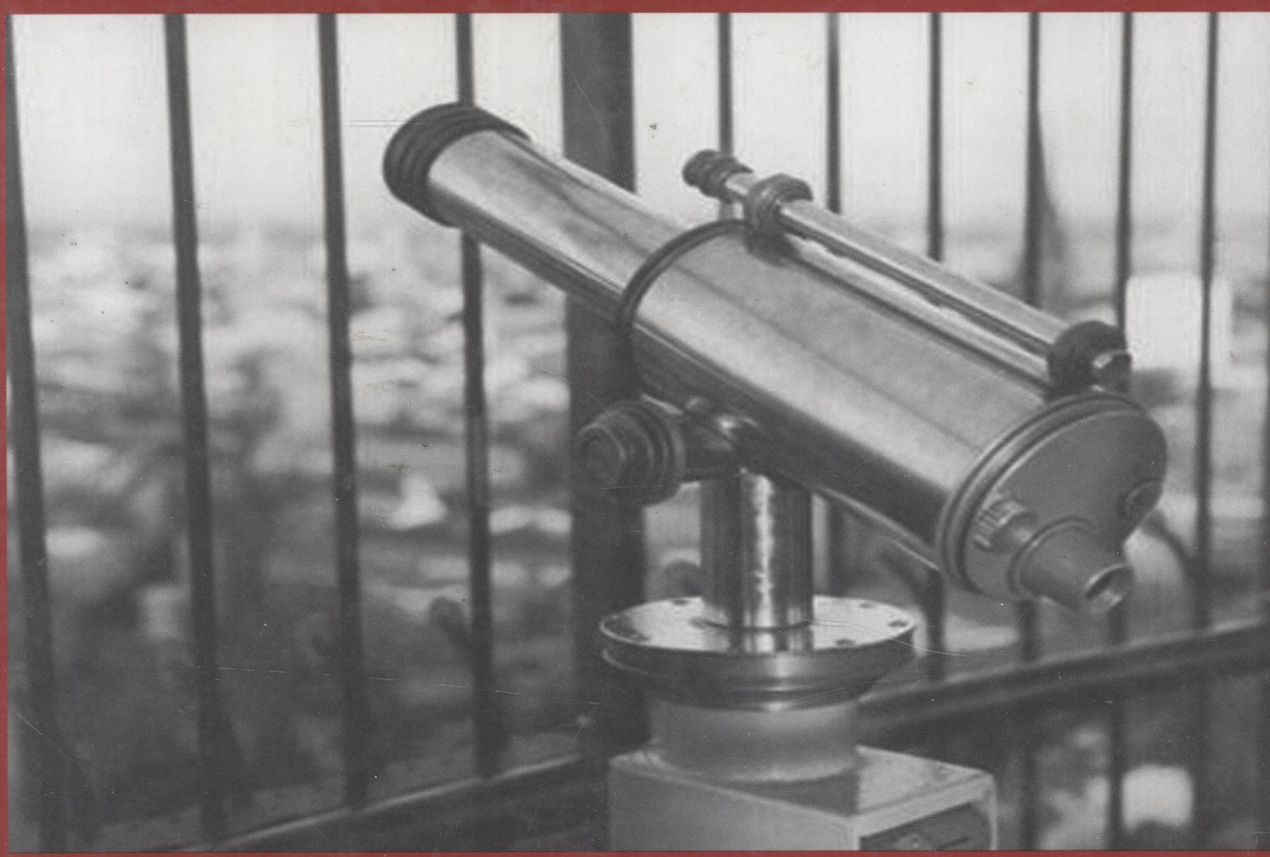
Mark T.D. Cronin
David J. Livingstone

Predicting Chemical Toxicity and Fate



CRC PRESS

Predicting Chemical Toxicity and Fate

Mark T.D. Cronin
Liverpool John Moores University
Liverpool, England

David J. Livingstone
University of Portsmouth
Portsmouth, England

CRC PRESS

Boca Raton London New York Washington, D.C.

Library of Congress Cataloging-in-Publication Data

Predicting chemical toxicity and fate / edited by Mark T.D. Cronin and David J. Livingstone.
p. cm.
Includes bibliographical references and index.
ISBN 0-415-27180-0 (alk. paper)
1. Molecular toxicology. 2. Toxicological chemistry. 3. QSAR (Biochemistry).
I. Cronin, Mark T.D. II. Livingstone, D. (David). III. Title.

RA1220.3.P74 2004
615.9—dc22 2004043999

Visit the CRC Press Web site at www.crcpress.com

International Standard Book Number 0-415-27180-0
Library of Congress Card Number 2004043999
Printed in the United States of America 1 2 3 4 5 6 7 8 9 0
Printed on acid-free paper

Predicting Chemical Toxicity and Fate

Dedications

From MC

To AMC and CCFC, for the pleasure and the pain.

Contributors

Romualdo Benigni
Istituto Superiore di Sanità
Rome, Italy

Jacob G. Bundy
School of Biological Sciences
University of Aberdeen
Aberdeen, Scotland

Colin D. Campbell
Macaulay Institute
Aberdeen, Scotland

Robert D. Combes
Fund for the Replacement of Animals in Medical Experiments
Russell & Burch House
Nottingham, England

Mark T.D. Cronin
School of Pharmacy and Chemistry
Liverpool John Moores University
Liverpool, England

John C. Dearden
School of Pharmacy and Chemistry
Liverpool John Moores University
Liverpool, England

Judith C. Duffy
School of Pharmacy and Chemistry
Liverpool John Moores University
Liverpool, England

Hong Fang
Logicon ROW Sciences
Jefferson, AR, U.S.A.

Huixiao Hong
Logicon ROW Sciences
Jefferson, AR, U.S.A.

Peter R. Fisk
Peter Fisk Associates
Whitstable, England

Klaus L.E. Kaiser
TerraBase, Inc.
Hamilton, Ontario, Canada

David J. Livingstone
ChemQuest, Isle of Wight, England
Centre for Molecular Design, University of Portsmouth, Portsmouth, U.K.

Helena Maciel
School of Biological Sciences
University of Aberdeen
Aberdeen, U.K.

Louise McLaughlin
Peter Fisk Associates
Whitstable, England

Monika Nendza
Analytisches Laboratorium
Luhnstedt, Germany

Tatiana I. Netzeva
School of Pharmacy and Chemistry
Liverpool John Moores University
Liverpool, U.K.

Graeme I. Paton
School of Biological Sciences
University of Aberdeen
Aberdeen, Scotland

Martin P. Payne
LHASA Ltd.
Department of Chemistry
University of Leeds
Leeds, England

Roger Perkins
Logicon ROW Sciences
Jefferson, AR, U.S.A.

Rosemary A. Rodford
SoloSTAR Ltd.
Bedford, England

T. Wayne Schultz
College of Veterinary Medicine
University of Tennessee
Knoxville, TN, U.S.A.

Gerrit Schüürmann
Department of Chemical Ecotoxicology
UFZ Centre for Environmental Research
Leipzig, Germany

Daniel M. Sheehan
Food and Drug Administration's National Center for Toxicological Research, Jefferson, AR, U.S.A.
Daniel M. Sheehan and Associates, Little Rock, AR, U.S.A.

Weida Tong
Food and Drug Administration's National Center for Toxicological Research
Jefferson, AR, U.S.A.

Cornelius J. Van Leeuwen
Institute for Health and Consumer Protection, Joint Research Centre
European Commission
Ispra, Italy

Gilman D. Veith
International QSAR Foundation to Reduce Animal Testing
Duluth, MN, U.S.A.

Rosalind J. Wildey
Peter Fisk Associates
Whitstable, England

Andrew P. Worth
European Center for the Validation of Alternative Methods
Institute for Health and Consumer Protection, Joint Research Center
European Commission
Ispra, Italy

Qian Xie
Logicon ROW Sciences
Jefferson, AR, U.S.A.

Foreword

When Corwin Hansch and Al Leo encouraged me in applying quantitative structure-activity relationships (QSARs) to the screening of environmental hazards, the U.S. Toxic Substances Control Act was still only a concept, and most QSAR calculations were still being made with a pencil. Their encouragement included two principles for QSAR along with a word of caution. The principles were that QSAR ought to be based on well-defined endpoints of intrinsic chemical activities as well as on molecular descriptors that could be interpreted mechanistically. The word of caution was that bureaucracies founded on laboratory testing, whether private or a regulatory agency, will only begrudgingly accept QSAR as a strategic tool in designing chemicals and managing chemical risks. Looking back over the last three decades, the Hansch/Leo principles for QSAR development have been largely ignored, if not disputed, by the growing QSAR community, with the possible exception in Europe where QSAR acceptance criteria will require transparency and a mechanistic foundation. Only the skepticism toward QSAR itself by our testing-oriented society seems to have been steadfast over three decades. The increasing costs of testing have produced renewed interest in more strategic *in silico* methods at a time when QSAR has been freed from many early computational barriers. Now more than ever, the scientific community needs an expert summary of QSAR methods like this book.

The guiding principles for QSAR development were intended to aid in the discovery of useful and robust models. The literature is replete with more than 10,000 QSAR correlations and models, yet few of them are useful enough to sway the skeptics. Still, progress in QSAR research can be measured by its own critics and the changing nature of their skepticism. The "*yes-but*" skeptics are particularly instructive to me. In 1974, our research plans faced the criticism, "*yes, QSAR may be able to predict some chemical properties, but it will never be able to predict bioaccumulation of chemical residues.*" In 1981, we faced, "*okay, QSAR may be able to predict bioconcentration potential, but it will never be able to predict toxicity.*" When the acute toxicity models appeared, we were confronted by "*yes, QSAR may be able to predict some ecotoxicity endpoints, but it will never predict chronic toxicity in mammals.*" Today, as the first mechanistic QSAR models are emerging for chronic reproductive effects and mutagenicity, this historical perspective on the QSAR skeptics serves as benchmarks for progress, if not encouragement.

Chemical reactivity in biological systems is far more complex than 20th century computational capabilities could have allowed one to address in quantitative terms. The rapid progress in computing power over the last decade enabled a steady stream of new computational methods in QSAR to emerge. Unfortunately, these new capabilities were not matched with the generation of high-quality biological databases needed to reveal systematic variation within heterogeneous chemical inventories. While many combinatorial problems in QSAR are likely to challenge computer sciences for years, present computer capabilities are sufficient to make future QSAR progress limited mostly by the databases for relevant, well-defined endpoints.

Our QSAR program at the Duluth, MN, U.S.A., laboratory focused on well-defined ecotoxicological endpoints that could be used directly in regulatory decisions. Our proof-of-concept paper in 1979 for estimating the bioconcentration potential required only a minimal database. Since then, many researchers have contributed to the evolution of bioaccumulation models and to extend them from simple screening-level methods for new chemicals to more exact estimates of tissue residues for risk assessments. In contrast to the bioconcentration database, the creation of the Duluth ecotoxicity database involved a multimillion dollar investment and dozens of scientists over most of a decade. Finding chemicals with common toxicity pathways to build mechanistic structure-toxicity relationships required better diagnostic bioassays, including behavioral symptomology (fish acute toxicity syndromes) and joint-toxicity studies. Our first paper on acute toxicity in 1983 was delayed almost 3 years due to rejections from toxicological journals based on our use of the term "narcosis" in describing reversible, baseline lethality — a criticism that lingers today in the health

research community. The dozens of more recent supporting papers on baseline toxicity and the even larger toxicity database created by Terry Schultz at the University of Tennessee (Knoxville) should be sufficient to overcome the skeptics of acute toxicity predictions so that the full attention of effects research can focus on important chronic toxicity endpoints.

The European Chemical Industry Council-led analysis of the state of QSAR in Setubal, Portugal (March 2002), concluded that QSARs for biodegradability were still the largest research gap in exposure research. Developing QSARs for important chemical properties progressed rapidly in the 1980s, but developing structure-biodegradability models has been paralyzed by a lack of systematic databases. Fortunately, in 1985 Hiroshi Tadakoro at the Hita laboratory in Japan recognized the need for a biodegradation database, and his team devoted more than a decade to systematically testing chemicals using activated sludge. Almost immediately after the Hita database was made available, the first QSAR screening models for biodegradability began to appear at scientific meetings. Again, these advances illustrate the importance of generating systematic data on crucial endpoints in the overall progress of predictive methods. Finding such endpoints and understanding how they can be reliably used in risk management is the central research challenge for QSAR. Once identified, QSAR progress seems to depend only on government funding to generate the systematic data needed to build acceptable QSARs for the respective endpoints.

The estimation of lethality and biodegradability directly from chemical structure has been one of the important first steps in applying QSAR to risk management. Shifting our focus to chronic effects and persistence of chemicals will require us to cross some exciting new frontiers, not the least of which will be the merger of metabolism and effects models as QSAR is incorporated into systems biology. To meet these challenges, scores of chronic toxicity pathways will have to be described, and "-omics" technology promises to open new doors in clustering chemicals by common toxicity pathways for QSAR modeling. With metabolic activation a critical step in many pathways, metabonomics offers unprecedented capability for identifying the key molecular initiating events for chronic effects, many being the new well-defined endpoints QSAR needs for chronic hazard identification. It is hoped that this book will play an important role in advancing QSAR in the face of healthy skepticism, and will bring greater attention to the need for high-quality data in strategic testing.

Dr. Gilman Veith
Duluth, MN

Preface

The motivation for this book was stimulated by a one-day meeting, "Modelling Environmental Fate and Toxicity," organized by the BioActive Sciences Group of the Society of Chemical Industry. The meeting was chaired by Drs. Mark Cronin and Dave Livingstone and held in London on March 27, 2001. The speakers at the meeting were drawn from industry and academia and described how computational methods could be applied to predict the toxicity and fate of chemicals in the environment. The meeting itself was well attended and was particularly timely. It coincided with an upsurge of interest in this area due both to legislative changes and the commercial possibilities of predicting toxicity and fate.

We are moving into a new era that is computationally rich and data poor. Modeling of toxicity is much easier than it was a decade ago because of increased computational power and greater availability of software to calculate descriptors of molecules (some of which is freely downloadable). However, we must never lose sight of the fact that good models require high quality input data, and preferably large amounts of it. Neither should we forget that predictive techniques are empirical models to be used; they should not be seen as an academic exercise. In commissioning this book we attempted to bring together a collection of chapters that would assist future modelers develop meaningful predictive techniques. This was always hoped to be a practical and didactic book, there are plenty of published reviews in all areas covered in the book. All authors were encouraged to make recommendations for the use of the methods and techniques described. The editors support the recommendations and hope they will be applied and useful to the next generation of modelers.

Mark Cronin and Dave Livingstone
July 2003

Acknowledgments

The editors wish to thank the BioActive Sciences Group of the Society of Chemical Industry (London, England) for the original opportunity to put on the one-day meeting that stimulated this volume. Without the group's foresight, encouragement, and organization, none of this would have been achievable. We also wish to thank the authors who have cheerfully contributed to the book, accepted our criticism, and made helpful comments. Finally we are extremely grateful to Taylor and Francis for originally commissioning the book and CRC Press for the final opportunity to publish it.

List of Abbreviations

χ	Randić branching index, or molecular connectivity
κ	Kappa shape index
σ	Hammett constant
μ	Dipole moment
Ψ	Wave function characterizing the state of a system state
AAR	Activity-Activity Relationship
ADME	Absorption, Distribution, Metabolism, and Excretion
AFP	α-Feto Protein
AM1	Austin Model 1
$\mathbf{A_{max}}$	Maximum acceptor superdelocalizability
ANN	Artificial Neural Network
AO	Atomic Orbital
AR	Androgen Receptor
ARI	Automated Rule-Induction
ATSDR	Agency for Toxic Substances and Disease Registry
BBB	Blood-Brain Barrier
BCF	Bioconcentration factor
BESS	Biodegradability Evaluation and Simulation System
BgVV	(German) Federal Institute for Health Protection of Consumers and Veterinary Medicine
BMD	BenchMark Dose
BMF	Biomagnification Factor
BRM	Carcinogenic potency in mice
BRR	Carcinogenic potency in rats
B3LYP	Hybrid density functional theory *ab initio* calculation method
c	Cluster (molecular connectivity)
C	Concentration (of a drug or toxicant)
C	Corrosive
CADD	Computer-Aided Drug Design
CAS	Chemical Abstract Service
CASE	Computer Automated Structure Evaluation
CCOHS	Canadian Center for Occupational Health and Safety
CDER	Center for Drug Evaluation and Research
Cl	Clearance
CM	Classification Model
CM1	Charge Model 1
CM2	Charge Model 2
CODESSA	COmprehensive DEscriptors for Structural and Statistical Analysis
CoMFA	Comparative Molecular Field Analysis
COMPACT	Computerized Optimized Parametric Analysis of Chemical Toxicity
COREPA	Common REactivity PAttern
CPSA	Charged Partial Surface Area
CRADA	Cooperative Research and Development Agreement
CT	Classification Tree
CV	Cross Validation
D	Distribution coefficient
DBP	Disinfection By Products
DEREK	Deductive Estimation of Risk from Existing Knowledge

DES	Diethylstilbestrol
DFT	Density Functional Theory
DSC	Differential Scanning Calorimetry
DSL	Domestic Substance List
DSS	Decision Support System
DSSTox	Distributed Structure-Searchable Toxicity
E	Hepatic extraction ratio
EA	Electron Affinity
$E_{1/2}$	Half-wave oxidation potential
ΔE	Difference in the energies of the highest occupied and lowest unoccupied molecular orbitals
EC	Extended connectivity
ECB	European Chemical Bureau
EC_{50}	Concentration causing 50% reduction in a specified effect
ECOSAR	Syracuse Research Corporation program to predict environmental toxicities
ECVAM	European Centre for the Validation of Alternative Methods
EDC	Endocrine Disrupting Chemical
EDKB	Endocrine Disruptor Knowledge Base
EDPSD	Endocrine Disruption Priority Setting Database
EDSTAC	Endocrine Disruptors Screening and Testing Advisory Committee
E_{HOMO}	Energy of the Highest Occupied Molecular Orbital
E_{kin}	Kinetic energy of a system
E_{LUMO}	Energy of the Lowest Unoccupied Molecular Orbital
EN	Electronegativity
EPIWIN	Estimations Programs Interface for Windows
E_{pot}	potential energy of a system
ER	Estrogen Receptor
e-state	Electrotopological state index
E_{tot}	Total energy of a system
EU	European Union
FDA	Food and Drug Administration
GI	Gastrointestinal
GST	Glutathione S-Transferase
FIRM	Formal Inference-based Recursive Modeling
H	Harary index
H	Hamilton operator
HD	Hardness
HENRYWIN	Syracuse Research Corporation program to predict Henry's law constant
HESI	Health and Environmental Sciences Institute
HF	Hartree-Fock
ΔHF	Heat of Formation
HPLC	High Performance Liquid Chromatography
HPV	High Production Volume
HPVC	High production volume chemical
HQSAR	Hologram Quantitative Structure-Activity Relationship
ICG_{50}	Concentration causing 50% inhibition of growth
ILSI	International Life Sciences Institute
IP	Ionization potential
Ipb	Isopropylbenzene
Ind	Induction
ITC	Interagency Testing Committee

IUCLID	International Uniform Chemicals Information Database
JME	Java Molecular Editor
JRC	Joint Research Centre of the European Commision
K	Partition coefficient
K_a	Equilibrium acid ionisation
KBS	Knowledge-Based Systems
K_i	Inhibition constant
K_m	Binding constant
K_{mxa}	Cuticular Matrix-Air partition coefficient
KNN	*K*-Nearest Neighbors
K_{oa}	Octanol-Air partition coefficient
K_{oc}	Soil-Water partition coefficient normalised for organic carbon content
K_{om}	Soil-Organic matter partition coefficient
K_{ow}	Octanol-Water partition coefficient
K'_{ow}	Apparent Octanol-Water partition coefficient
KOWWIN	Syracuse Research Corporation program to predict octanol-water partition coefficient
K_p	Skin permeability coefficients
K_{pa}	Plant-Air partition coefficient
K_{vs}	Vegetation-Soil partition coefficient
LCAO	Linear Combinations of Atomic Orbitals
LC_{50}	Lethal Concentration for 50% of animals
LD_{50}	Lethal Dose for 50% of animals
LFER	Linear Free Energy Relationship
LSER	Linear Solvation Energy Relationship
LNO	Leave-*N*-Out
Log	Logarithm to base 10
LOO	Leave-One Out
LRA	Linear Regression Analysis
MHBP	Molecular Hydrogen Bond Potential
MLP	Multilayer Perceptron
MLPot	Molecular Lipophilicity Potential
MLR	Multiple Linear Regression
MNDO	Modified Neglect of Diatomic Overlap
MO	Molecular Orbital
MOPAC	Molecular Orbital PACkage
MP	Melting Point
MPBPVP	Syracuse Research Corporation program to predict melting point, boiling point, and vapor pressure
MR	Molar Refractivity
MS-WHIM	Molecular-Surface Weighted Holistic Invariant Molecular
MultiCASE	Multiple Computer Automated Structure Evaluation
MW	Molecular Weight
NC	Non-corrosive
NCI	National Cancer Institute
NCTR	National Center for Toxicological Research
NDDO	Neglect of Diatomic Differential Overlap
NIOSH	National Institute for Occupational Safety and Health
NN	Neural Network
NOEC	No Observed Effect Concentration
NR	Nuclear Receptor

NTP	National Toxicology Program
OECD	Organization for Economic Co-operation and Development
OM1	Orthogonalization Model 1
OM2	Orthogonalization Model 2
OPPT	Office of Pollution Prevention and Toxics
OPS	Optimum Prediction Space
ORMUCS	Ordered MUlticategorical Classification method using the Simplex technique
p	Path (molecular connectivity)
P	Partition coefficient
PAH	PolyAromatic Hydrocarbon
$\mathbf{P_{alkb}}$	Promoter of the *alkb* gene
PBPK	Physiologically Based PharmacoKinetic
PBT	Persistent, Bioaccumulative, and toxic
pc	Path-cluster (molecular connectivity)
PC	Principal Component
PCA	Principal Component Analysis
PCB	PolyChlorinated Biphenyl
PCR	Regression on Principal Components
P-gp	P-Glycoprotein
pH	Negative logarithm of the hydrated proton concentration
PLS	Partial Least Squares
PM	Prediction Model
PM3	Parameterized Model 3
PM5	Parameterized Model 5
PMN	PreManufacture Notification
PNN	Probabilistic Neural Network
PNSA	Partial Negative Surface Area
PPAR	Peroxisome Proliferator Activated Receptor
PPSA	Partial Positive Surface Area
PSA	Polar Surface Area
$\mathbf{Q^2}$	Leave-one out or cross-validated R^2
$\boldsymbol{Q_A}$	Net atomic charge on atom A
$\mathbf{Q_h}$	Hepatic blood flow
QM	Quantum Mechanical
QSAR	Quantitative Structure-Activity Relationship
QSBR	Quantitative Structure-Biodegradability Relationship
QSPKR	Quantitative Structure-Pharmacokinetic Relationship
QSPR	Quantitative Structure-Property Relationship
R and R^2	Multiple correlation coefficient and its square
RBA	Relative Binding Affinity
REACH	Registration, Evaluation, and Authorization of CHemicals
rms	Root-mean-square
RTECS	Registry of Toxic Effects of Chemicals
SA	(Sub-)Structural Alerts
SAS	Statistical Analysis System
$\mathbf{S_{aq}}$	Aqueous solubility
SAR	Structure-Activity Relationship
SCF	Self-Consistent Field
SDF	Structure Data File
SHBG	Sex Hormone Binding Globulin
SMILES	Simplified Molecular Line Entry System

SRC	Syracuse Research Corporation
t ½	Half-life
TD_{50}	Dose required to halve the probability of animals remaining tumourless
TGD	Technical Guidance Document
TI	Topological Indices
TOPKAT	TOxicity Prediction by Komputer-Assisted Technology
TPSA	Topological Polar Surface Area
TSCA	Toxic Substances Control Act
UNIFAC	Uniquac Functional-group Activity Coefficient (where UNIQUAC = Universal Quasi-Chemical)
U.S. EPA	United States Environmental Protection Agency
V_d	Volume of distribution
V_m	Molecular Volume
vol	molar volume in $cm^3\ mol^{-1}$
W	Wiener index
WLN	Wiswesser Line Notation
WMPT	Waste Minimization Prioritization Tool
WSKOWIN	Syracuse Research Corporation program to predict water solubility

Table of Contents