditis elegans*. *C. elegans* have a fully mapped genome (*C. elegans* Sequencing Consortium 1998), and more than 50% of human genes have functional orthologs in *C. elegans* (Harris et al. 2004). Genetic or genomic manipulation—such as knockouts, knockdown via RNAi, and transgenic strains—is routinely available. A variety of bioinformatics tools have been developed to support high-throughput genomic studies with this organism (Cho et al. 2014). In addition, a dedicated Web site (Wormbase) allows investigators access to microarray data and comprehensive data on gene structures, mutants, RNAi phenotypes, and protein–protein interactions (Chen et al. 2005).

Numerous studies have shown that *C. elegans* and humans share many essential biological characteristics. *C. elegans* has a rudimentary nervous system, exhibits behavior, and is capable of rudimentary learning and memory functions. The anatomy of *C. elegans* is well understood. It contains 959 somatic cells, including about 300 neurons that are microscopically visible. The developmental cell lineage and the neural wiring diagram of *C. elegans* have been completely mapped. Many vertebrate neurotransmitters are well conserved in this nematode (Villatte et al.

⁷*Drosophila* instinctively climb against gravity (geotaxis).

⁸Eclosion is the hatching of adults from the pupal stage.

1998; McVey et al. 2012). The presence of a functional nervous system has been exploited by neurotoxicologists to study the acute neurotoxic effects of pesticides and other chemicals on this nematode (Williams and Dusenbery 1990; Ruan et al. 2009; Avila et al. 2012; McVey et al. 2012; Meyer and Williams 2014). End points evaluated include changes in survival, behavior, locomotion, life span, cell death, neurotransmitter concentration, and function. Image-tracking systems have also been developed for assessing *C. elegans* locomotion (Feng et al. 2004). Melstrom and Williams (2007) reported a strong correlation between LC₅₀s determined for *C. elegans* and LD₅₀s identified in rodents after exposure to cholinesterase-inhibiting pesticides. End points examined by Melstrom and Williams included depression of acetylcholinesterase activity and decreased movement.

Zebrafish Models

As a vertebrate, zebrafish (*Danio rerio*) have substantial physiological, anatomic, and genetic homology with humans (Barbazuk et al. 2000; Howe et al. 2013; Chakravarthy et al. 2014). Zebrafish are amenable to gene manipulation, have a short generation time, and have well-characterized rapid developmental stages; these characteristics have led to their growing use in developmental-toxicity studies (de Esch et al. 2012; Raldúa and Piña 2014). Its application to the study of developmental toxicity has garnered the interest of regulatory bodies. For example, the Organisation for Economic Co-operation and Development (OECD) has recently formulated guidelines for using zebrafish embryos for testing acute toxicity of 119 chemicals and for developmental toxicology (OECD 2013a,b). Zebrafish can be bred in large numbers with minimal maintenance cost (Raldúa and Piña 2014). They are a cost-effective *in vivo* model for screening drugs and other chemicals and meet many objectives of a high-throughput screening assay (Taylor et al. 2010; Tsang 2010; Lessman 2011). High-throughput zebrafish assays have also been used for microarray and proteomic studies (Love et al. 2004). Because zebrafish larvae are transparent, they are ideal for *in vivo* imaging without the use of invasive techniques (Knudsen et al. 2011; Raldúa and Piña 2014). Another advantage of the zebrafish larva model is that up to 4 days after fertilization they are not treated as vertebrates by US Institutional Animal Care and Use Committees because they retain a yolk and higher-order neuronal functions are generally absent.

Zebrafish are used in safety pharmacology studies to screen for arrhythmogenicity (Langheinrich et al. 2003; Milan et al. 2003; Burns et al. 2005), nephrotoxicity (Hentschel et al. 2005; Wu et al. 2012), hepatotoxicity (Vliegenthart et al. 2014), and neurotoxicity (de Esch et al. 2012; Legradi et al. *in press*). Some drugs that affect human cardiac function and structure are known to have similar effects in zebrafish (Milan et al. 2003). Heart-specific expression of the green fluorescent protein in zebrafish has been accomplished by using the cardiac myosin light chain 2 promoter (Huang et al. 2003). Specialized equipment exists to monitor changes in zebrafish heart rate after chemical exposure (Burns et al. 2005; Simoneschi et al. 2014). Behavioral assays of swimming behaviors have been developed for zebrafish (Ali et al. 2012; Bichara et al. 2014). Driessen and co-workers (2013) have shown good concordance in histopathological responses and gene expression profiles between zebrafish embryos and mice exposed to known hepatotoxic chemicals. Yen et al. (2011) evaluated changes in zebrafish larva survival, acetylcholinesterase activity, and behavior after exposure to three organophosphorus pesticides. That type of study might be useful for nerve agents and other chemicals that have a similar mechanism of action. *In vivo* zebrafish assays with reverse dosimetry have also been used to develop human oral-dose hazard values (Perkins et al. 2013).

Limitations and Needs for Improvement of Nonmammalian *In Vivo* Animal Models

Nonrodent animal models have the potential to assist in characterizing the acute toxicity of chemicals. One advantage of such systems is their ability to identify whole-animal and organ-

level responses. The exploration of nontraditional *in vivo* models for assessing acute-toxicity potential, however, will need to consider species differences in metabolism and cellular targets and other issues related to interspecies and *in vitro*-to-*in vivo* extrapolations. Other factors to consider include differences in organ composition (multiple cell types), cell organization or structure, and gradual enzyme expression in different tissue regions. Another challenge is related to the extrapolation of aqueous (as in the case of zebrafish) or medium (as in the case of *C. elegans*) concentrations to exposure concentrations that are relevant to humans. Alternative animal model *in vivo* assays have considerably lower throughput than other assay systems considered by the committee and are likely to be used in the later stages of the assessment process. Little work has been performed regarding the applicability of such models to assess the acute toxicity of chemicals relevant to DOD, and they have been incompletely analyzed for their predictive validity with respect to acute toxic effects or identification of affected organ systems.

EMERGING TECHNOLOGIES

The drive to develop nonanimal methods for toxicity testing is still in its infancy, and new technologies are continually being developed. Most are aimed at commercial applications, particularly for safety assessment in the pharmaceutical and cosmetics industries, but a subset of the new technologies will also be useful for predicting acute toxicity. In this section, emerging technologies are broadly divided into ones that generate large amounts of information per sample by multiplexed or image-based measurements and ones, such as organ-on-a-chip and induced pluripotent stem (iPS) cell technologies, that aim to model human tissues more accurately.

Multiplexed Assays

Multiplexed assays allow the measurement of dozens, hundreds, or even thousands of end points simultaneously on a single sample. In general, they seek to provide more biological data per sample than traditional assays that measure a single end point. Conceptually, obtaining data on many end points is expected to help in deciphering chemical mechanisms and identifying new biological targets for development of more specific assays (Larson et al. 2011). The most used, and best understood, multiplexed readout is a gene-expression profile, in which the amounts of hundreds or thousands of mRNAs in a sample are measured in parallel (Fabian et al. 2011; Klapner et al. 2014). Microarrays have been popular for measuring gene expression, but the decreasing cost of DNA sequencing is leading to a gradual replacement of microarray and related technologies with RNAseq approaches. There is increasing interest in multiplexed measurement of micro-RNAs, whose expression also reflects the state of a cell or tissue.

Protein and metabolite measurements can also be multiplexed. For example, multiplexed immunoassays can be used to measure the concentration of tens or hundreds of cytokines or other proteins in a single sample. The main limitation of such assays is in developing high-quality antibody pairs to capture and quantify a given protein with high sensitivity and specificity. Given recent developments in proteomics technology, it is possible that mass-spectrometry-based measurements will gradually replace immunoassays for multiplexed protein measurements (Fu et al. 2010; Potts et al. 2011). Modern multiplexed proteomics methods allow quantification of thousands of proteins in tens of samples in a single spectrometry run (McAlister et al. 2014). However, protein measurements are inevitably more difficult and less sensitive than nucleic acid measurements because proteins cannot be amplified by replication and have different physical properties and abundances. Metabolites can also be profiled by using a coupled chromatography-mass spectrometry technique.

Questions remain regarding how useful multiplexed measurements will be for predicting acute toxicity, whether they are at the level of RNA, proteins, or metabolites. Most relevant information is currently available for mRNA profiling because it has the longest history, but de-

pending on the toxic mechanism, profiling at the protein or metabolite level might be equally or more informative. Published examples show that comparison of mRNA profiles allows identification of chemicals that have common actions on cells in culture, including mechanisms that could cause acute toxicity (Lamb et al. 2006; Ravindranath et al. 2015). Those examples are encouraging, but it is important to recognize that they highlight specific success stories, not systematic toxicity prediction.

Another relevant literature is on toxicogenomic approaches to predicting toxicity of chemicals in liver and other organs. Those studies typically involve treating rodents with chemicals, harvesting organs, and using microarrays combined with pathology reports to classify effects on the liver and other organs. Pharmaceutical companies and governments have invested a great deal of resources in this approach, and major databases have recently been released to the public. The results have been mixed: considerable improvement in understanding toxic actions and identifying biomarkers but far from a complete solution to the problem of predicting liver toxicity (Chen et al. 2012). A recently created database includes expression signatures for 1,000 genes, using the L1000 assays, for treatment of tens of cell lines with thousands of chemicals (Duan et al. 2014). Analysis of that dataset should help to clarify the potential value of highly multiplexed gene-expression signatures in predictive toxicology.

In principle, multiplexed measurements at the protein or metabolite level might be expected to reveal a chemical's mechanisms of action more effectively, and with more predictive value, than gene-level measurements. More ambitiously, a combination of multiplexed measurements of mRNA, protein, and metabolite levels in parallel would in principle cover the most ground with respect to producing mechanistic information relevant to prediction of acute toxicity. That kind of integrated -omics approach has been shown to improve understanding of specific toxic mechanisms (Wilmes et al. 2013, in press) but is expensive and unproven in its usefulness for systematic testing of acute-toxicity potential. However, the data can be used to design new high-throughput assays that are mechanistically relevant to critical processes or pathways targeted by chemical-warfare agents.

High-Content Screening Assays

High-content screening (HCS) assays make multiple measurements of cell biology at the level of single cells by using microscopy or other imaging technologies. Typically, cells grown on multiwell plates are treated with a chemical, stained with several fluorescent markers that report on various aspects of metabolism and organelle health, imaged, and detected with some type of automated algorithm. HCS technology can provide much information that is relevant to specific pathways or organelles in a single assay, and it is faster and less expensive than multiplexed assays of gene or protein expression. It has been used in drug development for some time but only recently applied to toxicology. Recent studies show high potential (O'Brien 2014; Persson et al. 2014). In particular, some mechanisms of acute toxicity that might be poorly detected in gene-expression assays, such as damage to cellular membranes or organelles, can be directly assessed with HCS assays. So far, too few studies have been published to evaluate this promising technology, and key questions, such as reproducibility between laboratories, need to be addressed.

Imaging is a useful way to screen for numerous mechanistic end points at the same time, such as cell death, apoptosis, oxidative stress, mitochondrial membrane potential, DNA damage, and cell-cycle inhibition. Algorithms have been developed, for example, to predict human liver injury. Predictivity is achieved by testing enough positive and negative chemicals in the system to know the specificity and sensitivity of the platform or any particular assay. In general, these assays have high specificity but low sensitivity. In fact, in the absence of consideration of exposure, predictivity of drug-induced liver injury rarely gets above 50% (Xu et al. 2008). In addition, because these mechanistic end points are common end points of cell injury and death, it is difficult to predict particular organ toxicities. For example, the liver toxicant troglitazone and the cardiac toxicant doxorubicin both cause oxidative stress and mitochondrial dysfunction in vitro.

Other factors that contribute to toxicity in humans—such as inflammation, use of multiple drugs, and genetics—are difficult to model in simple cell systems.

Organ-on-a-Chip, Microphysiological Systems, and Advanced Organotypic Assays

The field of tissue engineering has advanced rapidly in recent years, having been stimulated especially by advances in microfabrication technology, such as micropatterning and microfluidics. The traditional goal of the tissue-engineering field is to develop replacement organs, but a shorter-term goal of generating tissue models for drug and toxicity testing has emerged (Alepee et al. 2014; Jennings 2015). Increasingly, proponents of this kind of technology are promoting “organ-on-a-chip” models as the ultimate systems for determining toxicity mechanisms in organ systems (Huh et al. 2010, 2012; Esch et al. 2011; Godoy et al. 2013; NAS 2014; Pamies et al. 2014; Sung et al. 2014). Recently, Maschmeyer et al. (in press) reported creation of long-term microphysiological systems that more closely mimic the human liver, intestinal barrier, and skin *in vivo*. Those systems might have broader applications in toxicology.

The considerations in evaluating the potential of these technologies are similar to those already discussed for organotypic models given that they are a modern extension of the organotypic models. Typically, organ-on-a-chip models are much more complex and expensive than simple cell-culture models. In theory, and in some studies, their predictive value is higher than that of simple cell culture, but their reliability needs to be evaluated, and their cost per data point might exceed that of animal models, especially if human primary cells are needed.

Induced Pluripotent Stem Cell–Derived Primary Cells

An iPS cell is a type of human pluripotent stem cell that can be differentiated into multiple types of tissue cells in cell culture. iPS cells are derived from adult tissues by forced expression of stem-cell transcription factors—an approach pioneered by Shinya Yamanaka (Takahashi and Yamanaka 2006) that avoids use of cells derived from human embryos. iPS cells can in principle provide a renewable source of almost any human primary cell type without requiring an embryo donor. iPS cells should work well with organotypic cultures, organ-on-a-chip technologies, and other biomimetic approaches that provide more realistic cell-culture models (Mathur et al. 2013). Another advantage to using iPS cells, in principle, is that they can be derived from donors who have different genotypes and might respond differently to toxicants, for example, different cytochrome p450 alleles that cause differences in drug metabolism. Thus, the long-term potential for application of iPS cells to toxicity testing is high. Nevertheless, many hurdles must be overcome, including development of methods for reliable differentiation of iPS cells into cell types that are relevant to acute toxicity. The field is promising but unlikely to be useful for chemical testing in the next 5 years. One cell type that is relevant to acute toxicity and is relatively easy to generate from iPS cells is human cardiomyocytes (Kraushaar et al. 2012); cardiotoxicity testing will therefore be a field to monitor for application of iPS technology to toxicology.

METABOLIC CONSIDERATIONS

Chemical metabolism is an important *in vivo* biological process that should be considered during the interpretation of *in vitro* testing data. There is a large capacity for metabolism in the body; metabolizing enzymes are present in the liver, lung, nasal region, and other tissues. Metabolism can convert parent chemicals into toxic metabolites (metabolic activation), into nontoxic metabolites (metabolic inactivation), or into metabolites that are rapidly removed from circulation (detoxification). Some *in vitro* systems are metabolically competent and thus provide at least some metabolism in an assay. Others are less metabolically competent and should use parent chemicals and metabolites as test agents to ensure that the appropriate chemicals are assessed. The use of structural identification tools described in Chapter 3 can help to predict toxic metabo-

lites. However, it can be difficult to identify a biologically active molecule without knowing the metabolic pathway in advance. This section summarizes metabolic assays and considerations that can complement *in vitro* testing strategies that focus on activity of a parent chemical. Although the most effective classical chemical-warfare agents do not require metabolic activation, a clear understanding of the role of chemical metabolism in toxicity has the potential to refine acute-toxicity assessments when incorporated with other testing strategies.

Assessing Reactive-Metabolite Formation

Many structural alerts⁹ that indicate the formation of reactive metabolites have been identified. Reactive metabolite formation, however, is not necessarily sufficient for toxicity to occur; in fact, many marketed drugs, which are considered safe at therapeutic dosages, contain such alerts (Kalgutkar and Dalvie 2014). Thus, it is difficult to assign toxicity to metabolites in the absence of *in vitro* experimentation. An initial strategy that could be used in the assessment of metabolite toxicity could be to identify potential metabolites by nontesting approaches (such as the use of quantitative structure–activity relationships) and then to confirm the presence of predicted chemical moieties experimentally. Formation of human metabolites can be assessed with *in vitro* systems (such as incubations with liver microsomes or hepatocytes) coupled with mass-spectrometric detection and measurement of metabolite formation. Thus, screening for toxicity can be based directly on a known metabolite or determined analytically by measuring metabolites formed during an assay.

Some chemical moieties (such as quinones and epoxides) are known to elicit toxicity. Highly electrophilic metabolites are known to be reactive with reduced glutathione (GSH), an endogenous nucleophile that plays a key role in xenobiotic metabolism and detoxification. Detection of GSH conjugates in *in vitro* assays that incorporate hepatocytes or liver microsomes provide indirect evidence that a reactive metabolite was formed.

In Vitro Systems to Study the Role of Metabolism in Toxicity Potential

A variety of *in vitro* model systems have been developed to study metabolism and include “precision-cut tissue slices, subcellular fractions such as the microsomal fraction, primary cells in suspension, primary monolayers of cells in culture, continuous cell lines, immortalized primary cells, liver-derived cell lines re-expressing biotransformation enzymes and genetically engineered cell lines expressing biotransformation enzymes” (Combes et al. 2006). Tissue fractions that contain metabolic enzymes, such as microsomes or S9 fractions, can also be introduced into an assay system to increase its metabolic competence (Glatt et al. 1989). Encapsulation of S9 in hydrogel microbeads has been recently introduced into cytotoxicity assays as one method of reducing leakage of potentially toxic microsomal lipid peroxides (Yamamoto et al. 2011).

Despite the availability of *in vivo* rodent acute-toxicity data on some chemicals, the studies will not identify toxic metabolites and resulting toxicity that are elicited via human-specific metabolic pathways not present in laboratory animal species. One approach to evaluating species differences in metabolism is to use primary human hepatocytes or other human-origin *in vitro* systems. Another approach to elucidating metabolic pathways relies on evaluating cell responses in the presence and absence of a P450 inhibitor, such as 1-aminobenzotriazole (ABT). If the formation of metabolites is inhibited by ABT, isoform-specific inhibitors can be used to identify specific isoforms involved. An individual cDNA-expressed human P450 isoforms system can be used to confirm the results of P450 isoform-specific inhibitors.

⁹As noted in Chapter 3, a structural alert is a chemical structure that has been linked to toxicity or a specific toxicity end point.

Current State of Metabolic Competence in *In Vitro* Testing Approaches

Because of their robust proliferative potential and sensitivity to toxic effects, immortalized cell lines have been identified as the system of choice for high-throughput assays. However, most of the cell lines have little or no metabolic enzyme content (that is, they are metabolically incompetent) and respond differently from tissue slices, primary cells, or tissue that is exposed to the same stimuli *in vivo*. For example, breast-cancer cell lines, such as MCF-7, are more prone to toxic effects than normal breast cells. Furthermore, although the hepatoma-derived HepG2 cells have many liver-specific functions and express conjugating enzymes, they lack functional expression of almost all the relevant human xenobiotic metabolizing enzymes in the cytochrome P450 family (Donato et al. 2008). Primary cells that are derived directly from animal or human tissues are more metabolically competent than immortalized cell lines but have much shorter half-lives and require much more care to be sustained for toxicity screening. More recent advances in cell-culture technology have improved metabolic capability in *in vitro* systems. HepaRG cells cultured in 3-D spinner-bioreactors are an attractive tool for toxicological studies and show an expression of CYP450 enzymes and phase II metabolism that more closely mimics *in vivo* conditions (Leite et al. 2012). Multicellular 3-D human primary liver cell cultures that contain hepatocytes, fibroblasts, stellate cells, and Kupffer cells have also demonstrated increased metabolic activity in the presence of fluid flow (Esch et al. 2015). In summary, cell systems have their own advantages and drawbacks, and understanding their enzymatic content will be important for estimating the effect of metabolism on *in vitro* screening data.

In the absence of metabolic activity, *in vitro* assays might still be used effectively in screening for biological activity of the parent chemical. However, *in vitro* assays designed to assess chemical metabolism or enzyme involvement might not translate directly to *in vivo* effects in that they might not recapitulate the endogenous concentrations or physiological distribution of enzymes *in vivo*. Although high-throughput screening focuses on targeted assay systems, pharmacokinetic information can be integrated independently, as further discussed in Chapter 5.

ASSAY CONSIDERATIONS FOR IMPROVING PREDICTION OF ACUTE TOXICITY

The limitations of specific *in vitro* assays have been discussed above, and here the focus is on broad steps that could be taken to improve the ability of screening assays to predict acute toxicity. The largest improvement needed is the demonstration of a linkage of assay measurements to relevant mechanisms of toxicity that quantitatively reflect an *in vivo* toxicity phenotype in target cell types. Prediction of acute toxicity would also be improved if the route of exposure were considered in assay design. The committee assumes that the relevant exposure routes are dermal and inhalation, and most assays have been designed to model oral and intravenous exposures. The issues surrounding exposure route increase as assays become higher throughput and less metabolically competent. As mentioned earlier, future assays should also take chemical metabolism into account. Those and other considerations are discussed in more detail below.

Quantitative Linkage of Assay Measurements to *In Vivo* Phenotype

Validated alternative test methods are needed for evaluating the safety of chemicals, cosmetics, and drugs. To address that need, the EU ACuteTox program assessed the ability of *in vitro* and *in silico* tools and assays to predict specific organ and system toxicity (such as hematotoxicity, neurotoxicity, nephrotoxicity, and hepatotoxicity) and intestinal absorption, distribution, and metabolism. The first “prevalidation” phase of the ACuteTox project tested a set of 57 chemicals in 50 *in vitro* and *in silico* assays or approaches and used published toxicity data as the phenotypic anchor for statistical analyses. That phase identified eight assays that showed some value in predicting acute toxicity (Clemenson et al. 2006); Box 4-1 provides more detailed information. The ACuteTox project identified several broad kinds of improvement that would need to be con-

sidered in the design of future *in vitro* screening systems, namely, improved consideration of mechanistic data and increased use of pharmacokinetic data to enhance toxicity estimates (ACuteTox 2010).

One example of mechanistically informed assay design comes from the field of testing of mitochondrial toxicity. Mitochondrial toxicity is a major contributor to organ toxicity, such as toxicity in liver, kidney, heart, muscle, and the central nervous system (Dykins and Will 2008). Acute effects can be measured rapidly in 96-well formats by using isolated mitochondrial and soluble-oxygen sensor technology (Luxcel Biosciences, Cork, Ireland) or cell-based assays that evaluate mitochondrial respiration and glycolysis (Seahorse Biosciences, Bellerica, MA) (Will and Dykens 2014). Those assays potentially can be multiplexed with readouts of mitochondrial membrane potential dissipation or cytotoxicity (ATP content), and incubation times can be tailored to be from 1–24 hours after drug exposure (Porceddu et al. 2012). In the absence of pharmacokinetic data, most biochemical and cell-based assays remain primarily ranking tools for hazard identification, not predictors of true risk.

Some *in vitro* assays have been sufficiently validated that they can largely replace animal testing. For example, the European Centre for the Validation of Alternative Methods (ECVAM) Scientific Advisory Committee endorsed the EPISKIN test and the EpiDerm method as scientifically valid replacements in a tiered testing strategy for the rabbit skin irritation method and for identifying skin irritants, respectively (Spielmann et al. 2007). In addition, the use of the Cultex® Radial Flow System to assess acute pulmonary toxicity of fine dusts and nanoparticles could possibly be adapted to test for chemicals that might be relevant to DOD (Steinritz et al. 2013). *In vivo*–*in vitro* comparison of acute respiratory tract toxicity using human 3-D airway epithelial models and human A549 and murine 3T3 monolayer cell systems has also been reported (Sauer et al. 2013).

BOX 4-1 ACuteTox Testing Strategy

The goal of the EU ACuteTox project was to evaluate whether regulatory animal tests for acute systemic toxicity could be replaced with a combination of *in vitro* assays. The ACuteTox program assessed the correlation of concentrations for *in vitro* activity with effective doses derived from whole-animal studies and evaluated a series of assays and physicochemical properties to determine how well they predicted *in vivo* acute systemic toxicity. On the basis of statistical analysis, eight test methods were found to be promising for inclusion in the testing strategy and, therefore, selected for participation in the pre-validation study:

- The neutral red uptake assay that uses the 3T3 fibroblast cell line (3T3/NRU).
- The cytokine release assay that uses human whole blood (IL-1, IL-6, and TNF- α).
- Inhibition of colony-forming-unit efficiency in human cord blood–derived cells stimulated with CFU-GM (CBC/CFU-GM).
- Gene expression (GFAP, HSP-32, MBP, and NF-H) and uridine incorporation measuring total mRNA synthesis in primary rat brain aggregate cultures.
- A panel that measures oxidative stress (intracellular peroxidative activity, intracellular concentrations of superoxide anion, and oxidized DNA base 8-oxoguanine) and cytotoxicity screening (intracellular Ca²⁺ concentrations, mitochondrial membrane potential, and plasma membrane potential) in HepG2, SH-SY5Y, and A.704 cells.
- The MTT assay that uses primary rat hepatocytes.
- Kinetic parameters (volume of distribution, protein binding, clearance, and oral absorption using Caco-2 cells and neuronal networks) for estimating the oral dose on the basis of the effective concentration observed *in vitro*.
- Estimation of chemical passage through the blood–brain barrier using neuronal networks (for neurotoxicity assays).

Sources: ACuteTox (2010); Combes et al. (2006).

In summary, evaluations of *in vitro* assays for predicting acute toxicity have focused on nonmechanistic indicators of toxicity, such as cytotoxicity assays, or low-throughput measurements. Few *in vitro* assays have been developed with quantitative linkage to any phenotype (acute or chronic). Screening for acute toxicity by using mechanisms that are likely to cause debilitating injury (Table 2-1) will require assays that are purposefully selected for their biochemical targets and characterized for their potential value in predicting human toxicity.

Assay Considerations: Lessons Learned from High-Throughput Screening Programs

The recent investment in EPA's ToxCast program and the broader Tox21 Initiative has allowed important progress in the development and use of HTS platforms to assess biological activity and potential mechanisms of action for industrial and environmental chemicals. Such programs provide rapid screening of hundreds of chemicals for dozens of cellular targets and relevant pathways. However, some chemical, cellular, and assay conditions need to be considered if one is to use and interpret the data appropriately.

First, for various reasons, the nominal chemical concentration added to the assay well is not necessarily representative of the concentration at which chemical bioactivity is observed (Groothuis et al. 2015). Chemical purity must be confirmed and solubility in the assay medium checked to determine that the initial applied concentration is accurately known. Chemical stability also needs to be monitored over the course of the assay so that chemical stability and availability can be tracked. Labile chemicals can degrade rapidly on exposure to light, aqueous conditions, or other constituents of the media or *in vitro* system. In fact, MacArthur et al. (2009) conducted chemical stability studies with cytochrome P450 assays at NCGC and found decreased chemical potency over time and lower efficacy of older samples stored in dimethyl sulfoxide (DMSO). For that reason, test chemicals at NCGC are used for no longer than 4–6 months. If degradation does occur, assay bioactivity (or lack thereof) might be inaccurately attributed to the parent chemical. Alternatively, the chemical might bind to plastics, cellular constituents, or proteins in the *in vitro* system and render it unavailable to elicit any effect in the test system (Blauboer 2010).

Second, there are limitations of current cellular systems. The cells used in HTS assays typically are selected for their proliferative capacity, adherent properties, and ease of growth in high-throughput plates and systems (Shukla et al. 2010). Immortalized cancer cell lines—such as MCF-7 (breast cancer), A549 (lung cancer), and HepG2 (liver cancer)—are commonly used. It is possible—or perhaps likely—that assessments in the limited cellular space might fail to detect chemical activities or effects that might occur in normal (nontumor) differentiated cells. In addition, proliferative cell lines have a reduced ability to metabolize parent chemicals. To address those issues, new hepatic cell lines are being developed to be more metabolically competent and are discussed further in the next paragraph.

Third, assay reproducibility can be an issue. Chemical autofluorescence and cytotoxicity are common causes of assay interference that can lead to false positives or false negatives (Huang et al. 2011). Furthermore, cells can have different levels of activity or responsiveness, depending on whether they are primary cells, differentiated cells, or immortalized cells and on how many times they have been passaged.¹⁰ Variability in metabolic capabilities among various sources of isolated primary hepatocytes is well documented and is due to numerous factors, including isolation issues and donor variability. Recently, the HepaRG cell line has been introduced as a hepatic cell line that has a degree of metabolic competence and is amenable to use in a 96-well high-throughput system (Guillouzo et al. 2007). However, maintenance of the HepaRG cells in a differentiated state requires the use of high concentrations of DMSO, which has substantial cellular effects, including inhibition of metabolizing enzymes of the cytochrome P450 family and alteration of membrane

¹⁰Cultured cells are routinely transferred and replated (subcultured) to avoid the senescence associated with high cell density. Passage number refers to the number of times that the cells have been subcultured.

permeability and antioxidant status. Whereas the HepaRG cells are recognized as metabolically competent, the competence is on a much lower scale than that of fresh or cryopreserved primary hepatocytes (Kanebratt and Andersson 2008; Lubberstedt et al. 2011).

Fourth, interpretation of activity or effective concentrations from HTS assays should be carefully considered. The concentration at which bioactivity is observed should be considered in the context of the complete activity profile among all assays tested and the range of the dose-response relationships. Review of the ToxCast data has revealed that a burst of activity in many assays might result at concentrations close to or approaching cytotoxicity (EPA 2014b). Activity measurements at high concentrations probably represent nonspecific effects and offer little information about specific bioactivity. Likewise, a lack of response can be due to tested concentrations below bioactivity, lack of representation of the biological target, or assay unreliability. The finding that cytotoxicity of some chemicals varies with the cell type emphasizes the need to characterize biological activity over a broad concentration range (Xia et al. 2008).

Fifth, assays should always include positive-control reference chemicals, whose activity can be used to determine assay variability, sensitivity, and specificity. Some of those considerations were discussed by Judson et al. (2013) and Patlewicz et al. (2013). Indeed, efforts to characterize and document nonguideline *in vitro* assays, including high-throughput and high-content assays, have been made under the auspices of OECD, which has recently published a guidance document (OECD 2014).

FINDINGS AND RECOMMENDATIONS

- **Finding:** On the basis of the results of HTS programs, *in vitro* assays have demonstrated some predictive value for acute toxicity; therefore, an *in vitro* screening approach for predicting the potential for chemicals to cause acute, debilitating injuries is theoretically feasible.

- **Finding:** Current assays were not developed for predictions of acute toxicity (particularly lethality) and have generally not dealt with chemicals that are acutely toxic or debilitating. The few evaluations of *in vitro* assays to predict acute toxicity have focused primarily on non-mechanistic indicators of toxicity, such as cytotoxicity assays, or on low-throughput measurements.

- **Finding:** There is little evidence that results of *in vitro* assays are predictive of *in vivo* outcomes of concern to DOD. A screening program for acute toxicity will require the development of new *in vitro* assays that are mechanistically relevant to critical processes or pathways that are related to acute, debilitating toxicity.

- **Finding:** Evaluations of *in vitro* assays have focused on oral and intravenous exposure. The evaluation and development of *in vitro* assays that address dermal and inhalation exposures and contact toxicity will require additional research to understand absorption and permeability in the skin and lung.

- **Finding:** Most *in vitro* assays do not account for important pharmacokinetic characteristics, such as metabolism, that can influence *in vivo* toxicity. Although *in vitro* assays lacking metabolic capacity can effectively screen for biological activity of the parent chemical, the pharmacokinetic relationship between exposure and concentration at a target site needs to be addressed.

- **Finding:** *In vitro* testing and screening programs (Tox21, ToxCast, and ACuteTox) offer a number of useful lessons regarding assay reliability, chemical solubility or purity, standardized reference chemicals, dose-response experimental designs, and standardized data processing.

- **Recommendation:** The experience of HTS programs should be considered in the design of an *in vitro* screening program to predict acute, debilitating toxicity. Because of the potential need to include highly toxic agents, if only as reference chemicals, such a screening program will need to consider health, safety, and environmental issues associated with handling highly toxic and threat agents.

REFERENCES

- Abdo, N., M. Xia, C.C. Brown, O. Kosyk, R. Huang, S. Sakamuru, Y.H. Zhou, J. Jack, P. Gallins, K. Xia, Y. Li, W.A. Chiu, A. Motsinger-Reif, C.P. Austin, R.R. Tice, I. Rusyn, and F.A. Wright. 2015. Population-based in vitro hazard and concentration-response assessment of chemicals: The 1000 Genomes High-Throughput Screening Study. *Environ. Health Perspect.* 125(5):458-466.
- ACuteTox. 2010. ACuteTox: Prevalidation, Highlights and Main Achievements of the Project. News Letter [online]. Available: http://www.acutetox.eu/Final_newsletter_Oct_2010.pdf [accessed May 7, 2015].
- Alépée, N., S. Bessou-Touya, J. Cotovio, A. de Smedt, B. de Wever, C. Faller, P. Jones, B. Le Varlet, M. Marrec-Fairley, U. Pfannenbecker, M. Tailhardat, F. van Goethem, and P. McNamee. 2013. Cosmetics Europe multi-laboratory pre-validation of the SkinEthic reconstituted human corneal epithelium test method for the prediction of eye irritation. *Toxicol. In Vitro* 27(5):1476-1488.
- Alépée, N., A. Bahinski, M. Daneshian, B. De Wever, E. Fritsche, A. Goldberg, J. Hansmann, T. Hartung, J. Haycock, H. Hogberg, L. Hoelting, J.M. Kelm, S. Kadereit, E. McVey, R. Landsiedel, M. Leist, M. Lubberstedt, F. Noor, C. Pellevoisin, D. Petersohn, U. Pfannenbecker, K. Reisinger, T. Ramirez, B. Rothen-Rutishauser, M. Schafer-Korting, K. Zeilinger, and M.G. Zurich. 2014. State-of-the-art of 3D cultures (organs-on-a-chip) in safety testing and pathophysiology. *ALTEX* 31(4):441-477.
- Ali, S., D.L. Champagne, and M.K. Richardson. 2012. Behavioral profiling of zebrafish embryos exposed to a panel of 60 water-soluble compounds. *Behav. Brain Res.* 228(2):272-283.
- Attene-Ramos, M.S., N. Miller, R. Huang, S. Michael, M. Itkin, R.J. Kavlock, C.P. Austin, P. Shinn, A. Simeonov, R.R. Tice, and M. Xia. 2013. The Tox21 robotic platform for the assessment of environmental chemicals - from vision to reality. *Drug Discov. Today* 18(15-16):716-723.
- Austin, C., R. Kavlock., and R. Tice. 2008. Tox21: Putting a Lens on the Vision of Toxicity Testing in the 21st Century [online]. Available: <http://alltox.org/tox21-putting-a-lens-on-the-vision-of-toxicity-testing-in-the-21st-century/> [accessed March 19, 2014].
- Avila, D., K. Helmcke, and M. Aschner. 2012. The *Caenorhabditis elegans* model as a reliable tool in neurotoxicology. *Hum. Exp. Toxicol.* 1(3):236-243.
- Barbazuk, W.B., I. Korf, C. Kadavi, J. Heyen, S. Tate, E. Wun, J.A. Bedell, J.D. McPherson, and S.L. Johnson. 2000. The syntenic relationship of the zebrafish and human genomes. *Genome Res.* 10(9):1351-1358.
- Basketter, D., D. Jirova, and H. Kandárová. 2012. Review of skin irritation/corrosion hazards on the basis of human data: A regulatory perspective. *Interdiscip. Toxicol.* 5(2):98-104.
- Bichara, D., N.B. Calcaterra, S. Arranz, P. Armas, and S.H. Simonetta. 2014. Set-up of an infrared fast behavioral assay using zebrafish (*Danio rerio*) larvae, and its application in compound biotoxicity screening. *J. Appl. Toxicol.* 4(2):214-219.
- Bier, E. 2005. *Drosophila*, the golden bug, emerges as a tool for human genetics. *Nat. Rev. Genet.* 6(1):9-23.
- Blaauboer, B.J. 2010. Biokinetic modeling and in vitro-in vivo extrapolations. *Toxicol. Environ. Health B Crit. Rev.* 13(2-4):242-252.
- Burns, C.G., D.J. Milan, E.J. Grande, W. Rottbauer, C.A. MacRae, and M.C. Fishman. 2005. High-throughput assay for small molecules that modulate zebrafish embryonic heart rate. *Nat. Chem. Biol.* 1(5):263-264.
- C. elegans* Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science* 282(5396):2012-2018.
- Chakravarthy, S., S. Sadagopan, A. Nair, and S.K. Sukumaran. 2014. Zebrafish as an in vivo high-throughput model for genotoxicity. *Zebrafish* 11(2):154-166.
- Chen, M., M. Zhang, J. Borlak, and W. Tong. 2012. A decade of toxicogenomic research and its contribution to toxicological science. *Toxicol. Sci.* 130(2):217-228.
- Chen, N., T.W. Harris, I. Antoshechkin, C. Bastiani, T. Bieri, D. Blasiar, K. Bradnam, P. Canaran, J. Chan, C.K. Chen, W.J. Chen, F. Cunningham, P. Davis, E. Kenny, R. Kishore, D. Lawson, R. Lee, H.M. Muller, C. Nakamura, S. Pai, P. Ozersky, A. Petcherski, A. Rogers, A. Sabo, E.M. Schwarz, K. Van Auken, Q. Wang, R. Durbin, J. Spieth, P.W. Sternberg, and L.D. Stein. 2005. WormBase: A comprehensive data resource for *Caenorhabditis* biology and genomics. *Nucleic Acids Res.* 33(Database Issue):D383-D389.
- Cho, A., J. Shin, S. Hwang, C. Kim, H. Shim, H. Kim, H. Kim, and I. Lee. 2014. WormNet v3: A network-assisted hypothesis-generating server for *Caenorhabditis elegans*. *Nucleic Acids Res.* 42(Web Server issue):W76-W82.
- Clemedson, C. 2008. The European ACuteTox project: A modern integrative in vitro approach to better prediction of acute toxicity. *Clin. Pharmacol. Ther.* 84(2):200-202.

- Clemedson, C., B. Blaauboer, J.V. Castell, P. Prieto, L. Risteli, J.A. Vericat, and A. Wendel. 2006. ACuteTox – Optimisation and pre-validation of an in-vitro test strategy for predicting human acute toxicity. *ALTEX* 23(Suppl.):254-258.
- Clemedson, C., A. Kolman, and A. Forsby. 2007. The integrated acute systemic toxicity project (ACute-Tox) for the optimisation and validation of alternative in vitro tests. *ATLA* 35(1):33-38.
- Combes, R., S. Coecke, M. Jacobs, E. Brandon, U. Bernauer, and W. Janssens. 2006. The Use of Metabolizing Systems for in Vitro Testing of Endocrine Disruptors [online]. Available: <http://www.oecd.org/chemicalsafety/testing/36280446.pdf> [accessed May 7, 2015].
- Cotovio, J., M.H. Grandidier, P. Portes, R. Roguet, and G. Rubinstenn. 2005. The in vitro skin irritation of chemicals: Optimisation of the EPISKIN prediction model within the framework of the ECVAM validation process. *ATLA* 33(4):329-349.
- Cotovio, J., M. Grandidier, D. Lelièvre, R. Roguet, E. Tinois-Tessonnaud, and J. Leclaire. 2008. In vitro acute skin irritancy of chemicals using the validated EPISKIN model in a tiered strategy: Results and performances with 184 cosmetic ingredients. *AATEX* 14(Special Issue):351-358 [online]. Available: <http://altweb.jhsph.edu/wc6/paper351.pdf> [accessed March 19, 2015].
- de Esch, C., R. Slieker, A. Wolterbeek, R. Woutersen, and D. de Groot. 2012. Zebrafish as potential model for developmental neurotoxicity testing: A mini review. *Neurotoxicol. Teratol.* 34(6):545-553.
- Dix, D.J., K.A. Houck, M.T. Martin, A.M. Richard, R.W. Setzer, and R.J. Kavlock. 2007. The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol. Sci.* 95(1):5-12.
- Donato, M.T., A. Lahoz, J.V. Castell, and M.J. Gomez-Lechon. 2008. Cell lines: A tool for in vitro drug metabolism studies. *Curr. Drug Metab.* 9(1):1-11.
- Driessen, M., A.S. Kienhuis, J.L. Pennings, T.E. Pronk, E.J. van de Brandhof, M. Roodbergen, H.P. Spaink, B. van de Water, and L.T. van der Ven. 2013. Exploring the zebrafish embryo as an alternative model for the evaluation of liver toxicity by histopathology and expression profiling. *Arch. Toxicol.* 87(5):807-823.
- Drysdale, R. 2008. FlyBase: A database for the Drosophila research community. *Methods Mol. Biol.* 420:45-59.
- Duan, Q., C. Flynn, M. Niepel, M. Hafner, J.L. Muhlich, N.F. Fernandez, A.D. Rouillard, C.M. Tan, E.Y. Chen, T.R. Golub, P.K. Sorger, A. Subramanian, and A. Ma'ayan. 2014. LINCS Canvas Browser: Interactive web app to query, browse and interrogate LINCS L1000 gene expression signatures. *Nucleic Acids Res.* 42(Web Server Issue):W449-W460.
- Dykens, J.A., and Y. Will. 2008. *Drug-Induced Mitochondrial Dysfunction*. Hoboken, NJ: Wiley.
- Ekwall, B. 1983. Screening of toxic compounds in mammalian cell cultures. *Ann. N.Y. Acad. Sci.* 407:64-77.
- EPA (US Environmental Protection Agency). 2003. Framework for Computational Toxicology Research Program in ORD. EPA/600/R-03/065. Office of Research and Development, US Environmental Protection Agency, Washington, DC [online]. Available: http://epa.gov/comptox/download_files/basic_information/comptoxframework06_02_04.pdf [accessed May 6, 2015].
- EPA (US Environmental Protection Agency). 2013. Organotypic Culture Models for Predictive Toxicology Center: Synopsis of Program [online]. Available: http://epa.gov/ncer/rfa/2013/2013_star_ocm.html [accessed March 19, 2015].
- EPA (US Environmental Protection Agency). 2014a. ToxCast Assay Annotation Version 1.0 Data User Guide. National Center for Computational Toxicology, US Environmental Protection Agency, Research Triangle Park, NC. October 2014 [online]. Available: http://epa.gov/ncct/toxcast/files/ToxCast%20Assays/ToxCast_Assay_Annotation_Data_Users_Guide_20141021.pdf [accessed March 19, 2015].
- EPA (US Environmental Protection Agency). 2014b. Meeting on Scientific Issues Associated with Integrated Endocrine Bioactivity and Exposure based Prioritization and Screening, December 2-5, 2014. FIFRA Scientific Advisory Panel, US Environmental Protection Agency [online]. Available: <http://www.epa.gov/scipoly/sap/meetings/2014/december/120214transcript.pdf> [accessed March 26, 2015].
- Esch, M.B., T.L. King, and M.L. Shuler. 2011. The role of body-on-a-chip devices in drug and toxicity studies. *Annu. Rev. Biomed. Eng.* 13:55-72.
- Esch, M.B., J.M. Prot, Y.I. Wang, P. Miller, J.R. Llamas-Vidales, B.A. Naughton, D.R. Applegate, and M.L. Shuler. 2015. Multi-cellular 3D human primary liver cell culture elevates metabolic activity under fluidic flow. *Lab Chip.* 15(10):2269-2277.
- Fabian, G., N. Farago, L.Z. Feher, L.I. Nagy, S. Kulin, K. Kitajka, T. Bitó, V. Tubak, R.L. Katona, L. Tiszlavicz, and L.G. Puskas. 2011. High-density real-time PCR-based in vivo toxicogenomic screen to predict organ-specific toxicity. *Int. J. Mol. Sci.* 12(9):6116-6134.

- Feng, Z., C.J. Cronin, J.H. Wittig, Jr., P.W. Sternberg, and W.R. Schafer. 2004. An imaging system for standardized quantitative analysis of *C. elegans* behavior. *BMC Bioinformatics* 5:115.
- Forsby, A., A.K. Bal-Price, A. Camins, S. Coecke, N. Fabre, H. Gustafsson, P. Honegger, A. Kinsner-Ovaskainen, M. Pallas, V. Rimbau, E. Rodríguez-Farré, C. Suñol, J.A. Vericat, and M.G. Zurich. 2009. Neuronal in vitro models for the estimation of acute systemic toxicity. *Toxicol. In Vitro* 23(8):1564-1569.
- Fu, Q., F.S. Schoenhoff, W.J. Savage, P. Zhang, and J.E. Van Eyk. 2010. Multiplex assays for biomarker research and clinical application: Translational science coming of age. *Proteomics Clin. Appl.* 4(3):271-284.
- Giacomotto, J., and L. Ségalat. 2010. High-throughput screening and small animal models, where are we? *Br. J. Pharmacol.* 160(2):204-216.
- Glatt, H., R. Padykula, G.A. Berchtold, G. Ludewig, K.L. Platt, J. Klein, and Oesch. 1989. Multiple activation pathways of benzene leading to products with varying genotoxic characteristics. *Environ. Health Perspect.* 82:81-89.
- Godoy, P., N.J. Hewitt, U. Albrecht, M.E. Andersen, N. Ansari, S. Bhattacharya, J.G. Bode, J. Bolleyn, C. Borner, J. Böttger, A. Braeuning, R.A. Budinsky, B. Burkhardt, N.R. Cameron, G. Camussi, C.S. Cho, Y.J. Choi, J.C. Rowlands, U. Dahmen, G. Damm, O. Dirsch, M.T. Donato, J. Dong, S. Dooley, D. Drasdo, R. Eakins, K.S. Ferreira, V. Fonsato, J. Fraczek, R. Gebhardt, A. Gibson, M. Glanemann, C.E.P. Goldring, M.J. Gómez-Lechón, G.M.M. Groothuis, L. Gustavsson, C. Guyot, D. Hallifax, J.B. Hamad, A. Hayward, D. Häussinger, C. Hellerbrand, P. Hewitt, S. Hoehme, H.G. Holzhütter, J.B. Houston, J. Hrach, K. Ito, H. Jaeschke, V. Keitel, J.M. Kelm, B.K. Park, C. Kordes, G.A. Kullak-Ublick, E.L. LeCluyse, P. Lu, J. Luebke-Wheeler, A. Lutz, D.J. Maltman, M. Matz-Soja, P. McMullen, I. Merfort, S. Messner, C. Meyer, J. Mwinyi, D.J. Naisbitt, A.K. Nussler, P. Olinga, F. Pampaloni, J. Pi, L. Pluta, S.A. Przyborski, A. Ramachandran, V. Rogiers, C. Rowe, C. Schelcher, K. Schmich, M. Schwarz, B. Singh, E.H.K. Stelzer, B. Stieger, R. Stöber, Y. Sugiyama, C. Tetta, W.E. Thasler, T. Vanhaecke, M. Vinken, T.S. Weiss, A. Widera, C.G. Woods, J. J. Xu, K.M. Yarborough, and J.G. Hengstler. 2013. Recent advances in 2D and 3D in vitro systems using primary hepatocytes, alternative hepatocyte sources and non-parenchymal liver cells and their use in investigating mechanisms of hepatotoxicity, cell signaling and ADME. *Arch. Toxicol.* 87(8):1315-1530.
- Gómez-Lechón, M.J., L. Tolosa, J.V. Castell, and M.T. Donato. 2010. Mechanism-based selection of compounds for the development of innovative in vitro approaches to hepatotoxicity studies in the HINTOP project. *Toxicol. In Vitro* 24(7):1879-1889.
- Gregory, J.M., T.P. Barros, S. Meehan, C.M. Dobson, and L.M. Luheshi. 2012. The aggregation and neurotoxicity of TDP-43 and its ALS-associated 25 kDa fragment are differentially affected by molecular chaperones in *Drosophila*. *PLoS One* 7(2):e31899.
- Groothuis, F.A., M.B. Heringa, B. Nicol, J.L. Hermens, B.J. Blaauboer, and N.I. Kramer. 2015. Dose metric considerations in in vitro assays to improve quantitative in vitro-in vivo dose extrapolations. *Toxicology* 332(5):30-40.
- Guillouzo, A., A. Corlu, C. Aninat, D. Glaise, F. Morel, and C. Guguen-Guillouzo. 2007. The human hepatoma HepaRG cells: A highly differentiated model for studies of liver metabolism and toxicity of xenobiotics. *Chem. Biol. Interact.* 168(1):66-73.
- Halle, W. 2003. The Registry of Cytotoxicity: Toxicity testing in cell cultures to predict acute toxicity (LD50) and to reduce testing in animals. *ATLA* 31(2):89-198.
- Harris, T.W., N. Chen, F. Cunningham, M. Tello-Ruiz, I. Antoshechkin, C. Bastiani, T. Bieri, D. Blasiar, K. Bradnam, J. Chan, C.K. Chan, W.J. Chen, P. Davis, E. Kenny, R. Kishore, D. Lawson, R. Lee, H.M. Muller, C. Nakamura, P. Ozersky, A. Petcherski, A. Rogers, A. Sabo, E.M. Schwarz, K. Van Auken, Q. Wang, R. Durbin, J. Spieth, P.W. Sternberg, and L.D. Stein. 2004. WormBase: A multi-species resource for nematode biology and genomics. *Nucleic Acids Res.* 32(Database Issue):D411-D417.
- Hentschel, D.M., K.M. Park, L. Cilenti, A.S. Zervos, I. Drummond, and J.V. Bonventre. 2005. Acute renal failure in zebrafish: A novel system to study a complex disease. *Am. J. Physiol. Renal. Physiol.* 288(5):F923-F929.
- Honegger, P., F. Monnet-Tschudi, and M.G. Zurich. 2009. Brain Aggregate Cell Cultures as an In Vitro Model for Neurotoxicity Testing [online]. Available: <http://alttox.org/brain-aggregate-cell-cultures-as-an-in-vitro-model-for-neurotoxicity-testing/> [accessed March 19, 2015].
- Houck, K.A., D.J. Dix, R.S. Judson, R.J. Kavlock, J. Yang, and E.L. Berg. 2009. Profiling bioactivity of the ToxCast chemical library using BioMAP primary human cell systems. *J. Biomol. Screen.* 14(9):1054-1066.

- Howe, K., M.D. Clark, C.F. Torroja, J. Tarrance, C. Berthelot, M. Muffato, J.E. Collins, S. Humphray, K. McLaren, L. Matthews, S. McLaren, L. Sealy, M. Caccamo, C. Churcher, C. Scott, J.C. Barrett, R. Koch, G.J. Rauch, S. White, W. Chow, B. Kilian, L.T. Quintais, J.A. Guerra-Assunção, Y. Zhou, Y. Gu, J. Yen, J.H. Vogel, T. Eyre, S. Redmond, R. Banerjee, J. Chi, B. Fu, E. Langley, S.F. Maguire, G.K. Laird, D. Lloyd, E. Kenyon, S. Donaldson, H. Sehra, J. Almeida-King, J. Loveland, S. Trevanion, M. Jones, M. Quail, D. Willey, A. Hunt, J. Burton, S. Sims, K. McLay, B. Plumb, J. Davis, C. Clee, K. Oliver, R. Clark, C. Riddle, D. Elliot, G. Threadgold, G. Harden, D. Ware, S. Begum, B. Mortimore, G. Kerry, P. Heath, B. Phillimore, A. Tracey, N. Corby, M. Dunn, C. Johnson, J. Wood, S. Clark, S. Pelan, G. Griffiths, M. Smith, R. Glithero, P. Howden, N. Barker, C. Lloyd, C. Stevens, J. Harley, K. Holt, G. Panagiotidis, J. Lovell, H. Beasley, C. Henderson, D. Gordon, K. Auger, D. Wrigh, J. Collins, C. Raisen, L. Dyer, K. Leung, I. Robertson, K. Ambridge, D. Leongamornlert, S. McGuire, R. Gilderthorp, C. Griffiths, D. Manthravadi, S. Nichol, G. Barker, S. Whitehead, M. Kay, J. Brown, C. Murnane, E. Gray, M. Humphries, N. Sycamore, D. Barker, D. Saunders, J. Wallis, A. Babbage, S. Hammond, M. Mashreghi-Mohammadi, L. Barr, S. Martin, P. Wray, A. Ellington, N. Matthews, M. Ellwood, R. Woodmansey, G. Clark, J. Cooper, A. Tromans, D. Grafham, C. Skuce, R. Pandian, R. Andrews, E. Harrison, A. Kimberley, J. Garnett, N. Fosker, R. Hall, P. Garner, D. Kelly, C. Bird, S. Palmer, I. Gehring, A. Berger, C.M. Dooley, Z. Ersan-Urün, C. Eser, H. Geiger, M. Geisler, L. Karotki, A. Kirn, J. Konantz, M. Konantz, M. Oberländer, S. Rudolph-Geiger, M. Teucke, C. Lanz, G. Raddatz, K. Osoegawa, B. Zhu, A. Rapp, S. Widaa, C. Langford, F. Yang, S.C. Schuster, N.P. Carter, J. Harrow, Z. Ning, J. Herrero, S.M. Searle, A. Enright, R. Geisler, R.H. Plasterk, C. Lee, M. Westerfield, P.J. de Jong, L.I. Zon, J.H. Postlethwait, C. Nüsslein-Volhard, T.J. Hubbard, H. Roest Crollius, J. Rogers, and D.L. Stemple. 2013. The zebrafish reference genome sequence and its relationship to the human genome. *Nature* 496(7446):498-503.
- Huang, C.J., C.T. Tu, C.D. Hsiao, F.J. Hsieh, and H.J. Tsai. 2003. Germ-line transmission of a myocardium-specific GFP transgene reveals critical regulatory elements in the cardiac myosin light chain 2 promoter of zebrafish. *Dev. Dyn.* 228(1):30-40.
- Huang, R., N. Southall, M.H. Cho, M. Xia, J. Inglese, and C.P. Austin. 2008. Characterization of diversity in toxicity mechanism using in vitro cytotoxicity assays in quantitative high throughput screening. *Chem. Res. Toxicol.* 21(3):659-667.
- Huang, R., M. Xia, M.H. Cho, S. Sakamuru, P. Shinn, K.A. Houck, D.J. Dix, R.S. Judson, K.I. Witt, R.J. Kavlock, R.R. Tice, and C.P. Austin. 2011. Chemical genomics profiling of environmental chemical modulation of human nuclear receptors. *Environ. Health Perspect.* 119(8):1142-1148.
- Huh, D., B.D. Matthews, A. Mammoto, M. Montoya-Zavala, H.Y. Hsin, and D.E. Ingber. 2010. Reconstituting organ-level lung functions on a chip. *Science* 328(5986):1662-1668.
- Huh, D., D.C. Leslie, B.D. Matthews, J.P. Fraser, S. Jurek, G.A. Hamilton, K.S. Thorneloe, M.A. McAlexander, and D.E. Ingber. 2012. A human disease model of drug toxicity-induced pulmonary edema in a lung-on-a-chip microdevice. *Sci. Transl. Med.* 4(159):159ra147.
- Jennings, P. 2015. The future of in vitro toxicology. *Toxicol. In Vitro* 29(6):1217-1221.
- Judson, R.S., K.A. Houck, R.J. Kavlock, T.B. Knudsen, M.T. Martin, H.M. Mortensen, D.M. Reif, D.M. Rotroff, I. Shah, A.M. Richard, and D.J. Dix. 2010. In vitro screening of environmental chemicals for targeted testing prioritization: The ToxCast project. *Environ. Health Perspect.* 118(4):485-492.
- Judson, R., R. Kavlock, M. Martin, D. Reif, K. Houck, T. Knudsen, A. Richard, R.R. Tice, M. Whelan, M. Xia, R. Huang, C. Austin, G. Daston, T. Hartung, J.R. Fowle III, W. Wooge, W. Tong, and D. Dix. 2013. Perspectives on validation of high-throughput assays supporting 21st century toxicity testing. *ALTEX* 30(1):51-56.
- Kalgutkar, A.S., and D. Dalvie. 2014. Predicting toxicities of reactive metabolite-positive drug candidates. *Annu. Rev. Pharmacol. Toxicol.* 55:35-54.
- Kanebratt, K.P., and T.B. Andersson. 2008. Evaluation of HepaRG cells as an in vitro model for human drug metabolism studies. *Drug Metab. Dispos.* 36(7):1444-1452.
- Kavlock, R.J., C.P. Austin, and R.R. Tice. 2009. Toxicity testing in the 21st century: Implications for human health risk assessment. *Risk Anal.* 29(4):485-487.
- Kavlock, R.J., K. Chandler, K. Houck, S. Hunter, R. Judson, N. Kleinstreuer, T. Knudsen, M. Martin, S. Padilla, D. Reif, A. Richard, D. Rotroff, N. Sipes, and D. Dix. 2012. Update on EPA's ToxCast program: Providing high throughput decision support tools for chemical risk management. *Chem. Res. Toxicol.* 25(7):1287-1302.
- Khetani, S.R., and S.N. Bhatia. 2008. Microscale culture of human liver cells for drug development. *Nat. Biotechnol.* 26(1):120-126.

- Khetani, S.R., C. Kanchagar, O. Ukairo, S. Krzyzewski, A. Moore, J. Shi, S. Aoyama, M. Aleo, and Y. Will. 2013. Use of micropatterned cocultures to detect compounds that cause drug-induced liver injury in humans. *Toxicol. Sci.* 132(1):107-117.
- Kinsner-Ovaskainen, A., P. Prieto, S. Stanzel, and A. Kopp-Schneider. 2013. Selection of test methods to be included in a testing strategy to predict acute oral toxicity: An approach based on statistical analysis of data collected in phase 1 of the ACuteTox project. *Toxicol. In Vitro* 27(4):1377-1394.
- Klaper, R., D. Arndt, J. Bozich, and G. Dominguez. 2014. Molecular interactions of nanomaterials and organisms: Defining biomarkers for toxicity and high-throughput screening using traditional and next-generation sequencing approaches. *Analyst*. 139(5):882-895.
- Kleinstreuer, N.C., J. Yang, E.L. Berg, T.B. Knudsen, A.M. Richard, M.T. Martin, D.M. Reif, R.S. Judson, M. Polokoff, D.J. Dix, R.J. Kavlock, and K.A. Houck. 2014. Phenotypic screening of the ToxCast chemical library to classify toxic and therapeutic mechanisms. *Nat. Biotechnol.* 32(6):583-591.
- Knudsen, T.B., R.J. Kavlock, G.P. Daston, D. Stedman, M. Hixon, and J.H. Kim. 2011. Developmental toxicity testing for safety assessment: New approaches and technologies. *Birth Defects Res. B Dev. Reprod. Toxicol.* 92(5):413-420.
- Kraushaar, U., T. Meyer, D. Hess, L. Gepstein, C.L. Mummery, S.R. Braam, and E. Guenther. 2012. Cardiac safety pharmacology: From human ether-a-gogo related gene channel block towards induced pluripotent stem cell based disease models. *Expert Opin. Drug Saf.* 11(2):285-298.
- Krewski, D., M.E., Andersen, E. Mantus, and L. Zeise. 2009. Toxicity testing in the 21st century: Implications for human health risk assessment. *Risk Anal.* 29(4):474-479.
- Lamb, J., E.D. Crawford, D. Peck, J.W. Modell, I.C. Blat, M.J. Wrobel, J. Lerner, J.P. Brunet, A. Subramanian, K.N. Ross, M. Reich, H. Hieronymus, G. Wei, S.A. Armstrong, S.J. Haggarty, P.A. Clemons, R. Wei, S.A. Carr, E.S. Lander, and T.R. Golub. 2006. The Connectivity Map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313(5795):1929-1935.
- Langheinrich, U., G. Vacun, and T. Wagner. 2003. Zebrafish embryos express an orthologue of HERG and are sensitive toward a range of QT-prolonging drugs inducing severe arrhythmia. *Toxicol. Appl. Pharmacol.* 193(3):370-382.
- Larson, B., P. Banks, S. Shultz, M. Sobol, and J.J. Cali. 2011. The utility of semi-automating multiplexed assays for ADME/Tox applications. *Comb. Chem. High Throughput Screen.* 14(8):658-668.
- Legradi, J., N. El Abdellaoui, M. van Pomeran, and J. Legler. In press. Comparability of behavioural assays using zebrafish larvae to assess neurotoxicity. *Environ. Sci. Pollut. Res.*
- Leite, S.B., I. Wilk-Zasadna, J.M. Zaldivar, E. Airola, M.A. Reis-Fernandes, M. Mennecozzi, C. Guguen-Guillouzo, C. Chesne, C. Guillou, P.A. Alves, and S. Coecke. 2012. Three-dimensional HepaRG model as an attractive tool for toxicity testing. *Toxicol. Sci.* 130(1):106-116.
- Lessman, C.A. 2011. The developing zebrafish (Danio rerio): A vertebrate model for high-throughput screening of chemical libraries. *Birth Defects Res. Part C Embryo Today* 93(3):268-280.
- Limbourg, B., and M. Zalokar. 1973. Permeabilization of Drosophila eggs. *Dev. Biol.* 35(2):382-387.
- Lin, Z., and Y. Will. 2012. Evaluation of drugs with specific organ toxicities in organ-specific cell lines. *Toxicol. Sci.* 126(1):114-127.
- Love, D.R., F.B. Pichler, A. Dodd, B.R. Copp, and D.R. Greenwood. 2004. Technology for high-throughput screens: The present and future using zebrafish. *Curr. Opin. Biotechnol.* 15(6):564-571.
- Lubberstedt, M., U. Muller-Vieira, M. Mayer, K.M. Biemel, F. Knospel, D. Knobeloch, A.K. Nussler, J.C. Gerlach, and K. Zeilinger. 2011. HepaRG human hepatic cell line utility as a surrogate for primary human hepatocytes in drug metabolism assessment in vitro. *J. Pharmacol. Toxicol. Methods* 63(1):59-68.
- MacArthur, R., W. Leister, H. Veith, P. Shinn, N. Southall, C.P. Austin, J. Inglese, and D.S. Auld. 2009. Monitoring compound integrity with cytochrome P450 assays and qHTS. *J. Biomol. Screen* 14(5):538-546.
- Marroquin, L.D., J. Hynes, J.A. Dykens, J.D. Jamieson, and Y. Will. 2007. Circumventing the Crabtree effect: Replacing media glucose with galactose increases susceptibility of HepG2 cells to mitochondrial toxicants. *Toxicol. Sci.* 97(2):539-547.
- Maschmeyer, I., T. Hasenberg, A. Jaenicke, M. Lindner, A.K. Lorenz, J. Zech, L.A. Garbe, F. Sonntag, P. Hayden, S. Ayehunie, R. Lauster, U. Marx, E.M. Materne. In press. Chip-based human liver-intestine and liver-skin co-cultures - A first step toward systemic repeated dose substance testing in vitro. *Eur. J. Pharm. Biopharm.*
- Mathur, A., P. Loskill, S. Hong, J. Lee, S.G. Marcus, L. Dumont, B.R. Conklin, H. Willenbring, L.P. Lee, and K.E. Healy. 2013. Human induced pluripotent stem cell-based microphysiological tissue models of myocardium and liver for drug development. *Stem Cell Res. Ther.* 4(Suppl. 1):S14.

- Mazzio, E.A., Y.I. Soliman, and K.F. Soliman. 2010. Variable toxicological response to the loss of OXPHOS through 1-methyl-4-phenylpyridinium-induced mitochondrial damage and anoxia in diverse neural immortal cell lines. *Cell Biol. Toxicol.* 26(6):527-539.
- McAlister, G.C., D.P. Nusinow, M.P. Jedrychowski, M. Wühr, E.L. Huttlin, B.K. Erickson, R. Rad, W. Haas, and S.P. Gygi. 2014. MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal. Chem.* 86(14):7150-7158.
- McVey, K.A., J.A. Mink, I.B. Snapp, W.S. Timberlake, C.E. Todt, R. Negga, and V.A. Fitsanakis. 2012. *Caenorhabditis elegans*: An emerging model system for pesticide neurotoxicity. *J. Environ. Anal. Toxicol.* 2:125, doi: 10.4172/2161-0525.S4-003.
- Melstrom, P.C., and P.L. Williams. 2007. Reversible AChE inhibitors in *C. elegans* vs. rats, mice. *Biochem. Biophys. Res. Commun.* 357(1):200-205.
- Messner, S., I. Agarkova, W. Moritz, and J.M. Kelm. 2013. Multi-cell type human liver in microtissues for hepatotoxicity testing. *Arch. Toxicol.* 87(1):209-213.
- Meyer, D., and P.L. Williams. 2014. Toxicity testing of neurotoxic pesticides in *Caenorhabditis elegans*. *J. Toxicol. Environ. Health B Crit. Rev.* 17(5):284-306.
- Milan, D.J., T.A. Peterson, J.N. Ruskin, R.T. Peterson, and C.A. MacRae. 2003. Drugs that induce repolarization abnormalities cause bradycardia in zebrafish. *Circulation.* 107(10):1355-1358.
- NAS (National Academy of Sciences). 2014. The Potential of the Tissue Chip for Environmental Health Studies. Emerging Science For Environmental Health Decisions Newsletter Issue II [online]. Available: <http://nas-sites.org/emergingscience/files/2013/05/tissue-chips-final.pdf> [accessed March 20, 2015].
- Nichols, C.D. 2006. *Drosophila melanogaster* neurobiology, neuropharmacology, and how the fly can inform central nervous system drug discovery. *Pharmacol. Ther.* 112(3):677-700.
- NIEHS (National Institute of Environmental Health Sciences). 2001a. Report of the International Workshop on In Vitro Methods for Assessing Acute Systemic Toxicity. NIH Publication 01-4499. National Toxicology Program [online]. Available: <http://ntp.niehs.nih.gov/pubhealth/evalatm/test-method-evaluations/acute-systemic-tox/in-vitro/wksp-rpt-recs/index.html#Complete-Report> [accessed March 20, 2015].
- NIEHS (National Institute of Environmental Health Sciences). 2001b. Guidance Document on Using In Vitro Data to Estimate In Vivo Starting Doses for Acute Toxicity Based on Recommendations from an International Workshop Organized by ICCVAM and NICEA TM. NIH Publication 01-4500. National Toxicology Program [online]. Available: <http://ntp.niehs.nih.gov/pubhealth/evalatm/test-method-evaluations/acute-systemic-tox/in-vitro/guidance-document/index.html> [accessed March 20, 2015].
- NRC (National Research Council). 2007. *Toxicity Testing in the 21st Century*. Washington, DC: The National Academies Press.
- NTP (National Toxicology Program). 2004. A National Toxicology Program for the 21st Century: A Roadmap for the Future [online]. Available: https://ntp.niehs.nih.gov/ntp/about_ntp/ntpvision/ntproadmap_508.pdf [accessed March 23, 2015].
- O'Brien, P. 2014. High-content analysis in toxicology: Screening substances for human toxicity potential, elucidating subcellular mechanisms and in vivo use as translational safety biomarkers. *Basic Clin. Pharmacol. Toxicol.* 115(1):4-17.
- O'Brien, P.J., W. Irwin, D. Diaz, E. Howard-Cofield, C.M. Krejsa, M.R. Slaughter, B. Gao, N. Kaludercic, A. Angeline, P. Bernardi, P. Brain, and C. Hougham. 2006. High concordance of drug-induced human hepatotoxicity with in vitro cytotoxicity measured in a novel cell-based model using high content screening. *Arch. Toxicol.* 80(9):580-604.
- OECD (Organization for Economic Co-operation and Development). 2013a. Test No. 210: Fish, Early-life Stage Toxicity Test. OECD Guidelines for the Testing of Chemicals, Section 2. Paris, France: OECD Publishing.
- OECD (Organization for Economic Co-operation and Development). 2013b. Test No. 236: Fish Embryo Acute Toxicity (FET) Test. OECD Guidelines for the Testing of Chemicals, Section 2. Paris, France: OECD Publishing.
- OECD (Organization for Economic Co-operation and Development). 2014. Guidance Document for Describing Non-guideline In Vitro Test Methods. Series on Testing and Assessment No. 211 [online]. Available: [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO\(2014\)35&doclanguage=en](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2014)35&doclanguage=en) [accessed March 24, 2015].
- Pamies, D., T. Hartung, and H.T. Hogberg. 2014. Biological and medical applications of a brain-on-a-chip. *Exp. Biol. Med.* 239(9):1096-1107.
- Patlewicz, G., T. Simon, K. Goyak, R.D. Phillips, J.C. Rowlands, S.D. Seidel, and R.A. Becker. 2013. Use and validation of HT/HC assays to support 21st century toxicity evaluations. *Regul. Toxicol. Pharmacol.* 65(2):259-268.

- Perkins, E.J., G.T. Ankley, K.M. Crofton, N. Garcia-Reyero, C.A. LaLone, M.S. Johnson, J.E. Tietge, and D.L. Villeneuve. 2013. Current perspectives on the use of alternative species in human health and ecological hazard assessments. *Environ. Health Perspect.* 121(9):1002-1010.
- Persson, M., A. Løye, M. Jacquet, N. Mow, A. Thougard, T. Mow, and J. Hornberg. 2014. High-content analysis/screening for predictive toxicology: Application to hepatotoxicity and genotoxicity. *Basic Clin. Pharmacol. Toxicol.* 115(1):18-23.
- Peterson, R.T., R. Nass, W.A. Boyd, J.H. Freedman, K. Dong, and T. Narahashi. 2008. Use of non-mammalian alternative models for neurotoxicological study. *Neurotoxicology* 29(3):546-555.
- Pfannenbecker, U., S. Bessou-Touya, C. Faller, J. Harbell, T. Jacob, H. Raabe, M. Tailhardat, N. Alepee, A. De Smedt, B. De Wever, P. Jones, Y. Kaluzhny, B. Le Varlet, P. McNamee, M. Marrec-Fairley, and F. Van Goethem. 2013. Cosmetics Europe multi-laboratory pre-validation of the epicular reconstituted human tissue test method for the prediction of eye irritation. *Toxicol. In Vitro* 27(2):619-626.
- Podratz, J.L., N.P. Staff, D. Froemel, A. Wallner, F. Wabnig, A.J. Bieber, A. Tang, and A.J. Windebank. 2011. *Drosophila melanogaster*: A new model to study cisplatin-induced neurotoxicity. *Neurobiol. Dis.* 43(2):330-337.
- Podratz, J.L., N.P. Staff, J.B. Boesche, N.J. Giorno, M.E. Hainy, S.A. Herring, M.T. Klennert, C. Milaster, S.E. Nowakowski, R.G. Krug II, Y. Peng, and A.J. Windebank. 2013. An automated climbing apparatus to measure chemotherapy-induced neurotoxicity in *Drosophila melanogaster*. *Fly (Austin)* 7(3):187-192.
- Porceddu, M., N. Buron, C. Roussel, G. Labbe, B. Fromenty, and A. Borgne-Sanchez. 2012. Prediction of liver injury induced by chemicals in human with a multiparametric assay on isolated mouse liver mitochondria. *Toxicol. Sci.* 129(2):332-345.
- Potts, S.J., T.D. Johnson, F.A. Voelker, H. Lange, and G.D. Young. 2011. Multiplexed measurement of proteins in tissue in a clinical environment. *Appl. Immunohistochem. Mol. Morphol.* 19(6):494-498.
- Raldúa, D., and B. Piña. 2014. In vivo zebrafish assays for analyzing drug toxicity. *Expert Opin. Drug Metab. Toxicol.* 10(5):685-697.
- Rand, M.D. 2010. *Drosophila* neurotoxicology: The growing potential for *Drosophila* in neurotoxicology. *Neurotoxicol. Teratol.* 32(1):74-83.
- Ravindranath, A.C., N. Perualila-Tan, A. Kasim, G. Drakakis, S. Liggi, S.C. Brewerton, D. Mason, M.J. Bodkin, D.A. Evans, A. Bhagwat, W. Talloen, H.W. Göhlmann, Z. Shkedy, and A. Bender. 2015. Connecting gene expression data from connectivity map and in silico target predictions for small molecule mechanism-of-action analysis. *Mol. BioSyst.* 11(1):86-96.
- Ruan, Q.L., J.J. Ju, Y.H. Li, R. Liu, Y.P. Pu, L.H. Yin, and D.Y. Wang. 2009. Evaluation of pesticide toxicities with differing mechanisms using *Caenorhabditis elegans*. *J. Toxicol. Environ. Health A.* 72(11-12):746-751.
- Rubin, G.M., and E.B. Lewis. 2000. A brief history of *Drosophila*'s contributions to genome research. *Science* 24(5461):2216-2218.
- Sauer, U.G., S. Vogel, A. Hess, S.N. Kolle, L. Ma-Hock, B. van Ravenzwaay, and R. Landsiedel. 2013. In vivo-in vitro comparison of acute respiratory tract toxicity using human 3D airway epithelial models and human A549 and murine 3T3 monolayer cell systems. *Toxicol. In Vitro* 27(1):174-190.
- Sauer, U.G., A. Aumann, L. Ma-Hock, R. Landsiedel, and W. Wohlleben. 2015. Influence of dispersive agent on nanomaterial agglomeration and implications for biological effects in vivo or in vitro. *Toxicol. In Vitro* 29(1):182-186.
- Sawin-McCormack, E.P., M.B. Sokolowski, and A.R. Campos. 1995. Characterization and genetic analysis of *Drosophila melanogaster* photobehavior during larval development. *J. Neurogenet.* 10(2):119-135.
- Segalat, L. 2007. Invertebrate animal models of diseases as screening tools in drug discovery. *ACS Chem. Biol.* 2(4):231-236.
- Shukla, S.J., R. Huang, C.P. Austin, and M. Xia. 2010. The future of toxicity testing: A focus on in vitro methods using a quantitative high throughput screening platform. *Drug Discov. Today* 15(23-24):997-1007.
- Simoneschi, D., F. Simoneschi, and N.E. Todd. 2014. Assessment of cardiotoxicity and effects of malathion on the early development of zebrafish (*Danio rerio*) using computer vision for heart rate quantification. *Zebrafish* 11(3):275-280.
- Sipes, N.S., M.T. Martin, P. Kothiya, D.M. Reif, R.S. Judson, A.M. Richard, K.A. Houck, D.J. Dix, R.J. Kavlock, and T.B. Knudsen. 2013. Profiling 976 ToxCast chemicals across 331 enzymatic and receptor signaling assays. *Chem. Res. Toxicol.* 26(6):878-895.
- Sokolowski, M.B. 2001. *Drosophila*: Genetics meets behaviour. *Nat. Rev. Genet.* 2(11):879-890.

- Spielmann, H., E. Genschow, M. Liebsch, and W. Halle. 1999. Determination of the starting dose for acute oral toxicity (LD50) testing in the up and down procedure (UDP) from cytotoxicity data. *ATLA* 27(6):957-966.
- Spielmann, H., S. Hoffmann, M. Liebsch, P. Botham, J.H. Fentem, C. Eskes, R. Roguet, J. Cotovio, T. Cole, A. Worth, J. Heylings, P. Jones, C. Robles, H. Kandarova, A. Gamer, M. Remmele, R. Curren, H. Raabe, A. Cockshott, I. Gerner, and V. Zuang. 2007. The ECVAM international validation study on in vitro tests for acute skin irritation: Report on the validity of the EPISKIN and EpiDerm assays and on the Skin Integrity Function Test. *ATLA* 35(6):559-601.
- Steinritz, D., N. Möhle, C. Pohl, M. Papritz, B. Stenger, A. Schmidt, C.J. Kirkpatrick, H. Thiermann, R. Vogel, S. Hoffmann, and M. Aufderheide. 2013. Use of the Cultex® Radial Flow System as an in vitro exposure method to assess acute pulmonary toxicity of fine dusts and nanoparticles with special focus on the intra- and inter-laboratory reproducibility. *Chem. Biol. Interact.* 206(3):479-490.
- Sung, J.H., B. Srinivasan, M.B. Esch, W.T. McLamb, C. Bernabini, M.L. Shuler, and J.J. Hickman. 2014. Using physiologically-based pharmacokinetic-guided “body-on-a-chip” systems to predict mammalian response to drug and chemical exposure. *Exp. Biol. Med. (Maywood)* 239(9):1225-1239.
- Takahashi, K., and S. Yamanaka. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126 (4):663-676.
- Taylor, K.L., N.J. Grant, N.D. Temperley, and E.E. Patton. 2010. Small molecule screening in zebrafish: An in vivo approach to identifying new chemical tools and drug leads. *Cell Commun. Signal* 8:11.
- Thomas, R.S., M.B. Black, L. Li, E. Healy, T.M. Chu, W. Bao, M.E. Andersen, and R.D. Wolfinger. 2012. A comprehensive statistical analysis of predicting in vivo hazard using high-throughput in vitro screening. *Toxicol. Sci.* 128(2):398-417.
- Tsang, M. 2010. Zebrafish: A tool for chemical screens. *Birth Defects Res. Part C Embryo Today* 90(3):185-192.
- Villatte, F., V. Marcel, S. Estrada-Mondaca, and D. Fournier. 1998. Engineering sensitive acetylcholinesterase for detection of organophosphate and carbamate insecticides. *Biosens. Bioelectron.* 13(2):157-164.
- Vliegenthart, A.D., C.S. Tucker, J. Del Pozo, and J.W. Dear. 2014. Zebrafish as model organisms for studying drug induced liver injury. *Br. J. Clin. Pharmacol.* 78(6):1217-1227.
- Whitworth, A.J., P.D. Wes, and L.J. Pallanck. 2006. Drosophila models pioneer a new approach to drug discovery for Parkinson's disease. *Drug Discov. Today* 11(3-4):119-126.
- Will, Y., and J. Dykens. 2014. Mitochondrial toxicity assessment in industry—a decade of technology development and insight. *Expert Opin. Drug Metab. Toxicol.* 10(8):1061-1067.
- Williams, P.L., and D.B. Dusenbery. 1990. A promising indicator of neurobehavioral toxicity using the nematode *Caenorhabditis elegans* and computer tracking. *Toxicol. Ind. Health* 6(3-4):425-440.
- Wilmes, A., A. Limonciel, L. Aschauer, K. Moenks, C. Leonard, M.O. Leonard, J. Hamon, D. Carpi, S. Ruzek, A. Handler, O. Schmal, K. Herrgen, P. Bellwon, C. Burek, G.L. Truissi, P. Hewitt, E. Di Consiglio, E. Testai, B.J. Blaauw, C. Guillou, C.G. Huber, A. Lukas, W. Pfaller, S.O. Mueller, F.Y. Bois, W. Dekant, and P. Jennings. 2013. Application of integrated transcriptomic, proteomic and metabolomic profiling for the delineation of mechanisms of drug induced cell stress. *J. Proteomics.* 79:180-194.
- Wilmes, A., C. Bielow, C. Ranninger, P. Bellwon, L. Aschauer, A. Limonciel, H. Chassaigne, T. Kristl, S. Aiche, C.G. Huber, C. Guillou, P. Hewitt, M.O. Leonard, W. Dekant, F. Bois, and P. Jennings. In press. Mechanism of cisplatin proximal tubule toxicity revealed by integrating transcriptomics, proteomics, metabolomics and biokinetics. *Toxicology In Vitro*.
- Wink, S., S. Hiemstra, S. Huppelschoten, E. Danen, M. Niemeijer, G. Hendriks, H. Vrieling, B. Herpers, and B. van de Water. 2014. Quantitative high content imaging of cellular adaptive stress response pathways in toxicity for chemical safety assessment. *Chem. Res. Toxicol.* 27(3):338-355.
- Wu, T.S., J.J. Yang, F.Y. Yu, and B.H. Liu. 2012. Evaluation of nephrotoxic effects of mycotoxins, citrinin and patulin, on zebrafish (*Danio rerio*) embryos. *Food Chem. Toxicol.* 50(12):4398-4404.
- Xia, M., R. Huang, K.L. Witt, N. Southall, J. Fostel, M.H. Cho, A. Jadhav, C.S. Smith, J. Inglese, C.J. Portier, R.R. Tice, and C.P. Austin. 2008. Compound cytotoxicity profiling using quantitative high-throughput screening. *Environ. Health Perspect.* 116(3):284-291.
- Xu, J.J., P.V. Henstock, M.C. Dunn, A.R. Smith, J.R. Chabot, and D. de Graaf. 2008. Cellular imaging predictions of clinical drug-induced liver injury. *Toxicol. Sci.* 105(1):97-105.
- Yamamoto, N., K. Komori, K. Montagne, H. Matsui, H. Nakayama, S. Takeuchi, and Y. Sakai. 2011. Cytotoxicity evaluation of reactive metabolites using rat liver homogenate microsomes-encapsulated alginate gel microbeads. *J. Biosci. Bioenerg.* 111(4):454-458.

Yen, J., S. Donerly, E.D. Levin, and E.A. Linney. 2011. Differential acetylcholinesterase inhibition of chlorpyrifos, diazinon and parathion in larval zebrafish. *Neurotoxicol. Teratol.* 33(6):735-741.

5

Integration and Decision-Making for Predictive Toxicology

As described in Chapter 2, a robust integration and decision-making strategy is needed as part of the overall approach developed by the committee to predict acute, debilitating toxicity. Specifically, this chapter will describe general approaches and considerations for integrating data to make the categorization decisions outlined in the tiered prioritization strategy described in Chapter 2. Discussions of these topics are related to the task of evaluating chemicals for their potential to elicit acute toxicity. The committee has also noted a number of topics beyond its charge on which the Department of Defense (DOD) will need to make policy decisions in light of specific needs.

GENERAL APPROACH TO INTEGRATION AND DECISION-MAKING

Integration and decision-making to support prediction of the potential of chemicals to cause acute toxicity are needed at many levels. As described in Chapter 2, the goal at each tier of the prioritization strategy is to place chemicals in three categories: “high confidence of high toxicity,” “high confidence of low toxicity,” and “inadequate data.” Box 5-1 presents a simplified illustration of the process to base decisions on the results of a single model for a single end point. As illustrated in this simple case, categorization depends on defining clear benchmarks that set the boundaries for “high” and “low” toxicity and on taking uncertainty or confidence in each individual prediction into account. Key tasks for DOD will be determining the appropriate benchmarks for each end point that is relevant to the evaluation of acute toxicity and specifying the level of confidence appropriate to its needs.

In the more general case in which there are several predictions for multiple end points, the committee divided the integration and decision-making process into two parts, as illustrated in Figure 5-1:

- “Within–end point” integration and decision-making, which is based on integrating various data streams and predictions that inform a single acute-toxicity end point.
- “Cross–end point” integration and decision-making, which is based on integrating predictions from several acute-toxicity end points.

Within–End Point Integration and Decision-making

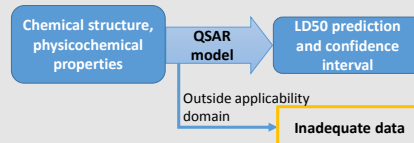
As discussed in Chapter 2, concern about potential acute toxicity spans a wide range of chemicals in terms of structures and physicochemical properties. Moreover, as shown in Chapters 3 and 4 and by others (Bauer-Mehren et al. 2012), no single prediction approach, whether a nontesting approach or an assay-based approach, is sufficient to capture the entire chemical domain. For some end points, such as a rat LD₅₀, multiple models and tools that have

various degrees of accuracy are available (see, for example, review by Diaz et al. 2015), and it might be of interest to integrate their predictions into a single summary prediction. Furthermore, a chemical's pharmacokinetics—absorption, distribution, metabolism, and excretion (ADME)—might contribute to its potency and toxicity. Therefore, even within an “end point” domain, there might be a need to integrate multiple databases, assays, models, and tools to develop an “integrated” prediction for that end point. The committee notes that using various integrative approaches might also help to identify biological responses that can explain chemical-induced adverse reactions (Bauer-Mehren et al. 2012).

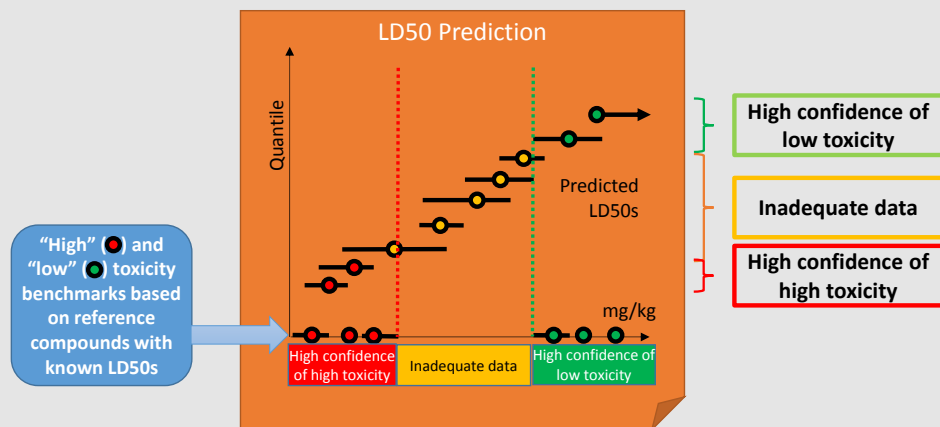
BOX 5-1 Simplified Illustration of Integration and Decision-making

Models and End Points: At Tier 1, a single quantitative structure–activity relationship (QSAR) model is being used to predict rat LD₅₀ values from chemical structure and physicochemical properties. No other models or end points are being considered. To illustrate a simplified approach, chemicals outside a specified applicability domain are placed in the “inadequate data” category for this example. For chemicals inside the applicability domain, the output of the model is an LD₅₀ estimate with a confidence interval that reflects uncertainty.

Integration: Because only a single model and a single end point are being considered, there is no integration of different predictions.



Category Benchmarks: A set of reference chemicals based on DOD interests that have known (possibly more than one) LD₅₀ values are selected to represent the “high toxicity” and “low toxicity” categories from which category benchmarks are derived. For example, the “high toxicity” benchmark could be defined as the highest LD₅₀ of the least toxic “high toxicity” reference chemical. For some end points, generic toxicity benchmarks are available, such as the European Union and Global Harmonized System categories for acute toxicity.



Decision-making: Decisions as to how to place chemicals into categories are based on the confidence bounds for each prediction:

- A chemical is categorized as “high confidence of high toxicity” if the upper confidence bound on the predicted LD₅₀ is less than or equal to the “high toxicity” benchmark.
- A chemical is categorized as “high confidence of low toxicity” if the lower confidence bound on the predicted LD₅₀ is greater than or equal to the “low toxicity” benchmark.

The remaining chemicals are categorized as having “inadequate data.”

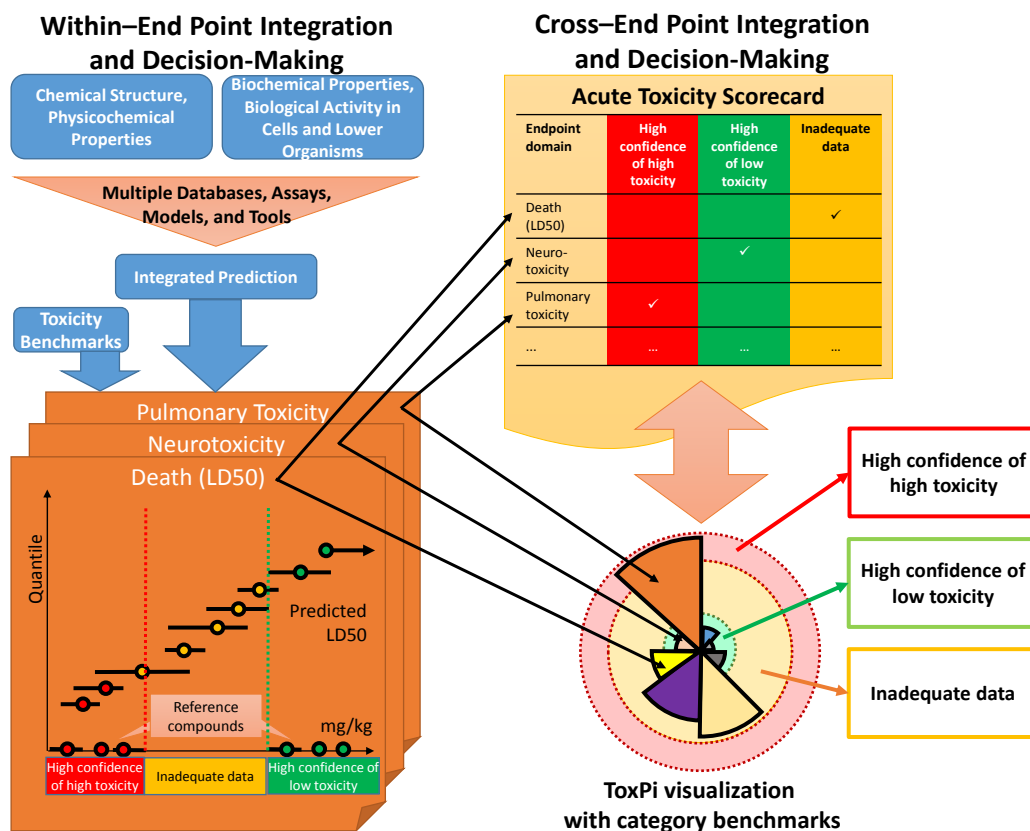


FIGURE 5-1 Illustration of a general approach to integration and decision-making for applying predictive approaches to acute, debilitating toxicity. “Within-end point” integration and decision-making has three basic steps: (1) developing an integrated prediction from potentially multiple databases, assays, models, and tools; (2) specifying toxicity benchmarks; and (3) placing chemicals into the appropriate category for the end point under consideration. “Cross-end point” integration and decision-making could consist of (1) a simple “scorecard” for a chemical in which each individual end point-specific decision is recorded or (2) more integrative approaches, such as ToxPi, that include the underlying toxicity end point predictions from which categories were assigned. The example presented here illustrates how information must be able to translate between within-end point and cross-end point integration.

A key task for DOD will be defining the most informative “end point domains” for its application, for example, whether to define an end point at a more general level, such as neurotoxicity, or at a more specific level, such as seizure or cholinesterase inhibition.

There are three steps in reaching a decision about a particular toxicity end point (the last two are the same as in the simple case described previously):

- (1) Integrating potentially multiple databases, assays, models, and tools into an “integrated prediction,” with its confidence interval, as to a chemical’s toxicity potential for that end point.
- (2) Specifying toxicity benchmarks that define the thresholds for what is considered “high” or “low” toxicity for that end point.
- (3) Placing chemicals into categories on the basis of the specified toxicity benchmarks, taking into account the confidence interval of the integrated prediction.

Some of the available methods for each step are described in greater detail below.

Cross-End Point Integration and Decision-Making

As described in Chapter 2, the concern about potential acute toxicity spans a wide array of toxic end points and biological mechanisms. From a military perspective, any acute toxicity that is severe enough to cause debilitation or death is enough to warrant concern. Thus, a simple approach to cross-end point integration would be simply to summarize the categorization results for each end point in a single scorecard, as shown in the upper right of Figure 5-1. Each end point would be evaluated as described above as to whether for a given type of acute toxicity (such as neurotoxicity) the chemical exhibited “high toxicity,” “low toxicity,” or “inadequate data.” Then, in integrating into an overall evaluation, a chemical will be sorted into the “high toxicity” bin if at least one of the end points is “high toxicity,” the “low toxicity” bin if *all the end points* are “low toxicity,” and the “inadequate data” bin if neither of the first two conditions applies. That approach also has the advantage of retaining the end point-specific information, so that future data generation can be targeted better. The approach is also consistent with a low tolerance for false negatives in that each end point identified as predictive of an acute toxic (debilitating or lethal) response serves as sufficient evidence to place a chemical into a “high toxicity” bin.

Cross-end point integration can also be visualized in a recombination approach (such as ToxPi, discussed in more detail below) and perhaps even augmented with information on toxicity benchmarks as illustrated in Figure 5-1. The recombination approach, however, suggests an alternative integration that would not depend strictly on a simple decision rule related to the categories for each end point. For example, the ToxPi approach (see lower right of Figure 5-1) could also provide a summary measure that consists of a weighted sum of individual toxicity end points. Thus, even if each individual end point is rated as “inadequate,” it is conceivable that the presence of multiple end points close to their benchmark thresholds would allow the chemical to be categorized as “high” or “low” on the basis of the summary measure even if no individual end point were so rated. Setting up appropriate decision rules would be a key policy question for DOD if it chose to implement the committee’s suggested approach for predicting acute, debilitating toxicity.

APPROACHES TO INTEGRATION

The general approaches for integrating databases, assays, models, and tools are illustrated in Figure 5-2 with LD₅₀ as an example end point. Broadly, they can be divided into approaches that combine individual predictions into an integrated prediction (upper panel, A) and approaches that first combine the underlying data from databases and assays before building an integrated model or tool (lower panel, B). Specific approaches are described in more detail below, particularly in relation to predicting acute toxicity; examples of applications to nontesting approaches (Chapter 3), biological assay-based approaches (Chapter 4), and combinations of the two are provided if possible.

Meta-Analytic Approaches

From a formal statistical perspective, the rich literature on meta-analysis offers guidance on how to aggregate information from multiple independent sources (Borenstein et al. 2009). It is expected that appropriate meta-analysis will help to improve the results from individual studies that might have been underpowered or have suffered from noise, bias, and absence of data. Meta-analysis can also help to reveal interesting patterns or relationships among studies and generate results that are statistically robust. The committee did not locate any examples of published meta-analyses of acute-toxicity predictions built from the types of data described in Chapters 3 and 4. However, for such an end point as LD₅₀, for which multiple tools and models are available in the same chemical domain (for example, the five models reviewed by Diaza et al. 2015), a meta-

analytic approach might be considered to combine results. In addition to combining individual predictions, meta-analytic approaches could be applied to individual categorization decisions (for example, see Box 5-2).

Three key issues should be considered in conducting any meta-analyses, including those applied to acute-toxicity end points. First, individual results should be investigated to determine whether the data can be combined reliably (Crowther et al. 2010). For example, results that are to be combined should be associated with a common predicted acute-toxicity end point. Often, decisions on which statistical techniques to use for meta-analysis are not as important as decisions on which studies are to be combined because later analysis will not be able to correct for an inappropriate combination of studies. Second, an effective meta-analysis should ensure that “better” results receive more weight during information aggregation (Crowther et al. 2010). For example, it is reasonable for predictions that have less variance or that are based on a larger sample size to contribute more heavily to the overall summary statistic in a meta-analysis. Other factors, such as biases and methodological strengths and weaknesses of individual approaches, can also be included, either qualitatively or quantitatively. Third, several statistical methods can be used to combine results. They include methods that are based on results of significance testing, such as p values or z scores, and fixed and random effect models that use summary statistics, such as the mean and standard error derived from individual results (Hedges and Olkin 1985; Borenstein et al. 2009). For all three issues, sensitivity analysis can be performed to assess the effects of various choices on the results of a meta-analysis (Higgins and Green 2011).

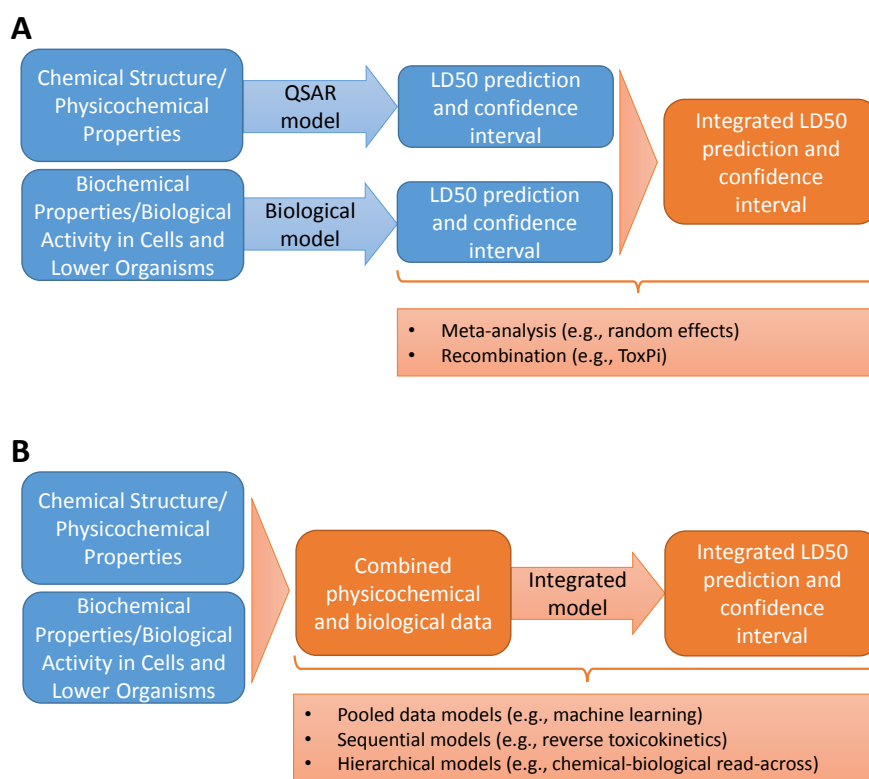


FIGURE 5-2 Approaches for integrating disparate datasets with LD₅₀ as an example. **A:** approaches that keep datasets separate (such as physicochemical data and assay-based biological-activity data) and integrate predictions from models developed for each dataset independently. **B:** approaches that combine datasets before modeling and develop a new “integrated” model that makes a prediction from the combined dataset.

Recombination-Based Approaches

Several approaches for data integration have recently been developed to handle new problems presented by high-dimensional toxicity data, such as information from different biological assays. Because the results from different data streams are not strictly comparable, formal meta-analytic approaches are not immediately applicable. There are a variety of methods for weighting multiple streams of evidence differently, from largely qualitative, expert-judgment approaches (for example, Hill criteria) to quantitative statistical frameworks that formalize weighting schemes (Linkov et al. 2015). Given the need to categorize chemicals, intermediate approaches that are quantitative and incorporate expert judgment are likely to be most useful in predicting acute toxicity. The Toxicological Prioritization Index (ToxPi) developed by Reif et al. (2010) provides a useful illustration of how to combine multiple data streams (discussed at length in NRC 2014).

BOX 5-2 Example of Meta-Analytic Approach That Uses Irreproducible Discovery Rates to Integrate Categorization Decisions

Meta-analytic approaches, such as the irreproducible discovery rate (IDR) (Li et al. 2011), can be used to measure the reproducibility of results from two independent studies and to filter noise in the results. In the example below, two studies (Study I and Study II) have made toxicity predictions, and the results are categorized into “low toxicity,” “inadequate data,” or “high toxicity,” on the basis of simulated category benchmarks with thresholds at values of [0–10], (10–60), and [60–100], respectively. However, many chemicals show inconsistent categorizations between the two studies. To integrate the two results, IDR considers the reproducibility between studies to assign a chemical-specific reproducibility measure. The far-right column presents the reproducibility measure as $1 - \text{local IDR}$, where the chemicals that have the highest values are highlighted in red. Note that taking reproducibility among studies into account can change the categorization of “high toxicity.”

	Chemicals	Initial Toxicity Prediction Scores		1 – local IDR
		Study I	Study II	
	1	99	100	0.99
	2	72	45	0.53
Category Benchmarks by initial	3	70	58	0.91
toxicity prediction scores:	4	66	0	0.00
- High toxicity:	5	65	38	0.19
[60–100]	6	62	49	0.59
- Inadequate data:	7	56	15	0.00
(10–60)	8	51	60	0.62
- Low toxicity:	9	51	72	0.79
[0–10]	10	47	59	0.45
	11	40	53	0.17
The ones categorized as “high	12	37	42	0.06
toxicity” by initial prediction scores	13	36	31	0.03
are highlighted in Blue	14	30	20	0.01
	15	30	20	0.01
The ones with top 6 (1-local IDR)	16	29	35	0.02
scores (or the 6 most reproducible	17	24	28	0.01
results) are highlighted in Red	18	15	28	0.00
	19	14	43	0.00
	20	8	41	0.00

The ToxPi combines data streams (physicochemical data, biological assays, or both) into a relative index to facilitate prioritization or categorization. At its most basic, a summary ranking is derived for each chemical on the basis of a weighted sum of rankings for different data sources. Although advanced statistical approaches could be used to group and weight individual pieces of evidence empirically, published applications have used substantial expert judgment and taken into account the specific sources of data being integrated and the context of integration. For example, binding assays for several cytochrome P450 (CYP) isoforms might be run to assess “xenobiotic metabolism,” but isoforms expressed in a target tissue of interest might warrant more weight. Thus, using the ToxPi approach and reference chemicals can provide an overall, integrated ranking that can be used for categorization decisions (see Figure 5-3). In addition, the ToxPi provides a visualization of the individual component ranks and so can be used to support a “multicriteria” decision-making scenario (Pavan and Worth 2008) in which categorization decisions involve integration of multiple, possibly conflicting criteria. For example, placing chemicals in the “high confidence of high toxicity” bin could necessitate synthesizing results when individual pieces of evidence serve as flags of high alert (such as solid evidence from a single assay that is deemed highly predictive of acute toxicity). The use of ToxPi for cross-end point integration is illustrated in the lower right of Figure 5-1, where axes are flagged if activity surpasses a benchmark threshold. ToxPi can also be applied to within-end point integration.

Pooled Data-Based Approaches

The basic idea behind pooled data-based approaches is the use of the same types of non-testing approaches (such as read-across and QSAR) to make chemical-based or biologically based predictions. Thus, the biological assay results are simply treated as additional “biological descriptors” that can be used with physicochemical descriptors in building quantitative models. Several examples of this approach applied to acute toxicity are described in Chapter 3 (Lee et al. 2010; Sedykh et al. 2011); they retrospectively apply statistical methods, such as k-nearest neighbor, random forest, or multiple linear regression to a combined dataset of chemical-structure data and literature-derived cytotoxicity data. The same approach could be used more prospectively, in which new biological activity data generated in Tier 2 are combined with chemical-structure information used previously in Tier 1 to develop an integrated prediction based on pooling of both datasets.

Hierarchical Modeling Approaches

Hierarchical approaches can build-in information on chemical structure or global performance to inform modeling decisions (Wilson et al. 2014). For example, as noted in Chapter 3, several groups have used biological data to stratify chemicals into clusters; more localized modeling, such as the use of QSARs, was then applied to each (Zhu et al. 2009; Zhang 2011; Lounkine et al. 2012). Zhu et al. (2009) applied such an approach to acute toxicity. Specifically, they grouped chemicals into those with and without good correlation between *in vitro* cytotoxicity IC_{50} s and *in vivo* rodent LD_{50} s. They then built one QSAR model to assign chemicals to each group and two additional QSAR models to predict LD_{50} s for each group. They also compared their approach with the commercial TOPKAT software, using a set of chemicals that was outside the training set of both approaches. They found that their two-step hierarchy of QSAR models had a greater correlation coefficient and smaller mean absolute error. That type of stratification of QSAR models with biological information might be a fruitful integration approach that can be tried with other acute-toxicity end points.

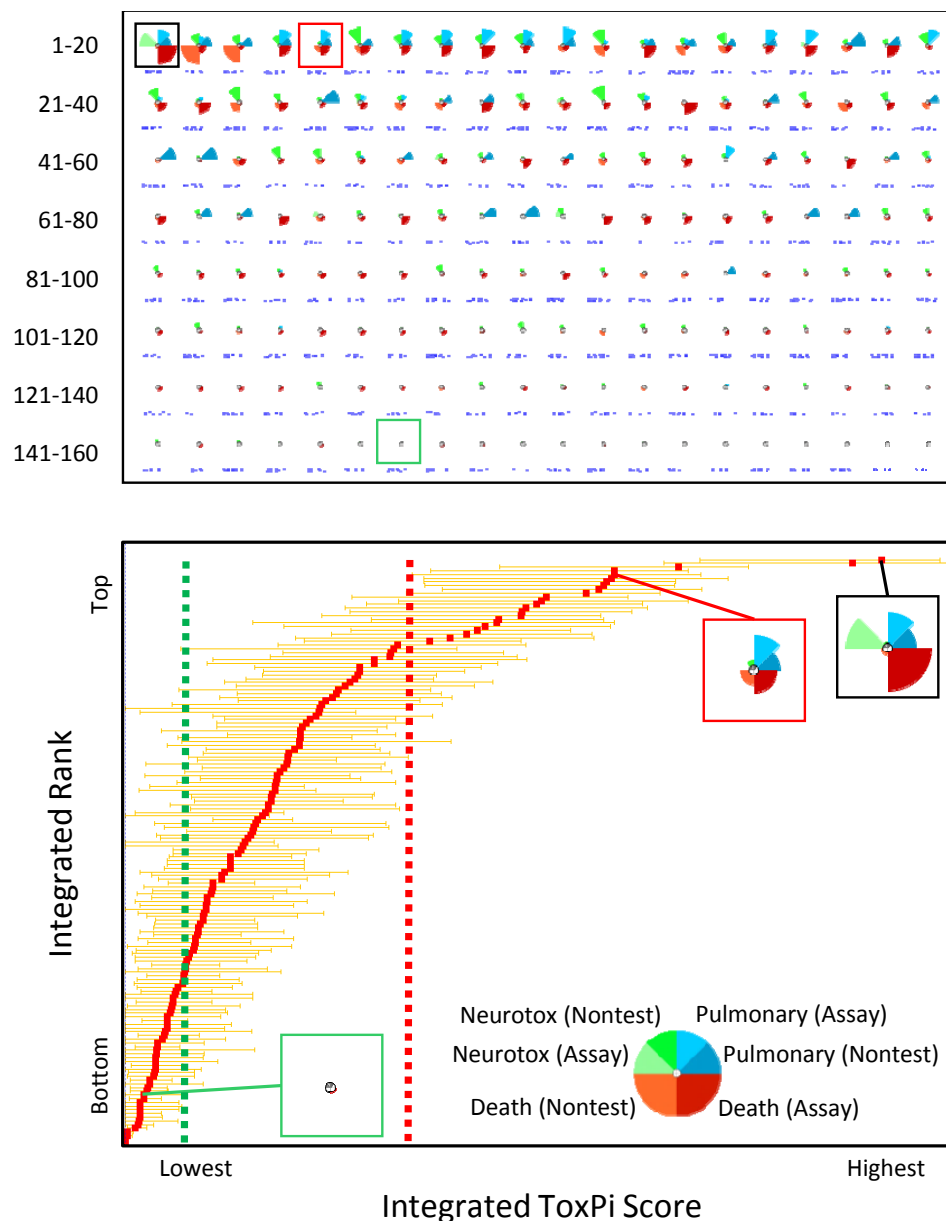


FIGURE 5-3 ToxPi model for integration of acute-toxicity potential. In this simulated example, data from nontesting approaches (Chapter 3) and assays (Chapter 4) have been integrated into a cross-end point ToxPi model for pulmonary toxicity, neurotoxicity, and death. The key (lower panel, inset) shows that evidence for the two “Death” slices have been given extra weight in determining the overall integrated ranking for acute toxicity. For each chemical profile, the distance of a slice from the origin indicates the relative potency. The “longer” slices indicate chemicals that are more potent than chemicals that have “shorter” slices or those deemed inactive (indicated by absence of a given slice). The upper panel shows all profiles in rank order. The lower panel translates the profiles into a plot of the integrated ranks vs scores, with 95% confidence intervals extending from the red square representing each chemical. The black-framed profile is the chemical that has the highest overall acute-toxicity potential (that is, rank = 1). The red-framed profile is the reference chemical for “high confidence of high toxicity,” and the green-framed profile is the reference chemical for “high confidence of low toxicity.” Note that the confidence intervals extending from each reference chemical define the category thresholds (vertical dashed red and green lines).

Sequential Modeling Approaches

Another approach to model-based integration is to connect models sequentially, that is, by using outputs from one model as part of the inputs into another model. Such an approach can be developed when there is a hypothesized chain of events that leads to acute toxicity. Specifically, as described in Chapter 3, nontesting approaches could be developed for initial or intermediate events along a mechanistic pathway. The predictions could then be inputs into models that predict acute toxicity on the basis of biological assay data on the intermediate events. As a result, predictions of acute toxicity that incorporate biological data could be made for chemicals for which biological assays have not yet been conducted.

For example, Chapter 3 discussed how chemical toxicity often results from nonspecific alterations in cell function; thus, *in vitro* cytotoxicity is likely to be a strong indicator of *in vivo* toxicity. Conceivably, a sequential model could be developed that combines a model that uses physicochemical data to predict cytotoxicity (reviewed in Chapter 3) with a model that uses cytotoxicity data to predict LD_{50} s (reviewed in Chapter 4). In the future, one might envision using physicochemical information to predict bioactivity in more specific biological assays and then using existing models that use bioactivity measurements to predict *in vivo* acute toxicity.

Another area in which sequential modeling is common is *in vitro*–*in vivo* extrapolation (IVIVE). In particular, the results of *in vitro* assays constitute inputs into a reverse-dosimetry or reverse-toxicokinetic approach to derive the external dose that will result in the internal serum concentration equivalent to the bioactive concentration in the *in vitro* assay (Rotroff et al. 2010; Wetmore et al. 2012). Integration of ADME specifically is discussed further below.

Sequential modeling is not restricted to single assay results. The US Environmental Protection Agency (EPA) ToxCast program has generated *in vitro* high-throughput screening data on several assay technologies that assess multiple pathways, genes, and responses over hundreds of end points (Judson et al. 2010; Kavlock et al. 2012). The built-in redundancy in end points allows assays to be aggregated into a pathway context, so that multiple assay results are combined into a summary measure of pathway activity before a reverse toxicokinetic model is applied (Judson et al. 2011). Other researchers have attempted to integrate assays anchored to pathways to arrive at a summary outcome, specifically for skin sensitization (see, for example, Jaworska et al. 2013; Patlewicz et al. 2014; van der Veen et al. 2014).

Integrating Toxicokinetics to Determine Acute-Toxicity Potential

The recent investment in *in vitro* and high-throughput screening (HTS) strategies to inform chemical toxicity testing has led to increased debate about relating the resulting data (potency values derived on the basis of the nominal testing concentration ranges used in the wells of assay plates) to values that would be informative in predicting *in vivo* human health hazard. Some assay designs do not consider *in vivo* toxicokinetic processes, which ultimately dictate the extent of chemical toxicity or potency in animals or humans. No matter how active or potent a chemical might be in some *in vitro* assays, if it is not absorbed into the human body on exposure it will not be bioavailable to elicit any effect. Similarly, a chemical that is cleared rapidly might not be present in the body long enough to elicit an effect. Chapters 3 and 4 describe how modeling of biological (toxicokinetic) processes can reduce the gaps observed between results of *in vitro* assay and toxic response in humans or animals. Consideration of the potential effect of *in vivo* toxicokinetic processes improves interpretation of the results of *in vitro* testing and the overall predictive-toxicology approach.

Development of toxicokinetic and dosimetric tools to inform extrapolations—among species, dose ranges, and experimental systems (for example, *in vitro* to *in vivo*)—has been widespread over the last 30 years. Forward dosimetry with physiologically based pharmacokinetic modeling originally introduced in the assessment of volatile organic chemicals (Andersen et al. 1987) provides a strategy that relies on pharmacokinetic knowledge to relate a known external

exposure to an internal blood or target-tissue dose. Alternatively, reverse dosimetry is often used to relate a known internal dose (either from blood biomonitoring data or from an *in vitro* assay bioactivity concentration) to an external chemical dose (Tan et al. 2007; Lyons et al. 2008; Wetmore et al. 2012). The focus of much of the reverse-dosimetry work has been on exposure scenarios (chronic, low-level, repeat exposures) of concern to EPA or other regulatory bodies that have similar public-health protection mandates (Basketter et al. 2012; Wetmore 2015). Whereas it is useful to understand the dosimetric strategies described, it should be noted that not all tools are directly applicable for DOD's purpose in which the exposure will be acute (probably at one time) and that acutely toxic chemicals might not require consideration of ADME to predict elicitation of debilitating effects. The following discussion provides a brief summary of the different components of ADME, tools available to predict the components, and probable effect of ADME on future attempts by DOD to predict acute toxicity.

Absorption or Bioavailability. Absorption via relevant routes of exposure (oral, dermal, and inhalation) and resulting bioavailability are important parameters for which there are predictive testing and nontesting tools (see Chapters 3 and 4). A conservative assumption of 100% absorption would be the most protective and should be assumed for chemicals for which available methods do not apply or do not provide sufficient predictivity. Computer models that describe the biochemical behavior of uptake by skin or gastrointestinal cells (McKone 1990; Jamei et al. 2009a; Rauma et al. 2013) can be applied in a sequential approach. However, the uptake models have been developed mostly for single-chemical exposure and rarely for acute conditions. If 100% absorption was initially assumed for a chemical and predictive tools indicate an adjustment away from that conservative default, the chemical might need to be downgraded or reassessed because lower absorption would indicate a lower potency or lower likelihood of acute toxicity. Chemicals that have other toxicity alerts that place them in the "high confidence of high toxicity" bin and identify them as readily absorbed should be noted because confirmation of high bioavailability will influence their ranking in that bin.

Distribution. Rapid accumulation throughout the body or in target tissues (such as skin, brain, and lungs) could lead to a substantial shift in acute-toxicity potential of a chemical. Some organophosphates, for example, are highly lipophilic and are readily distributed into fat and other tissues. Presence in the fat and slow release from the site will delay or prolong acetylcholinesterase inhibition (Karalliedde et al. 2003) and could lead to greater toxicity than that of chemicals that have lower distribution but a similar effect in an *in vitro* assay. Similarly, a chemical that is demonstrably neurotoxic in an *in vitro* assay and is identified as crossing the blood-brain barrier rapidly will likely be a more potent neurotoxicant than one that has similar *in vitro* toxicity and does not cross the blood-brain barrier. Blood transporters or lipoproteins can also shift the ability of a chemical to reach a toxicity target. Chemicals that bind to transporter proteins might reach target sites more easily than chemicals that are distributed solely by diffusion. Tools for predicting or assessing tissue partitioning and partition coefficients are available and can aid in predicting ADME behavior (Poulin and Haddad 2012, 2013). Distribution and metabolism (see below) are probably the two most important components to explore in modulating acute-toxicity potential. Distribution will indicate the amount of a chemical that is available to reach an *in vivo* targeted site. Modeling tools and simple *in vitro* experiments can estimate tissue or cell partitioning and thus provide a more accurate estimate of assay concentration that will generate a response.

Metabolism. Metabolism can lead to formation of a substantially more toxic metabolite, formation of an equally toxic metabolite, or detoxification of the parent chemical. Examples of chemicals that can be bioactivated into more toxic metabolites are the neurotoxic organophosphate insecticides, such as fenthion, parathion, diazinon, malathion, and chlorpyrifos, which are metabolized to potent oxon metabolites (Eisler 2007). Thus, metabolism is one ADME property that could substantially shift the potency of a chemical. Nontesting approaches for predicting metabolism are in various stages of development (see Chapter 3) and might be used to elucidate metabolites that could contribute to a parent chemical's acute-toxicity potential. The best way to

incorporate the potential effect of toxic metabolite formation will depend largely on the chemical mechanism in relation to others that have similar acute-toxicity potencies and will need to be considered in relation to other toxicity information.

Excretion. The liver and kidney are the major organs responsible for excretion of chemicals from the body (via bile and urine). Physicochemical properties can be used to predict chemical excretion rates (Ghibellini et al. 2006; Shitara et al. 2006; Sharifi and Ghafourian 2014). Hydrophilic chemicals are more rapidly excreted from the body than lipophilic chemicals. The other ADME properties are more likely than excretion to have an important effect on chemical acute-toxicity potential. Reductions in excretion are likely to be secondary to renal toxicity, an end point that is likely to be specifically assessed with other testing and nontesting approaches.

In Vitro to In Vivo Extrapolation Modeling to Inform Tissue Dosimetry and Dosimetric Potential

Modeling approaches developed to inform dosimetry assessments use IVIVE. Measurements from in vitro assays and predictions from nontesting approaches can provide various model inputs (such as rate of absorption, metabolic activity, and tissue partitioning), which can be combined in a bottom-up approach to estimate an in vivo dose (Jamei et al. 2009b). The extrapolation of in vitro data typically assumes that metabolism of a parent chemical implies loss of potential for toxicity, which is not necessarily the case, but it is valuable in providing an understanding of the dosimetrics or bioaccumulative potential of a chemical.

Recently, a simplified IVIVE approach amenable to incorporation with HTS data was presented; it predicted external doses that are required to achieve steady-state blood concentrations similar to ones that elicit activity in in vitro HTS assays (Rotroff et al. 2010; Wetmore et al. 2012). The approach incorporated plasma-protein binding and hepatic metabolic and renal non-metabolic clearance (key determinants of chemical steady-state toxicokinetics) to estimate chemical steady-state behavior. The toxicokinetic measurements showed not only significant cross-species correlation but strong correlation between in vivo and in vitro values for several chemicals (Wetmore et al. 2013, Wetmore 2015). Steady-state concentrations are known to be more relevant for chronic and repeated exposure. Recent assessments by EPA demonstrated that in some cases steady-state concentration (C_{ss}) estimates were consistent with peak concentrations (C_{max}); this relationship was not observed for a subset of chemicals that are rapidly or slowly cleared (Wambaugh et al. in press). A steady-state toxicokinetic assumption, however, can be valid for acute-toxicity assessment.

Many of the tools developed to predict ADME or pharmacokinetic properties were developed by using pharmaceutical chemicals, which represent a smaller chemical space than the chemical domain of concern to DOD. Tools to predict absorption, tissue partitioning, protein binding, and hepatic clearance might perform well only in a narrowly defined $\log K_{ow}$ space. Chemicals of concern to DOD will span multiple domains, and attempts to categorize their pharmacokinetic behavior might be severely limited.

Ultimately, one should recognize that the most reactive acutely toxic chemicals that are of top concern to DOD are likely to be rapidly lethal well before some metabolic or excretory processes are initiated. However, for a large percentage of the chemical space of concern to DOD, application of some of the physicochemical and assay tools to predict chemical toxicokinetics can probably be incorporated into the decision-making process reasonably efficiently. Overall, integration of the tools in a tiered testing framework will aid in refining estimates of the toxicity of chemicals and enable an appropriate and streamlined decision-making process.

APPROACHES TO CATEGORIZATION

As discussed in Chapter 2, the committee's suggested prioritization strategy consists of a tiered approach to placing chemicals into three categories: "high confidence of high toxicity,"

“high confidence of low toxicity,” and “inadequate data.” As described earlier in this chapter, the process includes two key steps: setting benchmarks that define the thresholds for what would be considered high and low toxicity and assigning chemicals to categories, taking into account the confidence in the prediction.

Setting Benchmarks

The thresholds for high and low toxicity for a given toxicity end point can be defined in multiple ways, including the following:

- Using reference chemicals of high and low toxicity when the toxicity end point of interest has been measured or predicted. This approach was illustrated in the simplified example in Box 5-1 and in Figure 5-1 with LD₅₀ as the toxicity end point.
- Using reference chemicals of high and low toxicity via clustering approaches. For example, the “nearest neighbors” to the high toxicity reference chemicals would also be considered to have high toxicity.
- Using pre-existing exposure-based thresholds if they exist for the toxicity end point of interest. For example, the European Union (EU) and the Globally Harmonized System of Classification and Labeling of Chemicals (GHS) each have pre-existing categories for acute oral toxicity in terms of milligrams per kilogram.

All the approaches require that the toxicity end point be a numerical value, such as an LD₅₀ in milligrams per kilogram, and they do not work for qualitative or binary outputs, such as “active” and “inactive.”

Categorization Decisions under Uncertainty

With respect to assigning chemicals to categories and taking into account confidence, by definition, the assignments to the high and low toxicity categories need to have high confidence. It is therefore of paramount importance for the categorization process to characterize uncertainty in the prediction of the toxicity end point.

Uncertainty is an unavoidable aspect of knowledge discovery from large-scale heterogeneous data, such as those discussed in Chapters 3 and 4. It has two main sources: the data themselves and the modeling of those data. With respect to data, even physicochemical properties cannot be measured with perfect precision. Moreover, such data are usually obtained from databases, which can contain errors that are due to data entry or other processes related to creating and populating the databases (Fourches et al. 2010). For biological assays, uncertainties will vary considerably with the type of assay. Although automation of some assays has greatly improved their reproducibility, there is still variation among batches from a given assay and possibly even among chemicals in a given batch. Uncertainty arising from the modeling or analysis approach (different aspects of data handling and data analysis) varies with the model and among parameter values in a given model.

There is a rich literature on characterizing uncertainty in experimental and computational sciences that need not be recapitulated here. Suffice it to say that the existing approaches span a wide range that is based on both “frequentist” and “Bayesian” statistical principles and use analytical methods (for example, classical confidence intervals) and sampling or simulation-based methods (such as bootstrap and Monte Carlo). In some cases, specific guidance is available on characterizing uncertainty for a particular application, such as QSAR modeling, gene-expression analyses, and pharmacokinetic modeling (IPCS 2008).

For example, Chapter 3 discussed the availability of Organisation for Economic Co-operation and Development (OECD) and Registration, Evaluation, Authorization and Restriction of Chemi-

cals (REACH) guidance on evaluating the confidence in (Q)SAR models. Although internal and external cross-validation make up the current standard approach to characterizing uncertainty in QSAR predictions, alternative approaches are being pursued. For example, Gramacy and Pantaleo (2010) used a Markov Chain Monte Carlo sampling for the Bayesian Lasso model to assess prediction uncertainties in QSAR analyses. More discussion of other aspects of uncertainty in QSAR predictions can be found in Sahlin (2013). Parametric methods can also be used sometimes, but these might be less appealing for complex data from multiple sources because distributional assumptions are likely to be violated.

An additional useful tool for assessing confidence is sensitivity analysis, which can tell how the uncertainty in the toxicity prediction of a model or system can be apportioned to different sources of uncertainty in its inputs or, equivalently, how much each input is contributing to the uncertainty. For example, sensitivity analyses might be performed for several purposes:

- To test the robustness of results by random data perturbation.
- To decompose the prediction error by relating input and output variables in a model, which can help to understand the sources of variation in the model. Such analyses can identify inputs that are key contributors to uncertainty and thereby focus research or data generation on aspects that could improve or refine the model.
- To perform model selection through internal cross-validation or external validation. Internal cross-validation can also be used to quantify the degree of fitting or overfitting of a model. External validation (meta-analysis and related data streams) is preferred as an independent assessment of a model's predictive ability.

Finally, it should be noted that these approaches to assessment of uncertainty and sensitivity might be applied not only to toxicity end point predictions but to the categorization decisions themselves. Application to the categorization decision is illustrated in Box 5-2, where uncertainty is analyzed by using the irreproducible discovery rate.

FINDINGS

- **Finding:** The committee's recommended prioritization strategy will require integration at various stages, and there are many approaches to integration, including formal statistical methods (such as meta-analysis), less formal "recombination" methods (such as ToxPi), methods that pool datasets, methods that use datasets hierarchically, and methods that link models sequentially.
- **Finding:** There are several possible approaches to placing chemicals in categories of "high toxicity," "low toxicity," and "inadequate data," including quantitative thresholds based on reference chemicals (such as sarin), generic thresholds based on external criteria (such as those of the EU and the GHS), and clustering based on reference chemicals. The stability of and confidence in the results of any categorization will be enhanced if the uncertainties are properly quantified.
- **Finding:** Multiple levels of complexity require integration, and the committee has distinguished between integration of predictions among end points (cross-end point integration) and integration of different model predictions or data streams that inform a single end point (within-end point integration).
- **Finding:** Given the many types of end points that are relevant to acute, debilitating toxicity and the need to place chemicals into categories of "high toxicity," "low toxicity," and "inadequate data," the simplest approach to cross-end point integration would be to summarize the categorization results for each end point in a scorecard.

- **Finding:** Some of the key policy decisions that DOD will need to make to use the committee's recommended prioritization strategy are (1) the kinds of responses that would be considered appropriate end points (for example, neurotoxicity, seizures, or cholinesterase inhibition), (2) determination of high and low toxicity thresholds for each end point of interest, (3) the degree of confidence required to conclude high confidence, and (4) decision rules related to a determination of a summary conclusion of high or low toxicity on the basis of multiple end points.

- **Finding:** Toxicokinetic and ADME behavior can influence the prediction of a chemical's acute toxicity potential and resulting categorization, and this emphasizes the need to include such considerations into the tiered prioritization strategy.

REFERENCES

- Andersen, M.E., H.J. Clewell III, M.L. Gargas, F.A. Smith, and R.H. Reitz. 1987. Physiologically based pharmacokinetics and the risk assessment process for methylene chloride. *Toxicol. Appl. Pharmacol.* 87(2):185-205.
- Basketter, D.A., H. Clewell, I. Kimber, A. Rossi, B. Blaauboer, R. Burrier, M. Daneshian, C. Eskes, A. Goldberg, N. Hasiwa, S. Hoffmann, J. Jaworska, T.B. Knudsen, R. Landsiedel, M. Leist, P. Locke, G. Maxwell, J. McKim, E.A. McVey, G. Ouédraogo, G. Patlewicz, O. Pelkonen, E. Roggen, C. Rovida, I. Ruhdel, M. Schwarz, A. Schepky, G. Schoeters, N. Skinner, K. Trentz, M. Turner, P. Vanparys, J. Yager, J. Zurlo, and T. Hartung. 2012. A roadmap for the development of alternative (non-animal) methods for systemic toxicity testing - t4 report. *ALTEX* 29(1):3-91.
- Bauer-Mehren, A., E.M. van Mullingen, P. Avillach, M.D.C. Carrascosa, R. Garcia-Serna, J. Piñero, B. Singh, P. Lopes, J.L. Oliveira, G. Diallo, E.A. Helgee, S. Boyer, J. Mestres, F. Sanz, J.A. Kors, and L.I. Furlong. 2012. Automatic filtering and substantiation of drug safety signals. *PLoS Comput. Biol.* 8(4):e1002457.
- Borenstein, M., L.V. Hedges, J.P.T. Higgins, and H.R. Rothstein. 2009. *Introduction to Meta-Analysis*. Chichester, U.K: Wiley.
- Crowther, M., W. Lim, and M.A. Crowther. 2010. Systematic review and meta-analysis methodology. *Blood* 116(17):3140-3146.
- Díaz, R.G., S. Manganelli, A. Esposito, A. Roncaglioni, A. Manganaro, and E. Benfenati. 2015. Comparison of in silico tools for evaluating rat oral acute toxicity. *SAR QSAR Environ Res.* 26(1):1-27.
- Eisler, R. 2007. *Eisler's Encyclopedia of Environmentally Hazardous Priority Chemicals*. Amsterdam: Elsevier.
- Fourches, D., E. Muratov, and A. Tropsha. 2010. Trust, but verify: On the importance of chemical structure curation in cheminformatics and QSAR modeling research. *J. Chem. Inf. Model.* 50(7):1189-1204.
- Ghibellini, G., E.M. Leslie, and K.L. Brouwer. 2006. Methods to evaluate biliary excretion of drugs in humans: An updated review. *Mol. Pharm.* 3(3):198-211.
- Gramacy, R.B., E. Pantaleo. 2010. Shrinkage regression for multivariate inference with missing data, and an application to portfolio balancing. *Bayesian Anal.* 5(2):237-262.
- Hedges, L.V., and I. Olkin. 1985. *Statistical Methods for Meta-Analysis*. New York: Academic Press.
- Higgins, J.P.T., and S. Green, eds. 2011. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]*. The Cochrane Collaboration [online]. Available: <http://handbook.cochrane.org/> [accessed March 24, 2015].
- IPCS (International Programme on Chemical Safety). 2008. *Uncertainty and Data Quality in Exposure Assessment*. World Health Organization [online]. Available: http://www.who.int/ipcs/publications/methods/harmonization/exposure_assessment.pdf?ua=1 [accessed March 24, 2015].
- Jamei, M., D. Turner, J. Yang, S. Neuhoff, S. Polak, A. Rostami-Hodjegan, and G. Tucker. 2009a. Population-based mechanistic prediction of oral drug absorption. *AAPS J.* 11(2):225-237.
- Jamei, M., S. Marciniak, K. Feng, A. Barnett, G. Tucker, and A. Rostami-Hodjegan. 2009b. The Simcyp® population-based ADME simulator. *Exp. Opin. Drug Metab. Toxicol.* 5(2):211-223.
- Jaworska, J., Y. Dancik, P. Kern, F. Gerberick, and A. Natsch. 2013. Bayesian integrated testing strategy to assess skin sensitization potency: From theory to practice. *J. Appl. Toxicol.* 33(11):1353-1364.
- Judson, R.S., K.A. Houck, R.J. Kavlock, T.B. Knudsen, M.T. Martin, H.M. Mortensen, D.M. Reif, D.M. Rotroff, I. Shah, A.M. Richard, and D.J. Dix. 2010. In vitro screening of environmental chemicals for targeted testing prioritization: The ToxCast project. *Environ. Health Perspect.* 118(4):485-492.

- Judson, R.S., R.J. Kavlock, R.W. Setzer, E.A. Cohen Hubal, M.T. Martin, T.B. Knudsen, K.A. Houck, R.S. Thomas, B.A. Wetmore, and D.J. Dix. 2011. Estimating toxicity-related biological pathway altering doses for high-throughput chemical risk assessment. *Chem. Res. Toxicol.* 24(4):451-462.
- Karalliedde, L.D., P. Edwards, and T.C. Marrs. 2003. Variables influencing the toxic response to organophosphates in humans. *Food Chem. Toxicol.* 41(1):1-13.
- Kavlock, R., K. Chandler, K. Houck, S. Hunter, R. Judson, N. Kleinstreuer, T. Knudsen, M. Martin, S. Padilla, D. Reif, A. Richard, D. Rotroff, N. Sipes, and D. Dix. 2012. Update on EPA's ToxCast program: Providing high throughput decision support tools for chemical risk management. *Chem. Res. Toxicol.* 25(7):1287-1302.
- Lee, S., K. Park, H.S. Ahn, and D. Kim. 2010. Importance of structural information in predicting human acute toxicity from in vitro cytotoxicity data. *Toxicol. Appl. Pharmacol.* 246(1-2):38-48.
- Li, Q., J.B. Brown, H. Huang, and P.J. Bickel. 2011. Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* 5(3):1752-1779.
- Linkov, I., O. Massey, J. Keisler, I. Rusyn, and T. Hartung. 2015. From "weight of evidence" to quantitative data integration using multicriteria decision analysis and Bayesian methods. *ALTEX* 321(1):3-8.
- Lounkine, E., M.J. Keiser, S. Whitebread, D. Mikhailov, J. Hamon, J.L. Jenkins, P. Lavan, E. Weber, A.K. Doak, S. Côté, B.K. Shoichet, and L. Urban. 2012. Large-scale prediction and testing of drug activity on side-effect targets. *Nature* 486(7403):361-367.
- Lyons, M., R.S. Yang, A.N. Mayeno, and B. Reisfeld. 2008. Computational toxicology of chloroform: Reverse dosimetry using Bayesian inference, Markov Chain Monte Carlo simulation, and human biomonitoring data. *Environ. Health Perspect.* 116(8):1040-1046.
- McKone, T.E. 1990. Dermal uptake of organic chemicals from a soil matrix. *Risk Anal.* 10(3):407-419.
- NRC (National Research Council). 2014. *A Framework to Guide Selection of Chemical Alternatives*. Washington, DC: National Academies Press.
- Patlewicz, G., C. Kuseva, A. Kesova, I. Popova, T. Zhechev, T. Pavlov, D.W. Roberts, and O. Mekenyan. 2014. Towards AOP application--implementation of an integrated approach to testing and assessment (IATA) into a pipeline tool for skin sensitization. *Regul. Toxicol. Pharmacol.* 69(3):529-545.
- Pavan, M., and A. Worth. 2008. A Set of Case Studies to Illustrate the Applicability of DART (Decision Analysis by Ranking Techniques) in the Ranking of Chemicals. EUR 23481. Luxembourg: European Communities [online]. Available: https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/doc/EUR_23481_EN.pdf.
- Poulin, P., and S. Haddad. 2012. Advancing prediction of tissue distribution and volume of distribution of highly lipophilic compounds from a simplified tissue-composition-based model as a mechanistic animal alternative method. *J. Pharm. Sci.* 101(6):2250-2261.
- Poulin, P., and S. Haddad. 2013. Hepatocyte composition-based model as a mechanistic tool for predicting the cell suspension: Aqueous phase partition coefficient of drugs in in vitro metabolic studies. *J. Pharm. Sci.* 102(8):2806-2818.
- Rauma, M., A. Boman, and G. Johanson. 2013. Predicting the absorption of chemical vapours. *Adv. Drug Deliv. Rev.* 65(2):306-314.
- Reif, D.M., M.T. Martin, S.W. Tan, K.A. Houck, R.S. Judson, A.M. Richard, T.B. Knudsen, D.J. Dix, and R.J. Kavlock. 2010. Endocrine profiling and prioritization of environmental chemicals using ToxCast data. *Environ. Health Perspect.* 118(12):1714-1720.
- Rotroff, D.M., B.A. Wetmore, D.J. Dix, S.S. Ferguson, H.J. Clewell, K.A. Houck, E.L. Lecluyse, M.E. Andersen, R.S. Judson, C.M. Smith, M.A. Sochaski, R.J. Kavlock, F. Boellmann, M.T. Martin, D.M. Reif, J.F. Wambaugh, and R.S. Thomas. 2010. Incorporating human dosimetry and exposure into high-throughput in vitro toxicity screening. *Toxicol. Sci.* 117(2):348-358.
- Sahlin, U. 2013. Uncertainty in QSAR predictions. *ATLA* 41(1):111-125.
- Sedykh, A., H. Zhu, H. Tang, L. Zhang, A. Richard, I. Rusyn, and A. Tropsha. 2011. Use of in vitro HTS-derived concentration-response data as biological descriptors improves the accuracy of QSAR models of in vivo toxicity. *Environ. Health Perspect.* 119 (3):364-370.
- Sharifi, M., and T. Ghafourian. 2014. Estimation of biliary excretion of foreign compounds using properties of molecular structure. *AAPS J.* 16(1):65-78.
- Shitara, Y., T. Horie, and Y. Sugiyama. 2006. Transporters as a determinant of drug clearance and tissue distribution. *Eur. J. Pharm. Sci.* 27(5):425-446.
- Tan, Y.M., K.H. Liao, and H.J. Clewell III. 2007. Reverse dosimetry: Interpreting trihalomethanes biomonitoring data using physiologically based pharmacokinetic modeling. *J. Expo. Sci. Environ. Epidemiol.* 17(7):591-603.

- van der Veen, J.W., E. Rorije, R. Emter, A. Natsch, H. van Loveren, and J. Ezendam. 2014. Evaluating the performance of integrated approaches for hazard identification of skin sensitizing chemicals. *Regul. Toxicol. Pharmacol.* 69(3):371-379.
- Wambaugh, J.F., B.A. Wetmore, R. Pearce, C. Strope, R. Goldsmith, J.P. Sluka, A. Sedykh, A. Tropsha, S. Bosgra, I. Shah, R. Judson, R.S. Thomas, and R.W. Setzer. In press. Toxicokinetic triage for environmental chemicals. *Toxicol. Sci.*
- Wetmore, B.A. 2015. Quantitative in vitro-to-in vivo extrapolation in a high-throughput environment. *Toxicology* 332:94-101.
- Wetmore, B.A., J.F. Wambaugh, S.S. Ferguson, M.A. Sochaski, D.M. Rotroff, K. Freeman, H.J. Clewell III, D.J. Dix, M.E. Andersen, K.A. Houck, B. Allen, R.S. Judson, R. Singh, R.J. Kavlock, A.M. Richard, and R.S. Thomas. 2012. Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. *Toxicol. Sci.* 125(1):157-174.
- Wetmore, B.A., J.F. Wambaugh, S.S. Ferguson, L. Li, H.J. Clewell, III, R.S. Judson, K. Freeman, W. Bao, M.A. Sochaski, T.M. Chu, M.B. Black, E. Healy, B. Allen, M.E. Andersen, R.D. Wolfinger, and R.S. Thomas. 2013. The relative impact of incorporating pharmacokinetics in predicting in vivo hazard and mode-of-action from high-throughput in vitro toxicity assays. *Toxicol. Sci.* 132(2):327-346.
- Wilson, A., D.M. Reif, and B.J. Reich. 2014. Hierarchical dose-response modeling for high-throughput toxicity screening of environmental chemicals. *Biometrics* 70(1):237-246.
- Zhang, S. 2011. Computer-aided drug discovery and development. *Methods Mol. Biol.* 716:23-38.
- Zhu, H. L. Ye, A. Richard, A. Golbraikh, F.A. Wright, I. Rusyn, and A. Tropsha. 2009. A novel two-step hierarchical quantitative structure-activity relationship modeling work flow for predicting acute toxicity of chemicals in rodents. *Environ. Health Perspect.* 117(8):1257-1264.

6

Lessons Learned and Next Steps

The objective of the committee's conceptual framework as presented in Chapter 2 is to predict the potential of a chemical to cause acute toxicity to organ systems that could result in debilitating or lethal effects.¹ In developing its framework, the committee considered how to characterize the inherent toxicity of a chemical, evaluate metabolic and pharmacokinetic attributes that can modify chemical toxicity, and integrate information over different domains. As discussed in Chapter 3, the conceptual framework includes models that use physicochemical and biological data to make predictions about potential acute toxicity. The predictive models can be qualitative (such as structural alerts) or quantitative (such as quantitative structure–activity relationship [QSAR] models), and the resulting outputs themselves might be qualitative (for example, a toxic or nontoxic determination) or quantitative (for example, a potency estimate). Medium-throughput and high-throughput assays will also be needed to predict acute mammalian toxicity as reviewed in Chapter 4. The toxicity predictions ultimately will depend on the collection and integration of the physicochemical and biological data that might be indicators of potential acute toxicity as discussed in Chapter 5.

One primary goal of the committee's conceptual framework is to develop data sufficient to categorize chemicals on the basis of their predicted acute toxicity. The committee considered existing toxicity-based chemical classification schemes developed for industrial chemicals, agrochemicals, biocides, and pharmaceuticals that often use lethality data to estimate toxicity, such as the dose required to kill 50% of a population of test animals (LD_{50}) (Seidle et al. 2010). Developing LD_{50} s was considered a useful benchmarking approach for predicting acute toxicity of chemicals of interest to the Department of Defense (DOD) because it allows DOD to exclude low-toxicity chemicals and to focus its resources on more toxic chemicals of concern. The committee considered the need to develop mechanistically based assays and to use well-characterized chemicals as positive controls to improve the predictive validity of LD_{50} s and to identify potential organ toxicities.

In the present chapter, the committee briefly discusses some current programs that are evaluating or developing modern testing strategies, reviews its suggested tiered testing strategy, and highlights important lessons learned from current testing programs that should be considered by DOD in developing its future testing strategy. The chapter also provides the committee's overall conclusions and identifies several steps that DOD could take in the short term to medium term (3–10 years) to implement a program that uses modern approaches to identify chemicals that have the potential to induce life-threatening acute toxicity in deployed personnel.²

¹As defined in Chapters 1 and 2, organ systems included the cardiovascular, respiratory, hepatic, renal, skeletomuscular, immune, and nervous systems, including special senses (vision and hearing).

²The committee recognizes that a more detailed research plan is needed and that development of such a plan is noted in the statement of task as a potential Phase 2 of this project.

MODERN APPROACHES FOR THE ASSESSMENT OF ACUTE CHEMICAL TOXICITY

The committee explored whether DOD could adopt a modern testing strategy for the prediction of acute toxicity.³ In particular, the committee's statement of task required a focus on the assessment of existing high-throughput screening (HTS) methods to identify acutely (and likely highly) toxic chemicals with greater predictivity. The committee considered whether several projects were relevant for DOD's purposes. For example, it examined the European Centre for the Validation of Alternative Methods (ECVAM) ACuteTox project⁴ whose stated aim is to develop a strategy to replace all in vivo tests of acute oral toxicity. The ACuteTox effort considers in vitro methods that address specific mechanisms of action relevant to acute systemic toxicity (such as assays for neurotoxicity) and includes such computational methods as QSAR modeling and physiologically based biokinetic modeling.

Initially, the ACuteTox project selected and tested 97 reference chemicals with six basal cytotoxicity assays and compared the results with published human and animal in vivo data (Clothier et al. 2008; Sjöström et al. 2008). Later, 57 reference chemicals were tested in a number of functional tests that covered absorption, distribution, metabolism, excretion, and specific organ and system toxicity, such as hematotoxicity, neurotoxicity, nephrotoxicity, and hepatotoxicity (Kinsner-Ovaskainen et al. 2009). Standardized experimental design and data acquisition were used in the second phase of ACuteTox program (Kopp-Schneider et al. 2013). Concentration–response data were routinely collected (for example, IC₂₀, IC₅₀, EC₂₀, or EC₅₀ values) and served as the statistical basis of the development of testing strategies (Kinsner-Ovaskainen et al. 2013; Prieto et al. 2013b). The oral acute-toxicity category was predicted using a chemical's physicochemical properties, in silico modeling results, and values (such as IC₅₀) obtained from the in vitro studies (Kopp-Schneider et al. 2013). In general, the predictive validity seen in the efforts has been moderate to low. In addition, there has been little effort to assess the predictive validity for highly toxic chemicals that would be of concern to DOD.

Examples of large-scale US projects include the Toxicology Testing in the 21st Century (Tox21) partnership and the US Environmental Protection Agency (EPA) ToxCast program. The ToxCast program uses a large suite of in vitro biochemical (cell-free) and cell-based assays to evaluate chemicals and to analyze their bioactivity profiles computationally (Kavlock et al. 2012; Judson et al. 2014; Kleinstreuer et al. 2014). Phase I of the ToxCast program included about 300 conventional pesticide active ingredients that were tested in a battery of cell-free and cell-based assays to evaluate the ability of the assays to predict potential human toxicity (Judson et al. 2010). In Phase II, the chemical space was broadened to include chemicals that are used in consumer products and industrial processes and unmarketed drugs that were donated by pharmaceutical companies (Sipes et al. 2013). However, few ToxCast assays were designed specifically to assess acute toxicity. Furthermore, there have been few efforts to use HTS approaches to evaluate acute toxicity. Regardless, the ToxCast program has identified a number of important technical issues that could be considered in the design of a program relevant for DOD (Tice et al. 2013).

³Modern approaches are ones that do not rely primarily on traditional toxicology testing and include computational and in vitro or nontraditional in vivo assays.

⁴See <http://www.acutetox.eu/>. The ACuteTox consortium was initially funded with €8 million by the European Commission's 6th Framework Programme and included 35 academic, industrial, and government research institutes in 13 European countries (Clemmedson 2008). The project was divided into 10 work packages that included selection of reference chemicals, generation of animal and human in vivo databases, development of an Internet-based database for central management of all project data, adaptation of promising in vitro methods to robotic screening platforms, statistical analysis (identification of outliers and design of the preliminary algorithm or prediction model), development of in vitro assays for neurotoxicity, and construction and optimization of the in vitro testing strategy.

The committee also explored modern toxicity-testing strategies that are under development in the pharmaceutical industry. For example, the pharmaceutical industry is exploring ways to predict drug-induced liver injury (DILI). DILI is a low-incidence but important idiosyncratic cause of drug toxicity and a major reason for attrition during drug development or withdrawal from the market (Stephens et al. 2014). Traditional toxicity-testing strategies do not predict DILI reliably in patients: fewer than 55% and 25% of DILI drugs are predicted on the basis of the regulatory animal-toxicity studies and simple in vitro tests, respectively (Olson et al. 2000; Xu et al. 2004). Box 6-1 describes the recent development of predictive models that can be used in pre-clinical studies to detect DILI risk in humans.

The committee was unable to find a robust modern testing program that DOD could readily adopt for its purposes. Lessons learned from the ToxCast, ACuteTox, and industry programs do, however, provide a great deal of guidance for DOD in designing a system that uses HTS approaches and predictive models to interpret the performance of cell-free and cell-based assays in predicting acute toxicity.

IMPLEMENTATION OF THE COMMITTEE'S CONCEPTUAL MODEL FOR ASSESSMENT OF ACUTE CHEMICAL TOXICITY

As discussed in Chapter 2, the committee developed a conceptual framework that links chemical structure, physicochemical properties, biochemical properties, and biological activity to acute toxicity. Implementation of the conceptual framework will require DOD to support development of a suite of databases, assays, models, and tools that are based on in vitro, nonmammalian in vivo, and in silico approaches for predicting acute toxicity. As presented in Chapter 2, these databases, assays, models, and tools would be used as part of a **tiered prioritization strategy** to predict acute toxicity (see Figure 6-1). Relatively easily obtained nontesting data (such as existing toxicity data or physicochemical properties; see Chapter 3 for more details) could allow initial evaluation of

BOX 6-1 An Example of an Integrated Testing Strategy for Predicting Drug-Induced Liver Injury

To facilitate identification of DILI drugs early in the drug-development process, Chen et al. (2014a) developed a tool that combines exposure and physicochemical data and a panel of in vitro high-content screening (HCS) assays to predict DILI. Implementation of the “rule-of-two” model (RO2) that combines high exposure (≥ 100 mg/day) and high lipophilicity ($\log P \geq 3$) resulted in high specificity (95%) but low sensitivity (27%); that equates to a high false-negative rate. The HCS panel alone, which measured eight cellular end points (including apoptosis, cell loss, DNA damage, DNA fragmentation, and mitochondrial potential), was marginally more sensitive (39%). Integration of the RO2 model with the HCS assay panel increased the sensitivity to 55% (specificity remained at 95%).

Thoughtful consideration of the data sources or models to be integrated and of the compatibility of the data streams from these sources is required to increase the likelihood of success of the prediction strategy. Reported successes (Rusyn et al. 2012; Chen et al. 2014a) typically resulted from integration of models that were based on different data sources that captured a greater diversity of information. In Chen et al. (2014a), only six of the 27 positives were predicted in both the RO2 and HCS strategies; this indicates the complementarity of the two approaches. Incorporation of mechanistic information into DILI predictions can improve the model performance (Chen et al. 2014b).

Published hybrid approaches to integrating chemical descriptors with in vitro data have had little success in improving predictions obtained on the basis of in vitro data alone (Low et al. 2011; Zhu et al. 2014). Those approaches directly pool data from disparate sources for modeling purposes. Incorporation of different modeling strategies that retain more information from heterogeneous data structures in hybrid integrations would likely maximize the data available for assessment and result in improved predictivity. Similarly, use of adapters might facilitate data availability and integrity (Chen et al. 2014b).

a large number of chemicals in the initial tiers. Higher testing tiers incorporate a variety of assays (in vitro or nontraditional in vivo) whose biological complexity increases as a chemical moves from one tier to the next. At each tier, a decision is made as to whether further assessment of toxicity is needed. The ultimate goal of the effort is to sort chemicals into categories that would indicate the predicted acute-toxicity hazard (for example, low toxicity, high toxicity, or uncertain toxicity because of inadequate data) and help to prioritize chemicals for follow-up evaluation. DOD can also apply its expertise to use other factors that were not considered by the committee (such as chemical availability, ease of chemical synthesis, and weaponizability) to exclude chemicals from further testing. One goal of the overall approach is to reduce the number of chemicals that progress through each tier in an efficient and cost-effective manner.

DEVELOPMENT OF A MODERN TIERED APPROACH FOR PREDICTING ACUTE TOXICITY: INITIAL STEPS

There are some initial steps that DOD could take to implement the committee's tiered testing strategy. This section describes in greater detail some approaches that might be useful for DOD to pursue in implementing modern testing approaches for predicting acute toxicity.

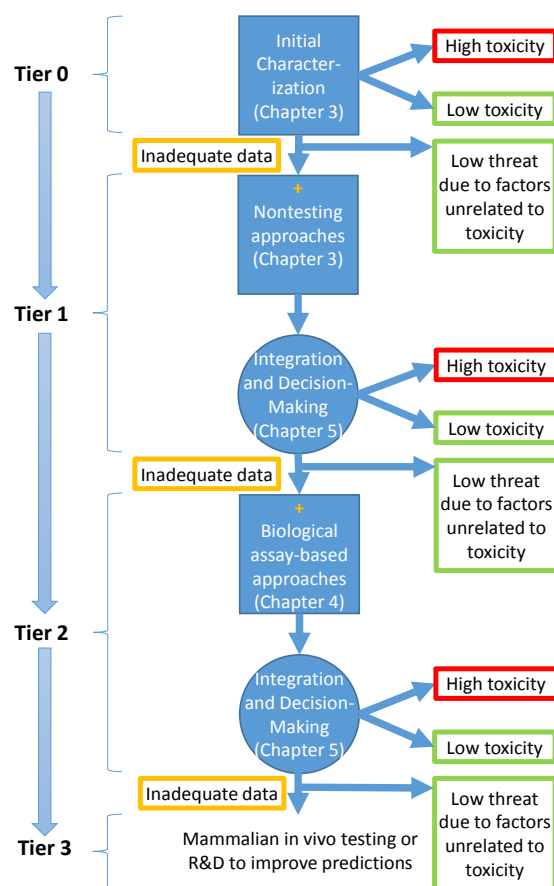


FIGURE 6-1 Prioritization strategy based on a tiered approach for using predictive-toxicology models and tools to evaluate agents for acute toxicity.

Computational Approaches

Chapter 3 discusses a variety of nontesting approaches that can be used to predict acute toxicity. Physicochemical data can be used to predict physical hazards or reactivity, such pharmacokinetic characteristics as metabolism and distribution, and likely routes of exposure. Those data can help to group chemicals by using chemical descriptors of various physicochemical or topological properties and to identify the biologically relevant chemical space (Lipinski and Hopkins 2004). Physicochemical data can also help to fill data gaps and support read-across approaches that apply data on a particular property or effect of a tested chemical to a similar untested chemical. A variety of *in silico* methods have been developed to predict the molecular sites where metabolism could occur and the types of metabolites that could be formed.⁵ Several available QSAR models use structural properties or physicochemical properties to predict acute oral LD₅₀s. Few such models are available for predicting inhalation LC₅₀s, and the committee was unable to identify models for predicting dermal LD₅₀s.

A few QSAR models predict neurotoxicity and cytotoxicity but not other end points that are relevant for acute, debilitating toxicity. Box 6-2 describes a QSAR model for the prediction of acetylcholinesterase (AChE) inhibition. Quantitative structure–toxicity relationship models have also been developed to evaluate the role of lipophilicity, polarity, molecular geometry, and quantum chemical descriptors for molecular orbital energy in the toxicity of organophosphate insecticides (Can 2014). Inhibition of AChE and other cholinesterases is an important step in the toxicity of nerve-gas agents (such as VX and soman) that are highly potent organophosphates; thus, the computational models that have been developed could be useful in a DOD testing strategy.

A large number of chemicals have reported LD₅₀ data, which cover a large portion of chemical space. Although LD₅₀ data are available on many chemicals, they are not informative about a chemical's mechanism of action. Such knowledge is important because acute toxicity might involve multiple biochemical mechanisms; this highlights the need for improved mechanistic insights about structure–toxicity relationships. Efforts to derive acute-toxicity QSAR models that have high predictive accuracy have fallen short because of mechanistic complexities.

High-Throughput Screens

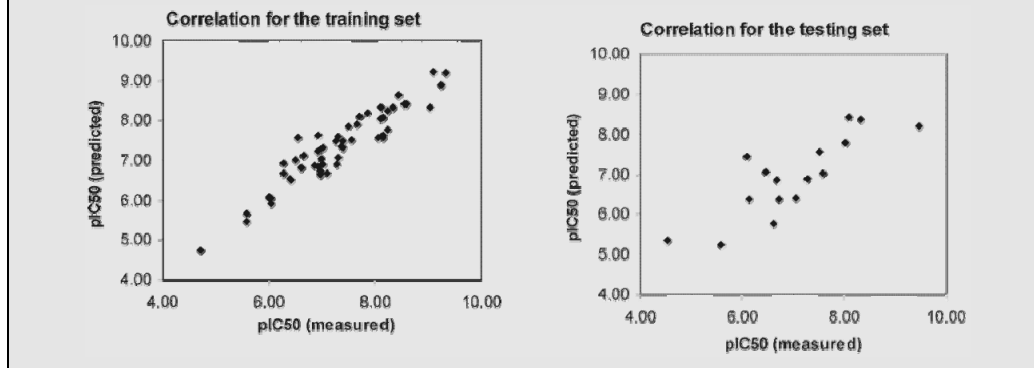
How the Tox21 and ToxCast data have been used to predict mammalian toxicity was demonstrated recently by the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), which evaluated whether HTS data could predict the results of an *in vivo* uterotrophic assay that screens for estrogen activity in rodents (see Box 6-3). The project is part of a larger NICEATM and EPA effort to develop a robust *in vitro* screen for chemicals with estrogenic or androgenic activity. The committee recognizes that estrogenic and androgenic end points are not relevant for the assessment of acute toxicities of concern to DOD, but the NICEATM and EPA efforts provide important insights. The committee found the targeted toxicity-prediction effort to be relatively successful (see Box 6-3); this finding is promising for the DOD goal of acute-toxicity prediction. However, estrogenic activity is, in principle, an easier target for prediction because it depends on specific gene-expression patterns for which sensitive and specific HTS assays can be designed. Nevertheless, the effort identified several needed features that can help to inform future DOD efforts, including the following:

- Performance standards for new assays that consider validated test methods, reference chemicals, and standards for assay accuracy and reliability (Wind and Stokes 2010; Stokes et al. 2012).

⁵Metabolites could also be identified through additional experimentation, for example, by using chemical analyses of biological test systems or by comparing results obtained from cell systems that use different levels of metabolic competence.

BOX 6-2 The Use of Computational Approaches for Evaluating Chemical-Induced Inhibition of Acetylcholinesterase

Receptor-specific scoring functions have been developed for predicting binding affinities of human acetylcholinesterase (AChE) inhibitors. A study performed by Guo et al. (2004) illustrates the general approach used to develop predictive (Q)SAR models. In this case, 69 chemicals with IC₅₀ data measured with a human AChE assay were selected for training and testing of the scoring function. The IC₅₀ of the 69 chemicals ranged from 0.33 to 30,000 nM. Docking calculations were carried out with published software (the Gold program). A weighted sum of electrostatic and van der Waals interactions between ligands and the receptor residues was calculated. Guo et al. examined the correlation of a calculated activity (pIC₅₀) with experimentally derived IC₅₀ data. The left figure shows the correlation for a 53-ligand training set (R² = 0.89). The right figure shows the results of using a novel 16-chemical test set (R² = 0.69).



- Suitable *in vivo* end points and study results that can be used to assess HTS assay performance (Rotroff et al. 2013); that is, there is a need for phenotypic anchoring of the *in vitro* results.
- Reference chemicals that have known biological activity (such as ER α in the cited NICEATM project) for evaluating the sensitivity, specificity, and predictivity of assays to identify agonists and antagonists for a biological target of interest (Huang et al. 2014).
- Appropriate data-integration models that can pool information from multiple assays (Rotroff et al. 2013); model performance can be evaluated by calculating sensitivity, specificity, and balanced accuracy for a specific set of criteria across chemical space (Rotroff et al. 2013; Cox et al. 2014).
- Recognition that misclassification of chemicals might occur (Rotroff et al. 2013; Cox et al. 2014). Possible explanations for misclassification of a chemical include a failure to test it at a high enough concentration to exhibit a response in ToxCast assays, inadequate metabolic competence of the test system, and species, tissue, or cell-type differences in response between the ToxCast assay and *in vivo* models that are used to assess HTS assay performance (Rotroff et al. 2013). Although not addressed directly by the ToxCast program, an additional source of misclassification of a chemical could be the variability in the *in vivo* data that are used to assess the ability of an assay to predict acute toxicity.

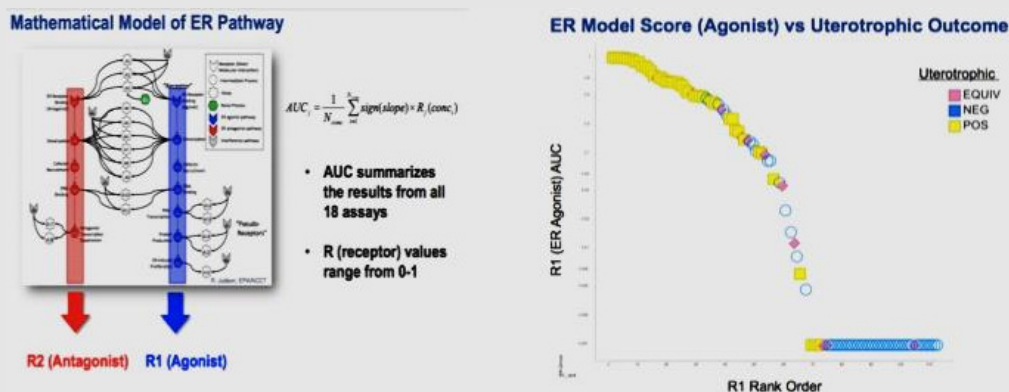
The example shown in Box 6-3 illustrates a current trend in the application of HTS testing as a replacement for some animal-based assays. In particular, the use of HTS assay data as part of an endocrine screening program has received some traction in the scientific community (Dix et al. 2007; Thomas et al. 2012; Rotroff et al. 2013; Thomas et al. 2013; Cox et al. 2014; Becker et al. 2015). In some cases (such as chemical interactions with the estrogen or androgen receptor system), adoption of HTS assays in the context of a tiered testing approach as a replacement for *in vivo* assays appears likely (van der Burg et al. 2015).

BOX 6-3 The Use of HTS Assays for Identifying Endocrine Disrupting Potential (Casey 2014)

In vivo phenotype: Estrogenic bioactivity was assessed with the EPA and Organisation for Economic Co-operation and Development (OECD) rodent uterotrophic assay in the ovariectomized rat, immature rat, or ovariectomized mouse. Duration of dosing (oral, subcutaneous, or intraperitoneal) was a minimum of 3 days. Suitable *in vivo* data on more than 100 chemicals were identified.

In vitro data: Results were obtained from 18 *in vitro* assays that measure estrogen receptor (ER)-mediated bioactivity. The *in vitro* assays (chosen from the EPA ToxCast or Tox21 program) probe perturbations of the ER pathway at multiple points (ER binding, receptor dimerization, DNA binding, RNA transcription, protein production, and cell proliferation) (Rotroff et al. 2014). Chemicals used in the evaluation included several reference chemicals of known activity against the ER (such as estradiol).

Outputs: A computational model was developed to calculate area-under-the-curve (AUC) scores for ER agonist (R1) and antagonist (R2) bioactivity. The scores were compared with the results of the uterotrophic assay results.



Conclusion: Data analyses showed good correlation between the calculated AUC scores (R1) and results of the uterotrophic assay. The study demonstrated that ToxCast *in vitro* assays perform adequately for prioritizing chemicals for further evaluation of ER activity, and the HTS assays are predictive of the likelihood of a positive or negative finding in more resource-intensive assays.

The predictiveness of many HTS assays, however, remains low (< 50%) (Thomas et al. 2012; Cox et al. 2014; Patlewicz et al. 2015). Many *in vivo* end points cannot be predicted any better by using HTS assay data than by using chemical descriptor information and QSAR models. Thomas et al. (2012) identified several factors that could account for the relatively poor ability of *in vitro* assays to predict *in vitro* responses, including (1) the failure of current *in vitro* assays to capture biochemical and cellular processes or properties in the *in vivo* tissues with adequate fidelity; (2) the possibility that *in vitro* assays do not capture biological context-specific outcomes reliably; (3) inadequate coverage of pathways, protein targets, and cell types; (4) substantial species differences between the *in vitro* assays and *in vivo* end points that are being predicted; and (5) the inadequate number of positive chemicals for each end point to capture the broad array of mechanisms that lead to *in vivo* toxicity.

Cytotoxicity Assays as General Indicators of Acute Toxicity

As discussed in Chapter 4, one approach to predicting acute toxicity relies on *in vitro* cyto-

toxicity assays that use human cells in culture (Seibert et al. 1996; Ekwall et al. 1998; Halle 2003; NICEATM 2006; Xia et al. 2008; Halwachs et al. 2013; Prieto et al. 2013a). That approach uses relatively simple assays and presumes that *in vivo* toxicity does not result primarily from impairment of specific functions of differentiated cells (Prieto et al. 2013a). In many cases, validation of the assays rests on an examination of the regression equation from the correlation of experimental IC_{50} ⁶ cytotoxicity values with published LD_{50} s,⁷ an approach that can be used to estimate unknown LD_{50} values from IC_{50} cytotoxicity values measured *in vitro* (Seibert et al. 1996). The committee examined the NICEATM and ECVAM validation study that assessed the predictive capacity of the BALB/3T3 neutral-red uptake cytotoxicity assay (see Box 6-4); that assay has been evaluated for its ability to identify chemicals that would be labeled as toxic or hazardous ($LD_{50} < 2,000$ mg/kg) (NTP 2014). The overall results of the studies have shown that there is a relatively good correlation of around 60–70% between *in vitro* cytotoxic concentrations (IC_{50} s) and rat oral LD_{50} s (JRC 2013).

Although cytotoxicity assays hold some promise for the prediction of acute toxicity, several important caveats and assay limitations need to be considered, including the following:

- Little evidence of assay performance exists for highly toxic chemicals. DOD is faced with identifying agents that have extremely low LD_{50} s or LC_{50} s. For example, acute oral LD_{50} values reported in mice for some agents of current concern to DOD are well below 1 mg/kg: botulinum toxin (0.001 μ g/kg), ricin (3 μ g/kg), VX (15 μ g/kg), anatoxin-a(s) (50 μ g/kg), soman (64 μ g/kg), and sarin (100 μ g/kg).⁸ In addition, those examples of highly toxic chemicals have different mechanisms of action, including inhibition of cholinesterase activity (for example, anatoxin-a(s), VX, soman, and sarin), ribosome inactivation (ricin), and blockade of acetylcholine secretion (botulinum toxin). Applying the BALB/3T3 NRU cytotoxicity assay to a set of 67 chemicals and using the rat LD_{50} data from the Registry of Cytotoxicity showed that predictions for highly toxic chemicals were generally poor—0/6 for chemicals with an LD_{50} below 5 mg/kg and 1/11 chemicals with an LD_{50} of 5–50 mg/kg (NICEATM 2006). Substances with LD_{50} s of 300–2,000 mg/kg were predicted better by this assay—about 81% accuracy (NICEATM 2006).

- Prediction of acute toxicity with cytotoxicity assays remains highly variable. There are several possible reasons why acute systemic toxicity of some chemicals is poorly predicted by basal cytotoxicity assays. First, toxicity might be due to tissue- or organ-specific effects caused by the chemical of interest. For example, substances with a mechanism of action on molecular targets (receptors, channels, and enzymes) are not modeled in most cell lines that are used for cytotoxicity assays, and this probably accounts for poor prediction of the toxicity of anatoxin, botulinum toxin, soman, and sarin, which perturb neuronal synapses. Second, even though a generally toxic mechanism is modeled by a cell line, the cell line could be much less sensitive to this mechanism than is some sensitive cell type in the human body. For example, ricin targets protein synthesis, which is needed by all cells, but in the human body, only specific organs are highly sensitive. Third, chemicals have restricted accessibility to target cells or tissues. Fourth, chemical toxicity might depend on bioactivation pathways that are absent in the test system. Fifth, a chemical might be quickly eliminated or detoxified through metabolism; thus, *in vitro* results might overpredict toxicity seen *in vivo*.

- Assay limitations can also contribute to reduced predictive power of *in vitro* cytotoxicity assays. For example, actual concentrations available to cells or to intracellular targets in *in vitro* tests might be much lower than the nominal concentrations.

⁶ IC_{50} is the concentration of a substance that causes 50% inhibition *in vitro*.

⁷In some cases, LD_{50} values are available from the Registry of Cytotoxicity (Halle 2003).

⁸ LD_{50} data available from Franz (1997).

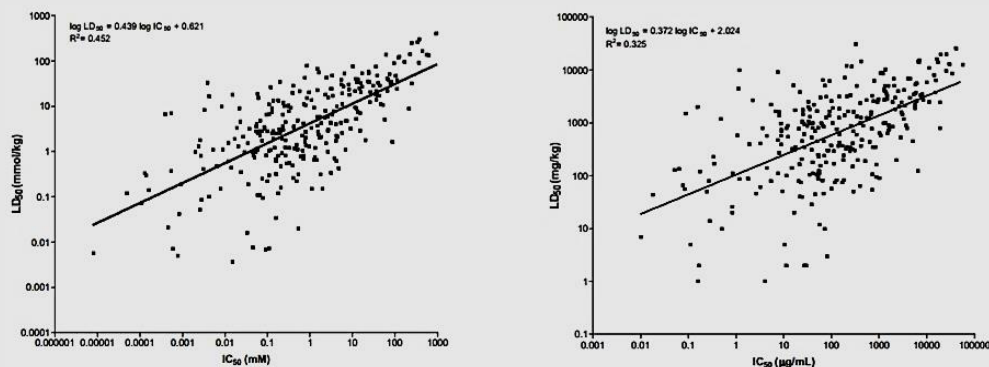
BOX 6-4 The Use of the BALB/3T3 Neutral-Red Uptake Cytotoxicity Assay to Predict Acute Toxicity

Assay: The 3T3 neutral-red uptake (NRU) cytotoxicity assay uses the BALB/c3T3 mouse fibroblast cell line and is based on the ability of viable cells to incorporate and bind the dye neutral red (NR). The assay is usually performed as a 96-well cytotoxicity-based assay that spectrophotometrically measures (Stokes et al. 2008) the concentration-dependent reduction in NRU by cells after exposure to a test material. The basic concept of basal cytotoxicity assays is that chemicals exert their toxic effects by disrupting structures and functions universal to all cells, such as cell membranes (Gennari et al. 2004). With the basal cytotoxicity assays, it is possible to quantify the cytotoxicity of a chemical by its IC_{50} value, that is, the concentration of the tested substance that decreases cell viability by 50% in the cell culture.

Chemicals: Chemicals tested in this assay included pharmaceuticals, pesticides, industrial chemicals, and food additives. The number of chemicals varied between test phases and ranged from about 70 to 300.

Predictive approach: The study assessed the predictive capacity of the assay in conjunction with a dichotomous prediction model that yielded only two categorical predictions: potential negative (predicted $LD_{50} > 2,000$ mg/kg) and potential positive (predicted $LD_{50} < 2,000$ mg/kg).

The figure below (from NICEATM 2006, p. 6-19) shows the regressions that used the Registry of Cytotoxicity (RC; Halle 2003) rat acute oral LD_{50} data on millimolar (LEFT) or weight (RIGHT) units for 282 test substances.



Overall assay performance: The 3T3 NRU method was shown to have a high sensitivity (92–96%) and a low false-negative rate (4–8%) (Prieto et al. 2013a). The assay also had a high false-positive rate or low specificity, which limited the usefulness of positive test results and led to a comparably low rate of identification of true negatives as such (40–44%). However, negative test results of the 3T3 NRU were largely accurate, that is, substances identified as negatives had low toxicity (low false-negative rate).

Assay limitations: The 3T3 NRU assay addresses specifically the toxicity mechanism of basal cytotoxicity; fibroblast cells cannot be used to evaluate interactions of chemicals with neuronal or cardiac receptors and ion channels and other tissue-specific molecular targets (NIEHS 2009). Furthermore, the cell line lacks metabolic competence associated with Phase I and Phase II biotransformation and so is sensitive to cytotoxicity induced by the parent chemical and not its metabolites.

The ACuteTox project identified several important technical considerations in the design of an in vitro testing strategy (Kopp-Schneider et al. 2013). Some of the more important considerations were the following:

- Select chemicals to span a wide range of outcomes of interest.
- Perform all assays with all chemicals.
- Be aware that chemical solubility might limit the concentrations that can be used in the test system.
 - Design experiments carefully to guarantee reliable and meaningful estimates.
 - Avoid overfitting of the prediction model, which occurs when a model is fitted exactly to the training data; this hinders performance in future screening efforts or applications.
 - Use cross-validation and bootstrapping to estimate error rates.

A QSAR model based on *in vitro* cytotoxicity data and oral LD₅₀ values from the Registry of Cytotoxicity has been developed for use in predicting acute toxicity in rodents (Freidig et al. 2007). The predictions from that model tend to overestimate toxicity; thus, substances that were predicted to have no or low toxicity with that model could be eliminated from additional *in vivo* testing.

Mechanistically Based Assays

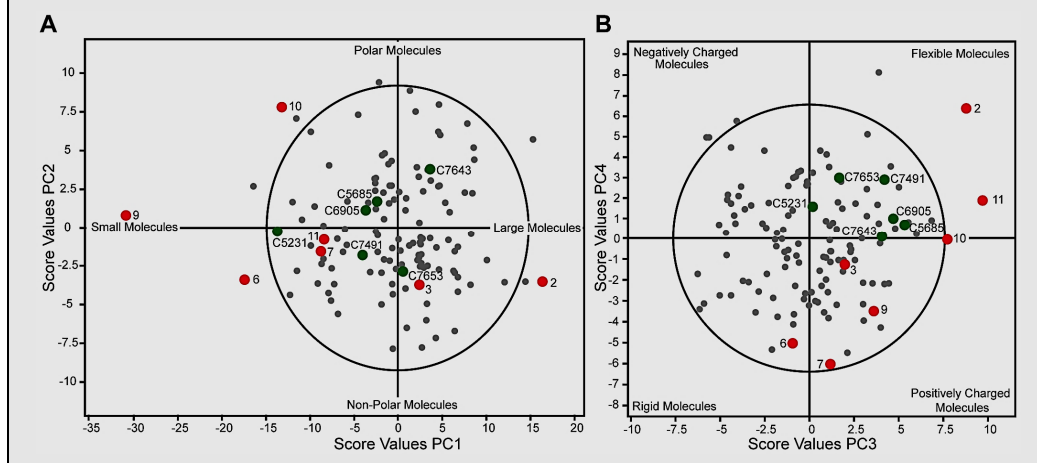
Another approach that has been used to predict acute chemical toxicity is based on the development of assays that evaluate a mechanism of action known to be critical for a chemical's toxic response. One mechanism-based example explored by the committee uses HTS assays and computational approaches to predict chemical-induced inhibition of AChE activity. Inhibition of AChE results in an accumulation of acetylcholine (ACh) at cholinergic synapses and associated clinical signs. A variety of agents, including nerve agents (such as VX and soman) and other toxic organophosphorus (OP) chemicals and some carbamates, inhibit AChE activity. Several molecular models have been developed that use reference chemicals to describe inhibitor binding with human (and other) AChE (Barril et al. 2001; Kua et al. 2002; Guo et al. 2004; Akula et al. 2006; Sopkova de Oliveira Santos et al. 2010; Gupta and Mohan 2011; Deb et al. 2012). The models evaluate inhibitor interactions at the two principal binding sites—catalytic anionic site (CAS) and a peripheral anionic site (PAS)—in the AChE enzyme. The nerve gases and other classical AChE inhibitors bind a phosphoryl group on a serine residue in the CAS (Deb et al. 2012). Another approach for assessing chemical–AChE interactions is to evaluate how the chemical of interest docks with the active site and other portions of the protein. Several structure–activity relationships (SARs) based on *pose*⁹ predictions for the interactions have been developed (Huang et al. 2010; Samadi et al. 2010). Results obtained with those computational approaches often depend on the type of ligand, the protein conformation, and the presence of water (Berg et al. 2011). Box 6-5 illustrates several examples of how an HTS approach can be used to screen chemicals as AChE inhibitors.

Similar mechanism-based screening efforts have been developed by the pharmaceutical industry and others to develop *in vitro* models that can predict *in vivo* biology in support of drug-safety assessment or drug discovery. For example, some investigators have examined the relationship between cytotoxicity data, basal gene-expression measurements, and a chemical's structure to identify putative molecular targets and the role of gene expression in cytotoxicity (Covell et al. 2005). Another example involves screening chemicals for their ability to inhibit mitochondrial complex I of the electron transport chain (Glover et al. 2007). Berg et al. (2006) successfully grouped chemicals by mechanism of action (for example, modulators of NFκB signaling or the phosphatidylinositol 3-kinase/Akt signal-transduction pathway) by using data from human cell-based HTS assays. Such approaches are consistent with the development of an adverse-outcome pathway (AOP), which is a linear conceptual model that links a molecular initiating event, key events, and an adverse outcome (Figure 3-1) (Ankley et al. 2010; Garcia-Reyero 2015).

⁹The term *pose* refers to the conformation and alignment of a molecule (Coats 2002).

BOX 6-5 The Use of HTS Assays to Evaluate Inhibition of Acetylcholinesterase

Berg et al. (2011) screened a chemical library consisting of 17,500 substances by using the colorimetric Ellman assay adapted to a 96-well format and a recombinant *human* AChE. The hydrolysis of acetylthiocholine iodide was monitored, and the average slope of the positive controls was set to 100% activity. At an assay concentration of 50 μM , 124 chemicals reduced the enzymatic activity of AChE by at least 70% in the single-replicate assays. The chemicals had a full dose–response curve determined to verify chemical inhibitory activity and identify potential false-positive results. A second set of chemicals that had *no* activity in the HTS but structural and physicochemical features similar to those of the positive hits was used to identify false negatives. The AChE inhibitors discovered in the screening campaign are chemically diverse, with molecular weights ranging from 234 to 596 Da, $\log P(o/w)$ from -1.16 to 8.14 , and 0–12 rotatable bonds. Five principal components (PCs) proved significant: they were related mainly to size, hydrophobicity, flexibility, charge (positive, neutral or negative), and electronic properties associated with halogens and aromatic elements (Berg et al. 2011). The figure below illustrates the chemical space as established by PC analysis of the physicochemical properties of the 124 hits (gray dots) that were identified in the HTS performed by Berg et al. (2011).



An AOP begins with a well-defined molecular initiating event and then describes the key events along a biological pathway that ultimately lead to an adverse outcome associated with chemical exposure. One has been developed for lethality associated with AChE-inhibiting organophosphate and carbamate pesticides (Russom et al. 2014). That AOP considers the role of metabolism (such as metabolic activation by cytochrome P450s to form oxon metabolites of some organophosphate pesticides) and links the main initiating event (cholinesterase inhibition) with various physiological cholinergic responses. Whether that or other AOPs in development (such as the ones for nonpolar narcosis and mitochondrial toxicity¹⁰) will be relevant for the highly toxic chemicals of concern to DOD is unknown but deserves further study.

OTHER CONSIDERATIONS

The committee has identified several broad technical considerations that are important, such as the use of reference chemicals and development of appropriate data-integration models.

¹⁰OECD provides additional information about developing AOPs at <http://www.oecd.org/chemicalsafety/testing/listsofprojectsontheopdevelopmentprogrammeworkplan.htm>.

DOD is faced with several additional challenges that will need to be considered as it develops a tiered testing program. They include technical challenges associated with the assessment of chemicals for direct debilitating portal-of-entry effects (such as skin blistering and pulmonary edema) because there are few *in vitro* systems for assessing the acute toxicity of inhaled chemicals or agents that produce skin blistering. Likewise, it is unknown whether toxicity screens developed to predict oral toxicity could also predict dermal LD₅₀s or inhalation LC₅₀s. Another important challenge that DOD faces is related to the broad array of chemicals (or chemical classes) that could require assessment. That issue might require DOD to develop distinct tiered approaches for different chemical classes, such as metals. DOD will also need to support the development of HTS assays for mechanisms of actions that are known or suspected to be involved in acute chemical toxicity. In the future, AOPs and QSARs could be developed to strengthen DOD's ability to predict acute chemical toxicity (Tollefsen et al. 2014). Finally, DOD will need to develop methods for integrating data across physicochemical and biological domains.

DOD has a history of using alternative test methods for the assessment of chemical-warfare agents. For example, the US Army's Medical Chemical Defense Research Program used a variety of assays to elucidate mechanisms of action and identify countermeasures for many of the classical chemical-warfare agents, such as sulfur mustard, lewisite, nerve agents, and hydrogen cyanide (see Siddell et al. 1997; Somani and Romano 2001). Considerable DOD effort went into being able to conduct the assays *safely* when working with agents, such as soman and VX, which have human dermal LD₅₀ estimates of 0.14 and 5 µg /kg of body weight, respectively. If the overt toxicity of a new or novel chemical is totally unknown and within an order of magnitude or so of a classical chemical-warfare agent's potency, serious consideration of the engineering controls necessary to conduct the assays will be required. Because few research facilities are equipped to work with known highly toxic chemical-warfare agents as reference chemicals, DOD might be required initially to use surrogate chemicals that have a shared chemical mechanism or clinical effect of more militarily relevant agents.

The committee expects that in the next 3–10 years any tiered testing approach will not be able to replace completely the need for follow-up targeted *in vivo* studies to confirm the toxicity of a chemical of interest. Indeed, the state of the science suggests that development of a predictive acute-toxicity program will require extensive DOD investment in development of fit-for-purpose assays, (Q)SAR and other computational modeling approaches, and *in vitro*–*in vivo* extrapolation and data-integration methods for the prediction of acute toxicity. To begin the investment, the committee recommends that DOD initiate pilot studies that evaluate chemical classes of highest concern with well-characterized reference chemicals. The pilot studies will allow it to develop the assays and tools that are needed to predict acute chemical toxicity efficiently and accurately and to evaluate the rate of false negatives and false positives. The pilot studies could also examine how generalizable the results of various assays and tools are from one chemical class¹¹ to another. That research would allow DOD to begin to address the size of the chemical space needed to make predictions about unknown chemicals. DOD could benefit from leveraging its efforts with existing federal activities, such as the ToxCast program. Collaboration would allow DOD to complete pilot studies more rapidly.

FINDINGS AND RECOMMENDATIONS

- **Finding:** On the basis of its review of the current state of the science, the committee concludes that development of a screening program to predict acute toxicity that incorporates such information as existing toxicity data, physicochemical properties, and results from *in silico*

¹¹In this context, chemical class is defined broadly to include structurally related chemicals, chemicals that have different mechanisms of action, and chemicals that have different toxic end points, such as hepatotoxicity and neurotoxicity.

modeling and in vitro testing is consistent with and supported by other testing frameworks that use modern toxicology data.

- **Finding:** The current state of the science for prediction of acute toxicity with computational approaches is seeing steady growth. An investment by DOD in computational approaches could yield benefits in characterizing the toxicity of chemicals on which toxicity data are sparse or lacking.

- **Finding:** Prediction of acute toxicity with HTS assays is in its infancy, and DOD will need to evaluate what assays or approaches might be applicable for evaluating acute toxicity for its system and consider the lessons learned from current large-scale HTS programs. Regardless, an investment by DOD in HTS approaches could yield benefits in characterizing the toxicity of chemicals on which few or no toxicity data are available. HTS approaches might prove useful in excluding chemicals that have low toxic potential (for example, rodent LD₅₀ greater than 1,000 mg/kg) from further testing and might also help to identify more toxic chemicals of concern for further testing.

- **Finding:** There are sufficient data to suggest that DOD could use simple cytotoxicity assays to identify chemicals that have low acute-toxicity potential; this would allow it to focus its attention on more toxic chemicals. Additional effort is needed to determine whether the assays are relevant for the identification of highly toxic chemicals that could be used against deployed troops. Furthermore, because simple cytotoxicity assays fail to predict toxicity of highly toxic chemicals that act on specific molecular targets, such as neuronal synapses, they would need to be supplemented with assays designed specifically to detect such effects.

- **Finding:** On the basis of scientific advances, the committee concludes that the development of appropriate cellular models and targeted mechanistically based assays could provide DOD with a useful resource for understanding and predicting potential toxicity of chemicals. In particular, having more explicit knowledge available on the numerous mechanisms of action that lead to acute systemic toxicity would be valuable in the design and validation of integrated prediction methods.

- **Recommendation:** DOD should initiate pilot studies that evaluate chemical classes of highest concern with well-characterized reference chemicals. The pilot studies will allow it to develop the assays and tools that are needed to predict acute chemical toxicity efficiently and accurately and to evaluate the rate of false negatives and false positives.

REFERENCES

- Akula, N., L. Lecanu, J. Greeson, and V. Papadopoulos. 2006. 3D QSAR studies of AChE inhibitors based on molecular docking scores and CoMFA. *Bioorg. Med. Chem. Lett.* 16(24):6277-6280.
- Ankley, G.T., R.S. Bennett, R.J. Erickson, D.J. Hoff, M.W. Hornung, R.D. Johnson, D.R. Mount, J.W. Nichols, C.L. Russom, P.K. Schmieder, J.A. Serrano, J.E. Tietge, and D.L. Villeneuve. 2010. Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environ. Toxicol. Chem.* 29(3):730-741.
- Barril, X., M. Orozco, and F.J. Luque. 2001. Towards improved acetylcholinesterase inhibitors: A structural and computational approach. *Mini Rev. Med. Chem.* 1(3):255-266.
- Becker, R.A., K.P. Friedman, T.W. Simon, M.S. Marty, G. Patlewicz, and J.C. Rowlands. 2015. An exposure: Activity profiling method for interpreting high-throughput screening data for estrogenic activity-Proof of concept. *Regul. Toxicol. Pharmacol.* 71(3):398-408.
- Berg, E.L., E.J. Kunkel, E. Hytopoulos, and I. Plavec. 2006. Characterization of compound mechanisms and secondary activities by BioMAP analysis. *J. Pharmacol. Toxicol. Meth.* 53(1):67-74.
- Berg, L., C.D. Andersson, E. Artursson, A. Hörnberg, A.K. Tunemalm, A. Linusson, and F. Ekström. 2011. Targeting acetylcholinesterase: Identification of chemical leads by high throughput screening, structure determination and molecular modeling. *PLoS One* 6(11):e26039.
- Can, A. 2014. Quantitative structure-toxicity relationship (QSTR) studies on the organophosphate insecticides. *Toxicol. Lett.* 230(3):434-443.

- Casey, W. 2014. Assessing the Biological Relevance of In Vitro Data: A Case Study Using Estrogen Pathway Signaling. Presentation at NTP Board of Scientific Counselors (BSC) Meeting, December 9-10, 2014 [online]. Available: http://ntp.niehs.nih.gov/ntp/about_ntp/bsc/2014/dec/presentations/04casey_bscdec2014_508.pdf [accessed April 6, 2015].
- Chen, M., C.W. Tung, Q. Shi, L. Guo, L. Shi, H. Fang, J. Borlak, and W. Tong. 2014a. A testing strategy to predict risk for drug-induced liver injury in humans using high-content screen assays and the 'rule-of-two' model. *Arch. Toxicol.* 88(7):1439-1449.
- Chen, M., H. Bisgin, L. Tong, H. Hong, H. Fang, J. Borlak, and W. Tong. 2014b. Toward predictive models for drug-induced liver injury in humans: Are we there yet? *Biomark. Med.* 8(2):201-213.
- Clemedson, C. 2008. The European ACuteTox project: A modern integrative in vitro approach to better prediction of acute toxicity. *Clin. Pharmacol. Ther.* 84(2):200-202.
- Clothier, R., P. Dierickx, T. Lakhansky, M. Fabre, M. Betanzos, R. Curren, M. Sjöström, H. Raabe, N. Bourne, V. Hernandez, J. Mainez, M. Owen, S. Watts, and R. Anthonissen. 2008. A database of IC50 values and principal component analysis of results from six basal cytotoxicity assays, for use in the modelling of the in vivo and in vitro data of the EU ACuteTox project. *ATLA* 36(5):503-519.
- Coats, E.A. 2002. Part III. 3D QSAR applications: The CoMFA steroids as a benchmark dataset for development of 3D QSAR methods. Pp. 197-213 in *3D QSAR in Drug Design, Vol. 3. Recent Advances*, H. Kubinyi, G. Folkers, and Y.C. Martin, eds. New York: Kluwer.
- Covell, D.G., A. Wallqvist, R. Huang, N. Thanki, A.A. Rabow, and X.J. Lu. 2005. Linking tumor cell cytotoxicity to mechanism of drug action: An integrated analysis of gene expression, small-molecule screening and structural databases. *Proteins* 59(3):403-433.
- Cox, L.A., D. Popken, M.S. Marty, J.C. Rowlands, G. Patlewicz, K.O. Goyak, and R.A. Becker. 2014. Developing scientific confidence in HTS-derived prediction models: Lessons learned from an endocrine case study. *Regul. Toxicol. Pharmacol.* 69(3):443-450.
- Deb, P.K., A. Sharma, P. Piplani, and R.R. Akkinapally. 2012. Molecular docking and receptor-specific 3D-QSAR studies of acetylcholinesterase inhibitors. *Mol. Divers.* 16(4):803-823.
- Dix, D.J., K.A. Houck, M.T. Martin, A.M. Richard, R.W. Setzer, and R.J. Kavlock. 2007. The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol. Sci.* 95(1):5-12.
- Ekwall, B., F.A. Barile, A. Castaño, C. Clemedson, R.H. Clothier, P. Dierickx, B. Ekwall, M. Ferro, G. Fiskesjö, L. Garza-Ocañas, M.J. Gómez-Lechón, M. Gülden, T. Hall, B. Isomaa, A. Kahru, G. Kerszman, U. Kristen, M. Kunimoto, S. Kärenlampi, L. Lewan, A. Loukianov, T. Ohno, G. Persoone, L. Romert, T.W. Sawyer, R. Shrivastava, H. Segner, A. Stamatii, N. Tanaka, M. Valentino, E. Walum, and F. Zucco. 1998. MEIC evaluation of acute systemic toxicity. Part VI. The prediction of human toxicity by rodent LD50 values and results from 61 in vitro methods. *ATLA* 26(Suppl. 2):617-658.
- Franz, D.R. 1997. Defense against Toxin Weapons. US Army Medical Research and Materiel Command [online]. Available: <http://www.usamriid.army.mil/education/defensetox/toxdefbook.pdf> [accessed December 24, 2014].
- Freidig, A.P., S. Dekkers, M. Verwei, E. Zvinavashe, J.G. Bessems, and J.J. van de Sandt. 2007. Development of a QSAR for worst case estimates of acute toxicity of chemically reactive compounds. *Toxicol. Lett.* 170(3):214-222.
- Garcia-Reyero, N. 2015. Are adverse outcome pathways here to stay? *Environ. Sci. Technol.* 49(1):3-9.
- Gennari, A., C. van den Berghe, S. Casati, J. Castell, C. Clemedson, S. Coecke, A. Colombo, R. Curren, G. Dal Negro, A. Goldberg, C. Gosmore, T. Hartung, I. Langezaal, I. Lessigiarska, W. Maas, I. Mangelsdorf, R. Parchment, P. Prieto, J. Riego Sintes, M. Ryan, G. Schmuck, K. Stitzel, W. Stokes, J.A. Vericat, and L. Gribaldo. 2004. Strategies to replace in vivo acute systemic toxicity testing. The report and recommendations of ECVAM Workshop 50. *ATLA* 32(4):437-459.
- Glover, C.J., A.A. Rabow, Y.G. Isgor, R.H. Shoemaker, and D.G. Covell. 2007. Data mining of NCI's anti-cancer screening database reveals mitochondrial complex I inhibitors cytotoxic to leukemia cell lines. *Biochem. Pharmacol.* 73(3):331-340.
- Guo, J., M.M. Hurley, J.B. Wright, and G.H. Lushington. 2004. A docking score function for estimating ligand-protein interactions: Application to acetylcholinesterase inhibition. *J. Med. Chem.* 47(22):5492-5500.
- Gupta, S., and C.G. Mohan. 2011. 3D-pharmacophore model based virtual screening to identify dual-binding site and selective acetylcholinesterase inhibitors. *Med. Chem. Res.* 20(9):1422-1430.
- Halle, W. 2003. The registry of cytotoxicity: Toxicity testing in cell cultures to predict acute toxicity (LD50) and to reduce testing in animals. *ATLA* 31(2):89-198.
- Halwachs, S., C. Lakoma, and W. Honscha. 2013. Validation of the hepatocyte-like HPCT-1E3 cell line as an in vitro model for the prediction of acute in vivo toxicity. *ATLA* 41(5):369-383.

110 *Application of Modern Toxicology Approaches for Predicting Acute Toxicity for Chemical Defense*

- Huang, L., A.D. Shi, F. He, and X.S. Li. 2010. Synthesis, biological evaluation, and molecular modeling of berberine derivatives as potent acetylcholinesterase inhibitors. *Bioorg. Med. Chem.* 18(3):1244-1251.
- Huang, R., S. Sakamuru, M.T. Martin, D.M. Reif, R.S. Judson, K.A. Houck, W. Casey, J.H. Hsieh, K.R. Shockley, P. Ceger, J. Fostel, K.L. Witt, W. Tong, D.M. Rotroff, T. Zhao, P. Shinn, A. Simeonov, D.J. Dix, C.P. Austin, R.J. Kavlock, R.R. Tice, and M. Xia. 2014. Profiling of the Tox21 10K compound library for agonists and antagonists of the estrogen receptor alpha signaling pathway. *Sci. Rep.* 4:5664.
- JRC (Joint Research Centre). 2013. EURL ECVAM Recommendation on the 3T3 Neutral Red Uptake Cytotoxicity Assay for Acute Oral Toxicity Testing. Report EUR 25946. Luxembourg: Office of European Union [online]. Available: https://eurl-ecvam.jrc.ec.europa.eu/eurl-ecvam-recommendations/file-s-3t3/ReqNo_JRC79556_lbna25946enn.pdf [accessed December 24, 2014].
- Judson, R.S., K.A. Houck, R.J. Kavlock, T.B. Knudsen, M.T. Martin, H.M. Mortensen, D.M. Reif, D.M. Rotroff, I. Shah, A.M. Richard, and D.J. Dix. 2010. In vitro screening of environmental chemicals for targeted testing prioritization: The ToxCast project. *Environ. Health Perspect.* 118(4):485-492.
- Judson, R., K. Houck, M. Martin, T. Knudsen, R.S. Thomas, N. Sipes, I. Shah, J. Wambaugh, and K. Crofton. 2014. In vitro and modelling approaches to risk assessment from the US Environmental Protection Agency ToxCast programme. *Basic Clin. Pharmacol. Toxicol.* 115(1):69-76.
- Kavlock, R., K. Chandler, K. Houck, S. Hunter, R. Judson, N. Kleinstreuer, T. Knudsen, M. Martin, S. Padilla, D. Reif, A. Richard, D. Rotroff, N. Sipes, and D. Dix. 2012. Update on EPA's ToxCast program: Providing high throughput decision support tools for chemical risk management. *Chem. Res. Toxicol.* 25(7):1287-1302.
- Kinsner-Ovaskainen, A., R. Rzepka, R. Rudowski, S. Coecke, T. Cole, and P. Prieto. 2009. Acutoxbase, an innovative database for in vitro acute toxicity studies. *Toxicol In Vitro* 23(3):476-485.
- Kinsner-Ovaskainen, A., P. Prieto, S. Stanzel, and Kopp-Schneider. 2013. Selection of test methods to be included in a testing strategy to predict acute oral toxicity: An approach based on statistical analysis of data collected in phase 1 of the ACuteTox project. *Toxicol. In Vitro* 27(4):1377-1394.
- Kleinstreuer, N.C., J. Yang, E.L. Berg, T.B. Knudsen, A.M. Richard, M.T. Martin, D.M. Reif, R.S. Judson, M. Polokoff, D.J. Dix, R.J. Kavlock, and K.A. Houck. 2014. Phenotypic screening of the ToxCast chemical library to classify toxic and therapeutic mechanisms. *Nat. Biotechnol.* 32(6):583-591.
- Kopp-Schneider, A., P. Prieto, A. Kinsner-Ovaskainen, and S. Stanzel. 2013. Design of a testing strategy using non-animal based test methods: Lessons learnt from the ACuteTox project. *Toxicol. In Vitro* 27(4):1395-1401.
- Kua, J., Y. Zhang, and J.A. Mccammon. 2002. Studying enzyme binding specificity in acetylcholinesterase using a combined molecular dynamics and multiple docking approach. *J. Am. Chem. Soc.* 124(28):8260-8267.
- Lipinski, C., and A. Hopkins. 2004. Navigating chemical space for biology and medicine. *Nature* 432(7019):855-861.
- Low, Y., T. Uehara, Y. Minowa, H. Yamada, Y. Ohno, T. Urushidani, A. Sedykh, E. Muratov, D. Fourches, H. Zhu, I. Rusyn, and A. Tropsha. 2011. Predicting drug-induced hepatotoxicity using QSAR and toxicogenomics approaches. *Chem. Res. Toxicol.* 24(8):1251-1262.
- NICEATM (National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods), 2006. In Vitro Cytotoxicity Test Methods for Estimating Acute Oral Systemic Toxicity. Background Review Document (BRD). NIH Publication No. 07-4518. National Institute of Health, National Institute of Environmental Health Sciences, Research Triangle Park, NC. November 2006 [online]. Available: <http://ntp.niehs.nih.gov/pubhealth/evalatm/test-method-evaluations/acute-systemic-tox/in-vitro-validation/brd/index.html> [accessed March 26, 2015].
- NIEHS (National Institute of Environmental Health Sciences). 2009. Report on the ICCVAM-ICEATM/ECVAM/JaCVAM Scientific Workshop on Acute Chemical Safety Testing: Advancing In Vitro Approaches and Humane Endpoints for Systemic Toxicity Evaluations. Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM), National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), National Institute of Environmental Health Sciences, National Toxicology Program, Research Triangle Park, NC [online]. Available: http://ntp.niehs.nih.gov/iccvam/docs/acute-tox_docs/workshop-feb09/acute-wksprpt.pdf [accessed April 6, 2015].
- NTP (National Toxicology Program). 2014. Validation Study of In Vitro Cytotoxicity Test Methods [online]. Available: <http://ntp.niehs.nih.gov/pubhealth/evalatm/test-method-evaluations/acute-systemic-tox/in-vitro-validation/> [accessed April 6, 2015].

- Olson, H., G. Betton, D. Robinson, K. Thomas, A. Monro, G. Kolaja, P. Lilly, J. Sanders, G. Sipes, W. Bracken, M. Dorato, K. Van Deun, P. Smith, B. Berger, and A. Heller. 2000. Concordance of the toxicity of pharmaceuticals in humans and in animals. *Regul. Toxicol. Pharmacol.* 32(1):56-67.
- Patlewicz, G., T.W. Simon, J.C. Rowlands, R.A. Budinsky, and R.A. Becker. 2015. Proposing a scientific confidence framework to help support the application of adverse outcome pathways for regulatory purposes. *Regul. Toxicol. Pharmacol.* 71(3):463-477.
- Prieto, P., T. Cole, R. Curren, R.M. Gibson, M. Liebsch, H. Raabe, A.M. Tuomainen, M. Whelan, and A. Kinsner-Ovaskainen. 2013a. Assessment of the predictive capacity of the 3T3 Neutral Red Uptake cytotoxicity test method to identify substances not classified for acute oral toxicity (LD50>2000 mg/kg): Results of an ECVAM validation study. *Regul. Toxicol. Pharmacol.* 65(3):344-365.
- Prieto, P., A. Kinsner-Ovaskainen, S. Stanzel, B. Albella, P. Artursson, N. Campillo, R. Cecchelli, L. Cerrato, L. Díaz, E. Di Consiglio, A. Guerra, L. Gombau, G. Herrera, P. Honegger, C. Landry, J.E. O'Connor, J.A. Páez, G. Quintas, R. Svensson, L. Turco, M.G. Zurich, M.J. Zurbano, and A. Kopp-Schneider. 2013b. The value of selected in vitro and in silico methods to predict acute oral toxicity in a regulatory context: Results from the European Project ACuteTox. *Toxicol. In Vitro* 27(4):1357-1376.
- Rotroff, D.M., D.J. Dix, K.A. Houck, T.B. Knudsen, M.T. Martin, K.W. McLaurin, D.M. Reif, M. Crofton, A.V. Singh, M. Xia, R. Huang, and R.S. Judson. 2013. Using in vitro high throughput screening assays to identify potential endocrine disrupting chemicals. *Environ. Health Perspect.* 121(1):7-14.
- Rotroff, D.M., D.J. Dix, K.A. Houck, T.B. Knudsen, M.T. Martin, K.W. McLaurin, D.M. Reif, K.M. Crofton, A.V. Singh, M. Xia, R. Huang, and R.S. Judson. 2014. Using in vitro high throughput screening assays to identify potential endocrine-disrupting chemicals. *Environ. Sci. Technol.* 48(15):8706-8716.
- Russom, C.L., C.A. LaLone, D.L. Villeneuve, and G.T. Ankley. 2014. Development of an adverse outcome pathway for acetylcholinesterase inhibition leading to acute mortality. *Environ. Toxicol. Chem.* 33(10):2157-2169.
- Rusyn, I., A. Sedykh, Y. Low, K.Z. Guyton, and A. Tropsha. 2012. Predictive modeling of chemical hazard by integrating numerical descriptors of chemical structures and short-term toxicity assay data. *Toxicol. Sci.* 127(1):1-9.
- Samadi, A., J. Marco-Contelles, E. Soriano, M. Álvarez-Pérez, M. Chioua, A. Romero, L. González-Lafuente, L. Gandía, J.M. Roda, M.G. López, M. Villarroya, A.G. García, and C. de los Ríos. 2010. Multipotent drugs with cholinergic and neuroprotective properties for the treatment of Alzheimer and neuronal vascular diseases. I. Synthesis, biological assessment, and molecular modeling of simple and readily available 2-aminopyridine-, and 2-chloropyridine-3,5-dicarbonitriles. *Bioorg. Med. Chem.* 18(16):5861-5872.
- Seibert, H., M. Balls, J.H. Fentem, V. Bianchi, R.H. Clothier, P.J. Dierickx, M. Garle, M.J. Gomez-Lechon, L. Gribaldo, M. Gulden, M. Liebsch, E. Rasmussen, R. Roguet, R. Shrivastana, and E. Walum, 1996. Acute toxicity testing in vitro and the classification and labelling of chemicals. The report and recommendations of ECVAM Workshop 16. *ATLA* 24:499-510.
- Seidle, T., S. Robinson, T. Holmes, S. Creton, P. Prieto, J. Scheel, and M. Chlebus. 2010. Cross-sector review of drivers and available 3Rs approaches for acute systemic toxicity testing. *Toxicol. Sci.* 116(2):382-396.
- Siddell, F.R., E.T. Takafuji, and D.R. Franz, eds. 1997. *Medical Aspects of Chemical and Biological Warfare*. Textbook of Medicine, Part 1. Warfare, Weaponry and the Casualty, Vol. 3. Office of Surgeon General, Borden Institute, Walter Reed Army Medical center, Washington, DC.
- Sipes, N.S., M.T. Martin, P. Kothiya, D.M. Reif, R.S. Judson, A.M. Richard, K.A. Houck, D.J. Dix, R.J. Kavlock, and T.B. Knudsen. 2013. Profiling 976 ToxCast chemicals across 331 enzymatic and receptor signaling assays. *Chem. Res. Toxicol.* 26(6):878-895.
- Sjöström, M., A. Kolman, C. Clemedson, and R. Clothier. 2008. Estimation of human blood LC50 values for use in modeling of in vitro-in vivo data of the ACuteTox project. *Toxicol. In Vitro* 22(5):1405-1411.
- Somani, S.M., and J.A. Romano, eds. 2001. *Chemical Warfare Agents: Toxicity at Low Levels*. Boca Raton, FL: CRC Press.
- Sopkova de Oliveira Santos, J., A. Lesnard, J.H. Agondanou, N. Dupont, A.M. Godard, S. Stiebing, C. Rochais, F. Fabis, P. Dallemagne, R. Bureau, and S. Rault. 2010. Virtual screening discovery of new acetylcholinesterase inhibitors issued from CERMN chemical library. *J. Chem. Inf. Model.* 50(3):422-428.
- Stephens, C., R.J. Andrade, and M.I. Lucena. 2014. Mechanisms of drug-induced liver injury. *Curr. Opin. Allergy Clin. Immunol.* 14(4):286-292.
- Stokes, W.S., S. Casati, J. Strickland, and M. Paris. 2008. Neutral red uptake cytotoxicity tests for estimating starting doses for acute oral toxicity tests. *Curr. Protoc. Toxicol.* (Suppl. 36):Unit 20.4.

- Stokes, W.S., J. Strickland, and W. Casey. 2012. Validation of the 21st century toxicology toolbox: Challenges, opportunities, and the way forward. Pp. 323-328 in Proceedings of the 8th World Congress on Alternatives and Animal Use in the Life Sciences, Montreal 2011. ALTEX Proceedings Vol. 1 [online]. Available: http://www.altex.ch/resources/WC8proceedings_full_issue.pdf [accessed March 27, 2015].
- Thomas, R.S., M.B. Black, L. Li, E. Healy, T.M. Chu, W. Bao, M.E. Andersen, and R.D. Wolfinger. 2012. A comprehensive statistical analysis of predicting in vivo hazard using high-throughput in vitro screening. *Toxicol. Sci.* 128(2):398-417.
- Thomas, R.S., M.A. Philbert, S.S. Auerbach, B.A. Wetmore, M.J. Devito, I. Cote, J.C. Rowlands, M.P. Whelan, S.M. Hays, M.E. Andersen, M.E. Meek, L.W. Reiter, J.C. Lambert, H.J. Clewell, III, M.L. Stephens, Q.J. Zhao, S.C. Wesselkamper, L. Flowers, E.W. Carney, T.P. Pastoor, D.D. Petersen, C.L. Yauk, and A. Nong. 2013. Incorporating new technologies into toxicity testing and risk assessment: Moving from 21st century vision to a data-driven framework. *Toxicol. Sci.* 136(1):4-18.
- Tice, R.R., C.P. Austin, R.J. Kavlock, and J.R. Bucher. 2013. Improving the human hazard characterization of chemicals: a Tox21 update. *Environ. Health Perspect.* 121(7):756-765.
- Tollefsen, K.E., S. Scholz, M.T. Cronin, S.W. Edwards, J. de Knecht, K. Crofton, N. Garcia-Reyero, T. Hartung, A. Worth, and G. Patlewicz. 2014. Applying Adverse Outcome Pathways (AOPs) to support Integrated Approaches to Testing and Assessment (IATA). *Regul. Toxicol. Pharmacol.* 70(3):629-640.
- van der Burg, B., E.B. Wedebye, D.R. Dietrich, J. Jaworska, I. Mangelsdorf, E. Paune, M. Schwarz, A.H. Piersma, and E.D. Kroese. 2015. The ChemScreen project to design a pragmatic alternative approach to predict reproductive toxicity of chemicals. *Reprod. Toxicol.* 55(1):114-123.
- Wind, M., and W.S. Stokes. 2010. Developing performance standards to expedite validation of innovative and improved test methods. *ALTEX* 27(Special Issue):97-102.
- Xia, M., R. Huang, K.L. Witt, N. Southall, J. Fostel, M.H. Cho, A. Jadhav, C.S. Smith, J. Inglese, C.J. Portier, R.R. Tice, and C.P. Austin. 2008. Compound cytotoxicity profiling using quantitative high-throughput screening. *Environ. Health Perspect.* 116(3):284-291.
- Xu, J.J., D. Diaz, and P.J. O'Brien. 2004. Applications of cytotoxicity assays and pre-lethal mechanistic assays for assessment of human hepatotoxicity potential. *Chem. Biol. Interact.* 150(1):115-128.
- Zhu, X, A. Sedykh, and S. Liu. 2014. Hybrid in silico models for drug-induced liver injury using chemical descriptors and in vitro cell-imaging information *J. Appl. Toxicol.* 34(3):281-288.

Appendix A

Biographical Information on the Committee on Predictive-Toxicology Approaches for Military Assessments of Acute Exposures

David C. Dorman (*Chair*) is a professor of toxicology in the Department of Molecular Biosciences of North Carolina State University. The primary objective of his research is to provide a refined understanding of chemically induced neurotoxicity in laboratory animals that will lead to improved assessment of potential neurotoxicity in humans. Dr. Dorman's research interests include neurotoxicology, nasal toxicology, pharmacokinetics, and cognition and olfaction in animals. He has served on multiple National Research Council committees and recently chaired the Committee on Design and Evaluation of Safer Chemical Substitutions: A Framework to Inform Government and Industry Decisions. He has served on other advisory boards for the US Navy, the National Aeronautics and Space Administration, and the US Department of Agriculture and is currently a member of the National Toxicology Program's Board of Scientific Counselors. He is an elected fellow of the Academy of Toxicological Sciences and a fellow of the American Association for the Advancement of Science. He received his DVM from Colorado State University. He completed a combined PhD and residency program in toxicology at the University of Illinois at Urbana-Champaign and is a diplomate of the American Board of Veterinary Toxicology and the American Board of Toxicology.

Weihshueh A. Chiu is a professor in the Department of Veterinary Integrative Biosciences in the College of Veterinary Medicine and Biomedical Sciences at Texas A&M University. Before joining Texas A&M University, Dr. Chiu worked at the US Environmental Protection Agency (EPA) for over 14 years, most recently as chief of the Toxicity Pathways Branch in the Integrated Risk Information System (IRIS) Division of the National Center for Environmental Assessment. His research has focused on human health risk assessment, particularly with respect to toxicokinetics, mechanisms of toxicity, physiologically based pharmacokinetic modeling, dose-response assessment, and characterizing uncertainty and variability. He led the development of EPA's 2011 IRIS assessment of trichloroethylene, which pioneered the use of probabilistic methods for characterizing uncertainty and variability in toxicokinetics and dose-response relationships. Dr. Chiu received his PhD in physics from Princeton University.

Haiyan Huang is an associate professor in the Department of Statistics at the University of California, Berkeley. She is interested in the interface between statistics and data-rich scientific disciplines, such as biology. Her research is focused on applied and theoretical statistics, high dimensional and integrative genomic data analysis, hierarchical multilabel classification, and translational bioinformatics. Dr. Huang earned a PhD in applied mathematics from the University of Southern California.

Andy Nong is the lead computational toxicologist of Health Canada. His research focuses on computer models of biological systems that can be applied to understand and predict the fate of a chemical dose in the body and its possible health effects. Dr. Nong also explores different types of computing approaches (pharmacokinetic models, benchmark dosing, chemical structure–activity regression, and systems-biology models) that can help to evaluate chemical safety and eventually support the screening of a larger set of chemicals that lead to similar health outcomes. Before coming to Health Canada, he worked as a research investigator with the Hamner Institutes for Health Sciences. Dr. Nong received his PhD in public health and toxicology from the University of Montreal.

Grace Patlewicz is a research chemist at the National Center for Computational Toxicology of the US Environmental Protection Agency (EPA). Before joining EPA, she was employed as a computational toxicologist by DuPont’s Haskell Global Centers for Health and Environmental Sciences where she acted as a focal point and technical lead for all (quantitative) structure–activity relationships ([Q]SAR) and read-across queries for product stewardship and regulatory purposes. A chemist and toxicologist by training, she started her career as a safety-evaluation scientist at Unilever before focusing her interests in computational toxicology and moving into a role that involved providing modeling and chemistry expertise for a variety of projects. While working for the (Q)SAR group at the European Commission’s Joint Research Centre, she was involved in many activities related to the development of technical guidance for Registration, Evaluation, Authorisation, and Restriction of Chemicals (REACH), including investigation of the feasibility of using computational approaches in the development of chemical categories, development and evaluation of (Q)SAR models for human health, and coordination of the technical development of software tools, such as Toxtree and Toxmatch. Dr. Patlewicz received her PhD in organic chemistry from the University of Santiago de Compostela in Spain.

David M. Reif is an associate professor of biological sciences at North Carolina State University and resident member of the Bioinformatics Research Center. His research focuses on the complex interactions between human health and the environment through the integrated analysis of high-dimensional data from diverse sources, including epidemiological studies of human health, high-throughput screening of environmental chemicals, and model organism data. Dr. Reif was previously a principal investigator with the US Environmental Protection Agency’s National Center for Computational Toxicology, where he led several statistical and bioinformatics efforts and collaborated on a variety of projects with federal, academic, and industry partners. Dr. Reif received his PhD in human genetics from Vanderbilt University.

John Wade is vice president and general manager of the Life Sciences Research Business in the National Security Division of Battelle. He is responsible for all Battelle’s animal-use laboratories and activities, in particular chemical–biological defense research, development, test, and evaluation and general toxicology test and evaluation services for both government and commercial customers. Dr. Wade has been a member-at-large of the NATO Human Factors and Medicine Panel as the US delegation’s chemical- and biological-defense expert since 2000. He also serves as the chairman emeritus of the National Defense Industrial Association’s Chemical and Biological Defense Division. Dr. Wade received his DVM from Michigan State University and his PhD in toxicology from the University of Kansas School of Medicine, and he was a diplomate of the American Board of Toxicology from 1991 to 2006.

Katrina Waters is the deputy division director for biological sciences at the Pacific Northwest National Laboratory. Her research interests are reconstruction of cell-response networks from integrated gene- and protein-expression data to enable predictive mechanistic modeling of disease and toxicity pathways. Dr. Waters serves on the US Food and Drug Administration National Center for Toxicological Research’s Science Advisory Board and the US Environmental Protec-

tion Agency's Scientific Advisory Panel on Methods for Prioritizing Endocrine Disrupting Chemicals. She is a member of the Society of Toxicology and adjunct faculty in the Department of Environmental and Molecular Toxicology at Oregon State University. Dr. Waters received a PhD in biochemistry from the University of Wisconsin.

Barbara Wetmore is a senior research investigator at The Hamner Institutes for Health Sciences. Her research focuses on integration of predictive modeling tools with high-throughput screening and other in vitro strategies to address issues of importance in chemical safety and risk assessment. Other research interests have been the application of genomic and proteomic tools to inform chemical mode-of-action assessments and biomarker discovery. She is vice-president-elect of the Society of Toxicology's In Vitro and Alternative Methods Specialty Section, and she has served as a study-section reviewer for the US Environmental Protection Agency and as an expert for the European Union Reference Laboratory for Alternatives to Animal Testing (EURL-ECVAM). Dr. Wetmore received her PhD in toxicology from North Carolina State University.

Yvonne Will is a senior director and the global science and technology lead for drug safety at Pfizer. Her research interests include mitochondrial, mechanistic, and cell-based toxicity assessment; drug-induced organ toxicities; and alternative in vitro and in vivo models. Dr. Will is the president of the Society of Toxicology's Drug Discovery Specialty Section and a member of the steering committee of the Technology Industry Consortium. She also serves on the editorial board of *Current Protocols in Toxicology* (John Wiley and Sons) and *Applied In Vitro Toxicology* (Mary Ann Liebert, Incy). Dr. Will received her PhD in biochemistry and biophysics from Oregon State University.

Appendix B

Available Data or Databases

There are many sources of acute-toxicity data. They include peer-reviewed literature and secondary sources, such as *The Merck Index* (O'Neil 2013). Several acute-toxicity databases can be easily searched by using chemical name, CAS Registry number, chemical structure, and other identifiers, and some can also be searched on the basis of the type of study, toxic effect, species, sex, dose, exposure duration, and route of exposure. Some databases will tabulate the results of a search. Comprehensive overviews of major acute-toxicity databases are available (Tsakovska et al. 2006; Lapenna et al. 2010). Brief overviews of a few are described here.

Several US federal agencies maintain databases of toxicity data. The public version of the US Environmental Protection Agency (EPA) Toxicity Reference Database (ToxRefDB) contains data from chronic, subchronic, developmental, and reproductive studies. It is also linked to the agency's ToxCast program. The EPA Aggregated Computational Toxicology Resource (AC-ToR) includes acute-toxicity data that are compiled from the Integrated Risk Information System (IRIS), Organisation for Economic Co-operation and Development (OECD) Summary reports, and Agency for Toxic Substances and Disease Registry documents. Unlike ToxRefDB, the AC-ToR database is not directly linked to acute-toxicity information on ToxCast or Tox21 chemicals. EPA also maintains the Toxic Substances Control Act Inventory and the SUPERLIST set of regulatory resources.

The National Library of Medicine (NLM) manages a network of databases called TOXNET®, which makes it possible to search for acute-toxicity information that is available in the Hazardous Substances Data Bank (HSDB). The most convenient means of accessing acute-toxicity information is through a chemical search that uses ChemIDplus, which has direct links to resources in NLM, other federal agencies, states, and scientific sites. The NLM databases contain records for more than 400,000 chemicals. NLM also maintains Web-site links at <http://sis.nlm.nih.gov/chem/alllocators.html> to other databases, such as the Canadian Domestic Substances List, the European Inventory of Existing Commercial Substances, the FDA Drugs@FDA system, and databases of the International Agency for Research on Cancer.

Several commercially available databases are also available. For example, Leadscope, Inc. markets a toxicity database that contains nearly 180,000 chemical structures and over 400,000 toxicity-study results derived from the US Food and Drug Administration Priority-based Assessment of Food Additives (PAFA) Database, the National Toxicology Program Chronic Database, the Registry of Toxic Effects of Chemical Substances (RTECS), and the DSSTox Carcinogenicity Potency Database (CPDB) (Leadscope 2012). Acute-toxicity data related to multiple exposure routes are available in the PAFA database and RTECS.

Several international databases are available. A multinational OECD database, eChemPortal¹, is a no-cost publicly available acute-toxicity database that can be searched by using a variety of chemical identifiers. One of its strengths is that it includes classification results based on the Globally Harmonized System of Classification and Labelling of Chemicals (GHS). The Europe-

¹See <http://www.echemportal.org/echemportal/substancesearch/substancesearchlink.action>.

an Chemicals Agency also manages an electronic database derived from Registration, Evaluation, Authorisation, and Restriction of Chemicals (REACH) registration dossiers for chemical substances manufactured or imported in Europe. The OECD QSAR Toolbox is a software tool that facilitates the development, evaluation, justification, and documentation of chemical categories for read-across. It contains regulatory inventories, toxicity data, and chemical-structure information that encode structure–activity relationship information.

REFERENCES

- Lapenna, S., M. Fuart-Gatnik, and A. Worth. 2010. Review of QSAR Models and Software Tools for Predicting Acute and Chronic Systemic Toxicity. EUR 24639. Luxembourg: Office of the European Union [online]. Available: http://publications.jrc.ec.europa.eu/repository/bitstream/JRC61930/eur_24639_en.pdf [accessed March 30, 2015].
- Leadscope. 2012. Comprehensive Source for Toxicity Data [online]. Available: http://www.leadscope.com/toxicity_database/ [accessed March 30, 2015].
- O’Neil, M.J. 2013. The Merck Index: An Encyclopedia of Chemicals, Drugs, and Biologicals, 15th Ed. Cambridge: Royal Society of Chemistry.
- Tsakovska, I., I. Lessigiarska, T. Netzeva, and A.P. Worth. 2006. Review of (Q)SARs for Mammalian Toxicity. EUR 22846 EN. Luxembourg: Office of the European Union.

