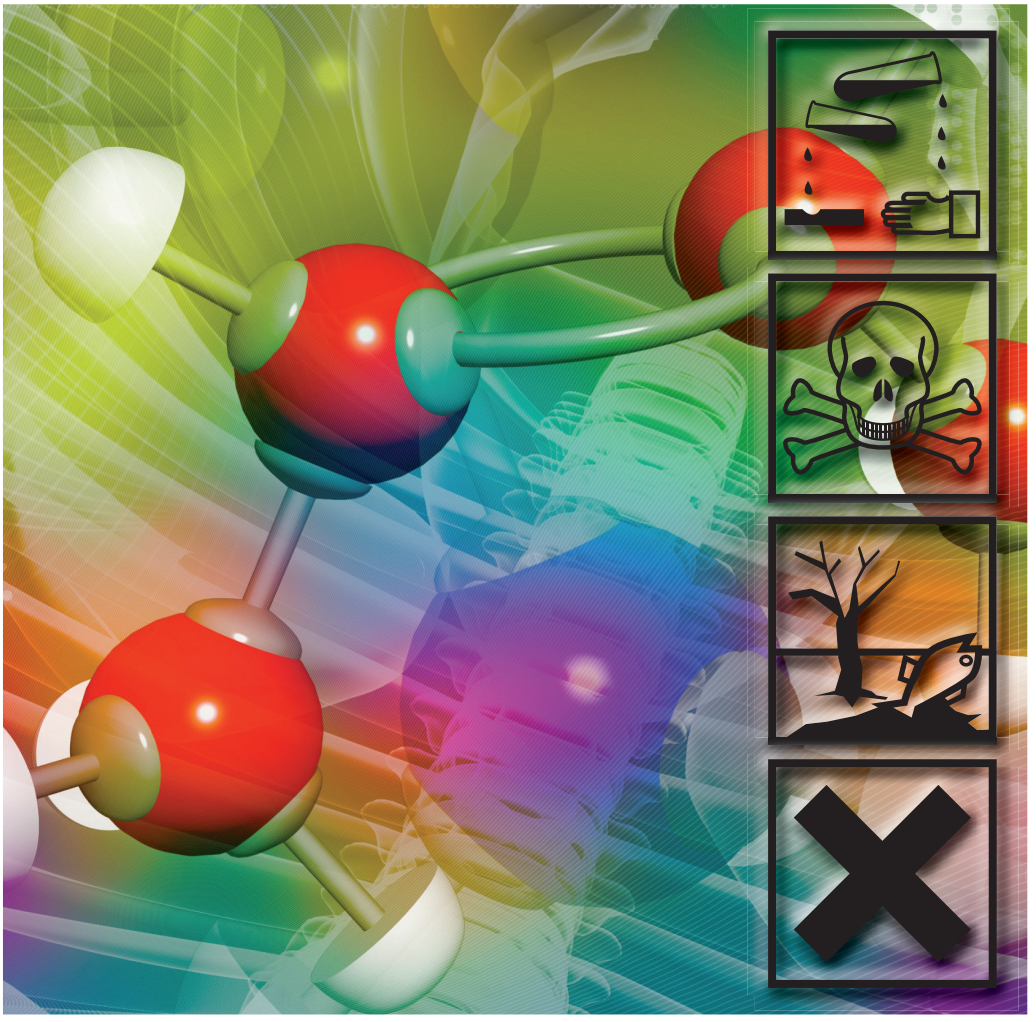
Issues in Toxicology

Edited by Mark T. D. Cronin and Judith C. Madden

In Silico Toxicology

Principles and Applications



RSC Publishing

In Silico Toxicology
Principles and Applications

Issues in Toxicology

Series Editors:
Professor Diana Anderson, *University of Bradford, UK*
Dr Michael D Waters, *Integrated Laboratory Systems, Inc, N Carolina, USA*
Dr Timothy C Marrs, *Edentox Associates, Kent, UK*

Titles in the Series:
1: Hair in Toxicology: An Important Bio-Monitor
2: Male-mediated Developmental Toxicity
3: Cytochrome P450: Role in the Metabolism and Toxicity of Drugs and other Xenobiotics
4: Bile Acids: Toxicology and Bioactivity
5: The Comet Assay in Toxicology
6: Silver in Healthcare
7: *In Silico* Toxicology: Principles and Applications

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

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

In Silico Toxicology
Principles and Applications

Edited by

Mark T. D. Cronin and Judith C. Madden
Liverpool John Moores University, Liverpool, UK

RSC Publishing

Issues in Toxicology No.7

ISBN: 978-1-84973-004-4
ISSN: 1757-7179

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

Published by The Royal Society of Chemistry,
Thomas Graham House, Science Park, Milton Road,
Cambridge CB4 0WF, UK

Registered Charity Number 207890

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

From JCM

To JGA and MMA

Preface

Over recent years, there have been great advances in the use of *in silico* tools to predict toxicity. Significant drivers for this include legislative demands for more information on chemicals to ensure their safe use and minimise toxicity to both human health and the environment, combined with an emphasis on the necessity to reduce and replace animal testing wherever possible.

The aim of this book was to provide a single 'from start to finish' guide to cover all aspects of developing and using *in silico* models in predictive toxicology. Hence the book covers all aspects of modelling from data collation and assessment of quality, to generating and interpreting physico-chemical descriptors, selecting the most appropriate statistical methods for analysis, consideration of applicability domain of the models and validation. Factors that may modulate toxicity such as external and internal exposure are presented and the use of expert systems, grouping and read-across approaches is discussed. *In silico* toxicology is becoming less of an academic exercise and more of a practical approach to filling the gaps in our knowledge concerning the toxicity of chemicals in use. The practical applications of these techniques in risk assessment are also considered in terms of their use in overall weight-of-evidence approaches and integrated testing strategies.

This book is intended to be a useful resource for those with experience in modelling who require more detail of the individual steps within the process, but should also provide the fundamental background information for those less experienced in *in silico* toxicology who wish to investigate the use of the techniques. Each chapter provides the reader with useful information on the individual aspects of the modelling approaches and relevant literature references are provided for those seeking more detail.

During the time of writing this book there are growing signs of change in toxicology and attitudes towards alternatives. This has been stimulated as much by the need to change in response to legislation as the vision to

Issues in Toxicology No.7
In Silico Toxicology: Principles and Applications
Edited by Mark T. D. Cronin and Judith C. Madden

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

incorporate the new technologies emerging in molecular biology, high throughput screening and toxicological pathways. The National Academy of Science's report, *Toxicity Testing in the 21st Century*, puts that vision into words, encapsulating the deficiencies of the current paradigm and the possibilities for a future without animal testing, but with real relevance for the effects of long term and low exposure of chemicals to humans. *In silico* techniques are a part of this process, whether it is in providing direct predictions of toxicity, rationalising the results, grouping chemicals with relevance to pathways or assessing the exposure to a chemical through toxicokinetic considerations. This book represents the state-of-the-art of *in silico* toxicology. We hope it is obvious to the reader that we have as many (if not more) tools and techniques as the toxicological data will support. What is needed in the near future is an illustration of how to use these *in silico* models. This includes integrating results and methods together to obtain greater certainty. Methods need to be developed to demonstrate the power of computational workflows and how they can provide estimates of the confidence in a prediction. *In silico* toxicology has an important role to play in toxicity testing in the 21st century; it must stand up and show the way to predicting effects.

Mark Cronin and Judith Madden

Contents

Issues in Toxicology No.7
In Silico Toxicology: Principles and Applications
Edited by Mark T. D. Cronin and Judith C. Madden

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

CHAPTER 1

In Silico *Toxicology*— *An Introduction*

M. T. D. CRONIN AND J. C. MADDEN

School of Pharmacy and Chemistry, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK

1.1 Introduction

Chemistry is a vital part of everyday life. In order for our interactions with chemicals to be safe, we must understand their properties. Traditional methods to determine the safety of chemicals are centred around toxicological assessment and testing, often using animals. There is, however, great interest and a need to develop alternatives to the traditional testing regime. Given the breadth and complexity of toxicological endpoints it is likely that, to ensure the safety of all chemicals, a variety of techniques will be required. This will require a paradigm shift in thinking, both in terms of acceptance of alternatives and the recognition that these alternatives will seldom be 'one for one' replacements.

In silico toxicology is viewed as one of the alternatives to animal testing. It is a broad term that is taken, in this book, to indicate a variety of computational techniques which relate the structure of a chemical to its toxicity or fate. The purpose of *in silico* toxicology is to provide techniques to retrieve relevant data and/or make predictions regarding the fate and effects of chemicals. In this sense the term '*in silico*' is used in the same manner as *in vitro* and *in vivo*, with '*silico*' relating to the computational nature of the work. There are, obviously, many advantages to *in silico* techniques, including their cost-effectiveness, speed compared with traditional testing, and reduction in animal use.

Issues in Toxicology No.7
In Silico Toxicology: Principles and Applications
Edited by Mark T. D. Cronin and Judith C. Madden

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

The science of *in silico* toxicology encompasses many techniques. These include:

- Use of existing data. If suitable data exist for a compound, there should be no requirement to initiate a new test or make a new prediction (unless prediction is for the purposes of model validation). If data are lacking for the chemical of interest, then other data can be used to develop (and subsequently evaluate) a new predictive model. Data sources include the ever increasing number of available databases as well as the open scientific literature. In addition, those working in industry may be able to utilise their own in-house data. More details on the retrieval and use of existing data are given in Chapter 3.
- Structure–activity relationships (SARs) are qualitative and can be used to demonstrate that a fragment of a molecule or a sub-structural feature is associated with a particular event. SARs become particularly powerful if they are formalised into structural alerts. A structural alert can be used to associate a particular toxicity endpoint with a specific molecular fragment such that, if the fragment is present in a new molecule, that molecule may elicit the same toxicity. The use of SARs and structural fragments is discussed in more detail in Chapters 8, 13, 16 and 19.
- There is a strong theme in this book towards forming groups of similar molecules. These groupings are also termed chemical categories. There are a number of approaches to 'categorise' a molecule including mechanistic profilers (structural alerts) and chemical similarity. Once a robust group of structures has been formed, it can be populated with toxicity data for those members of the group where experimental measurements are available. This allows for a read-across approach to be used to predict the toxicity of those members of the group for which no data are available. Various strategies for category formation and read across are discussed in Chapters 13–17.
- Quantitative structure–activity relationships (QSARs) provide a statistical relationship between the effects (toxicity and fate) of a chemical and its physico-chemical properties and structural characteristics. Linear regression analysis is often used but a variety of other multivariate statistical techniques are also used. The generation and use of QSARs are discussed in many chapters in this book.
- Expert systems (in the sense of *in silico* toxicology used in this book) are formalised and computerised software packages intended to make the use of SARs and QSARs easier. They usually provide an interface to enter a molecular structure and a suitable means of displaying the prediction (*i.e.* the result), in certain cases other supporting information is also given. Expert systems may be distributed on a commercial basis, although some are freely available. Increasingly they are integratable with other software. These types of software are described in more detail in Chapters 16, 17 and 19.
- Some *in silico* models can be derived to extrapolate the toxic effect measured in one species to predict toxicity in another species, thus reducing the requirement for further testing. This is discussed in more detail in Chapter 18.

- Models for other effects are increasingly becoming included under the remit of *in silico* toxicology. For example, a number of pieces of software can be used to estimate the likely exposure of an organism to a toxicant. Models exist for both external exposure (*i.e.* determining the amount present in the environment) and internal exposure (*i.e.* the amount taken up and distributed within an organism). These will not, themselves predict toxicity, but provide useful supplementary information for the overall risk assessment process. External and internal exposure scenarios are described in Chapters 20 and 21 respectively.

A variety of other tools and software also included within *in silico* toxicology are considered within this book. These include numerous applications for modelling the properties of chemicals (Chapters 5–8) defining the applicability domains of models (Chapter 12) and assisting with weight-of-evidence predictions (Chapter 22).

A note on terminology: it is increasingly common to see the use of the abbreviation (Q)SAR, indicating both SAR and QSAR. Where possible, the term (Q)SAR has been applied in this book and is intended to apply to both. A broad description of the use of alternatives to animal testing is provided by Hester and Harrison.[1] Alternatives can be generalised into *in vitro* tests and *in silico* approaches. In addition to activities such as optimising testing and reducing harm, these approaches form a framework within the 3Rs philosophy (Replacement, Refinement and Reduction) to replace animal tests. *In vitro* toxicology includes the use of cellular systems, -omics, *etc.* and while it will support the *in silico* approaches discussed in this book and is often referred to, it is not the main focus here.

1.2 Factors that Have Impacted on *In Silico* Toxicology: the Current State-of-the-Art

Much has been written on the history of QSAR and related techniques, and it is not the purpose to review much of it in this chapter.[2,3] However, it is worth considering some of the key factors to reach the current state-of-the-art. The initial factors listed below (Sections 1.2.1–1.2.4) are the drivers for the search of *in silico* toxicology; the remaining factors (Sections 1.2.5–1.2.7) have assisted in progress to the current state-of-the-art.

1.2.1 Environmental and Human Health

There is a need to ensure that any species exposed to a chemical is at minimal or no risk. The chemical cocktail to which man and environmental species is exposed potentially comprises a vast number of different substances, with more being added to that list annually. Whilst it is never the intention to allow exposure to a hazardous chemical at a concentration that may harm (with the exception of pesticides and pharmaceuticals *etc.*), for the vast majority of chemicals there is little knowledge regarding their effects. Traditional testing of

chemicals to assess toxicological properties has been heavily reliant on animal testing. Such tests require specialist facilities, are time-consuming and costly—even before animal welfare considerations are taken into account. It is widely acknowledged that to assess all the chemicals that are commonly used, animal testing will not solve the problem of ensuring that harmful chemicals are identified. Therefore, *in silico* alternatives that can make predictions from chemical structure alone have the potential to be very powerful tools.

1.2.2 Legislation

The manufacture and release of chemicals is carefully regulated across the world. The aim is to ensure safety through legislation. In addition, companies consider it a corporate responsibility to ensure the safety of their workers and consumers, and are highly aware of the possibility of litigation if they fail to do so. Each country and geographical region has a raft of legislation allowing the risk assessment and risk management of chemicals; it is beyond the scope of this volume to discuss this further and readers are referred to the excellent 'standard text' from van Leeuwen and Vermeire.[4]

No one single piece of legislation has promoted the use and development of *in silico* approaches. However, many have included it implicitly or explicitly. At the time of writing much work is being performed as a result of the European Union's Registration, Evaluation, Authorisation and restriction of Chemicals (REACH) Regulation, not to mention the Cosmetics Regulation. Elsewhere globally there has been similar legislation such as the Domestic Substances List in Canada and the Chemical Substances Control List in Japan.[5] Within all of these pieces of legislation there is the expectation that *in silico* toxicology will be applied. In each case new tools, methods and techniques have been developed as a direct response to the legislative requirements.

1.2.3 Commercial—Product Development

Linked to the needs to comply with chemicals legislation, businesses have long recognised the need to predict toxicological properties from structure. This provides many competitive advantages including possibilities to identify toxic compounds early on in the development pipeline, designing out toxicity in new molecules, registration of products with the use of fewer animals and hence at lower cost—in addition to the rationalisation of testing procedures. *In silico* approaches are broadly applied across many industries with a particular emphasis on the development of pharmaceuticals.[6] Therefore, there is a commercial need for reliable tools and approaches.

1.2.4 Societal

Not only does society desire safe chemicals, but it would prefer that animals were not used in the assessment of the properties of molecules. Therefore, there is commercial and consumer pressure to find and use alternatives. 'Computer models' are

often cited as a method to replace animal tests. There is an opportunity here to gain public support of this area of science. In so doing, this may improve the perception of what science can provide to society. As often happens, public opinion moves ahead of the science and what might be realistically achievable. Everyone reading this book and considering using these methods should be encouraged to promote their use and realistic expectations of what they can provide.

1.2.5 Commercial—Software

Sections 1.2.1–1.2.3 reveal a potentially huge marketplace for *in silico* approaches to predict toxicity. This has been realised by a number of companies which have developed commercial software systems; an overview of many of these is provided in Chapter 19 and elsewhere in this book. These companies have helped to raise the profile of *in silico* toxicology and make it more than an academic exercise. Many of the software products are now considered 'standard' and, whilst they cannot be considered the 'finished product' (the models will always need updating and refining), the commercial impact on *in silico* toxicology should never be underestimated. A competitive marketplace is being developed; a number of commercial companies are offering free 'taster' products with the assumption that the user may want more from the company. While often incomplete, these free products can provide invaluable training and educational tools, and allow the novice to familiarise themselves with the concepts and practice of *in silico* toxicology.

1.2.6 Computational—Hardware and Software

Anyone reading this book will appreciate the exceptional advances in computational power, software and networking capabilities in their lifetime. Both authors fondly remember performing early computational chemistry calculations by building the classic 'straw' model of a molecule which was manipulated manually to obtain a 'visual' optimum geometry before guessing at reasonable bond lengths and angles. The computational input file for the three-dimensional (3-D) structure was then written by hand allowing for optimisation of a single bond that was often an overnight calculation. Happy days indeed! The rapid progression of technology has, however, meant that calculations can be performed at previously unthought of rates and for vast inventories—millions of compounds very rapidly. *In silico* methods for computational chemistry and toxicology have been quick to embrace the new, affordable computational power and also use the internet to compile, organise and distribute information. Many of the tools described in this book would not have been possible ten or even five years ago.

1.2.7 New and Better Solutions to Complex (Toxicological) Problems

It is true to say that *in silico* models may be better able to predict certain endpoints than others. This is a result of a number of factors, in particular, the

Table 1.1 Summary of the different types of *in silico* model and software for the prediction of chemistry and toxicology.

In silico *tools and resources*	*Information retrieved or predicted*	*Description*	*Chapter in this book for further information*
Databases	Records of toxicological data and information (existing data rather than predictions)	Usually searchable by chemical identifier, substructure or similarity	3, 4
Calculation of physico-chemical properties (descriptors) to be used in models—may provide implicit information	Various physico-chemical properties (descriptors)	Fundamental properties such as log *P*, solubility, pK_a, *etc.*	5
Calculation of chemical structure-based properties—descriptors to be used in model	2-D properties	Various software calculates properties from 2-D structure, *e.g.* molecular connectivities	6
	Molecular orbital properties	Quantum chemical calculations requiring a 3-D optimised structure	7
Calculation of toxicological effects—direct prediction of toxicity	Structural alert based expert systems	Fragment systems relating a particular sub-structure in a molecule to a toxic effect	16, 19
	Multivariate and/or quantitative expert systems	Systems automating the QSAR approach to allow for seamless prediction of toxicity	19
	Grouping or category approaches	Formation of groups of molecules on the basis of rationally defined similarity	14, 15, 16, 17

Estimates of external exposure	Complex models for various exposure scenarios	May calculate potential for exposure from a variety of routes, *e.g.* inhalation, dermal, *etc.*	20
Predictions for internal exposure	Absorption, distribution, metabolism and elimination characteristics of chemicals	Assessment of factors that may modulate overall toxic potential by considering extent of internal exposure at site of action	21
Validation of models	Applicability domain definition	Various statistical methods for defining the domain of a (Q)SAR	12
	QSAR model reporting format	Automated methods to compile the details of a (Q)SAR suitable for validation	11
Integration of models	Pipelines for integration of models	Automated methods to link together algorithms into a predictive workflow	24

data available for modelling, the extent of knowledge concerning the mechanisms involved and the complexity of the endpoint. The current real challenge to find alternatives in toxicology is for the chronic, low-dose, long-term effects to mammals (and understanding the effects on man in particular).

The most difficult endpoints to address with alternatives include developmental toxicity and repeated dose toxicity. For these endpoints it is necessary to move away from the direct replacement dogma that has driven *in vitro* toxicology. There are likely to be many alternatives proposed, but one way forward is capturing the chemistry (*i.e.* structural attributes) of compounds associated with particular pathways which lead to toxicological events; more discussion is given in Chapter 14. In a related manner, this is (partially) the vision of the report, *Toxicity Testing in the 21st Century: A Vision and A Strategy*, published by the US National Research Council of the National Academies,[7] which has subsequently spawned the Tox21 collaboration between the US National Institute of Health (NIH) institutes and the US Environmental Protection Agency (EPA).[8]

1.3 Types of *In Silico* Models

There are many different types of models for computational chemistry and *in silico* toxicology. Table 1.1 shows a broad distinction between the types of models referred to in this book. As illustrated in Table 1.1, only some of the available software can be used to predict toxicity directly. Other models provide the building blocks to (Q)SAR formation and their application. How these may all fit together is described in more detail in Chapter 2.

1.4 Uses of *In Silico* Models

As suggested in Section 1.2 and Table 1.1, there are likely to be many potential uses of *in silico* models and related software. For the purposes of this book, these uses can be summarised in terms of which aspects of the life of the chemical or product they relate to. These are summarised in Table 1.2—but remember that this is only a small proportion of the possibilities for use.

With regards to the assessment of the toxicological properties of molecules, there will be much greater emphasis in the future on so-called Integrated Testing Strategies (ITS).[9] These are initially being used in response to legislative requirements, but future use will hopefully extend their application to provide realistic frameworks to replace animal tests. These are described in more detail in Chapter 23 and require a variety of *in silico* building blocks, *e.g.* use of existing databases, (Q)SAR predictions, *etc.* A part of ITS will undoubtedly be the combination of predictions from different methods.[10] This will be an important process, and whilst models are usually considered in isolation, it must be remembered that greater confidence will be obtained if all possible information is combined together.

Table 1.2 Uses of *in silico* methods to predict toxicity and properties through the various stages of chemical development.

Chemical life stage	*Use of* in silico *methods*
Development of substance, *i.e.* pre-patent or registration	*In silico* screening to eliminate potentially toxic compounds prior to synthesis
	Designing out of harmful features during development
	Prediction of properties to assist in formulation, *e.g.* solubility, melting point
New chemical, *i.e.* at registration	Assessment of toxicological profile or confirmation of animal tests
	Prediction of properties, *e.g.* log *P* required for registration
Existing chemical	Databases may be used to retrieve information on existing chemicals
	Prediction of toxic effects to assist in prioritisation of existing chemicals
	Grouping of compounds within inventories to assist in read-across

1.5 How to Use this Book

This brief chapter is not intended to do anything more than set the scene for the reader. It is anticipated that readers will be users and/or developers of models. Users of models will find the book a useful place to find basic definitions, obtain in-depth details of models and assess the different approaches for the *in silico* prediction of toxicity. Both novice and experienced modellers will find Chapter 2 the ideal starting place for tackling the problems of creating a meaningful workflow for *in silico* toxicology development. It is intended that this book will lead developers through the process of identifying relevant data, characterisation of molecules, development of significant (statistical) relationships, the interpretation and documentation for regulatory purposes—and allow them to be able to place the models in the context of current knowledge.

1.6 Acknowledgements

Funding from the European Union 6th Framework Programme OSIRIS Integrated Project (GOCE-037017-OSIRIS) and the ReProTect Integrated Project (LSHB-CT-2004-503257) is gratefully acknowledged.

References

1. R. E. Hester and R. M. Harrison, *Alternatives to Animal Testing,* Royal Society of Chemistry, Cambridge, 2006.
2. M. T. D. Cronin, in *Predicting Chemical Toxicity and Fate*, ed. M. T. D. Cronin and D. J. Livingstone, CRC Press, Boca Raton, FL, 2004, pp. 3–13.

3. M. T. D. Cronin, in *Recent Advances in QSAR Studies: Methods and Applications*, ed. T. Puzyn, J. Leszczynski and M. T. D. Cronin, Springer, Dordrecht, The Netherlands, 2010, pp. 3–11.
4. C. J. van Leeuwen and T. G. Vermeire, *Risk Assessment of Chemicals: An Introduction,* Springer, Dordrecht, The Netherlands, 2007.
5. A. P. Worth, in *Recent Advances in QSAR Studies: Methods and Applications*, ed. T. Puzyn, J. Leszczynski and M. T. D. Cronin, Springer, Dordrecht, The Netherlands, 2010, pp. 367–382.
6. S. Boyer, *Altern. Lab. Anim.*, 2009, **37**, 467.
7. National Research Council of the National Academies (NRC), *Toxicity Testing in the 21st Century: A Vision and a Strategy*, National Academies Press, Washington DC, 2007.
8. F. S. Collins, G. M. Gray and J. R. Bucher, *Science*, 2008, **319**, 907.
9. C. Grindon, R. Combes, M. T. D. Cronin, D. W. Roberts and J. F. Garrod, *Altern. Lab. Anim.*, 2008, **36**(Suppl 1), 7.
10. C. M. Ellison, J. C. Madden, P. Judson and M. T. D. Cronin, *Mol. Inform.*, 2010, **29**, 97.

CHAPTER 2

Introduction to QSAR and Other In Silico *Methods to Predict Toxicity*

J. C. MADDEN

School of Pharmacy and Chemistry, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK

2.1 Introduction

Scientific analysis of the world around us is based on collecting information on what is known, or can be measured, and structuring the information to enable investigation of why systems behave the way in which they do. The information must be organised into a framework from which the relationships between the different aspects of complex systems can be determined and how they interact to produce the effects observed. Scientifically, this current knowledge is used to create theories or models from which predictions of unknown phenomena can be made. As scientific knowledge advances and more information becomes available, this can be used to test the theories or models that have been built.

In silico prediction of toxicity is based on such scientific principles. Initially, information is gathered from previous observations such as collation of measured toxicities of a group of chemicals. The properties of these chemicals are investigated to establish which features are responsible for their toxic activity, *i.e.* to determine the relationship between the specific molecular properties of the compound and its associated toxicity. This information can then be used to build models that can explain why a given compound does (or does not) elicit a

Issues in Toxicology No.7
In Silico Toxicology: Principles and Applications
Edited by Mark T. D. Cronin and Judith C. Madden

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

particular effect and to predict the effects likely to be elicited by compounds for which the measured data are not available. As more data become available, the validity of the models can be tested and adjustments made as necessary. This is an iterative process allowing for continual refinement of predictive models.

In silico models use computational methods to predict activity of compounds based on knowledge of their chemical structure and selected properties. The properties themselves (*e.g.* physico-chemical or structural properties) may be computationally calculated using a range of software, or determined experimentally. *In silico* methods include (quantitative) structure–activity relationships [(Q)SARs], expert systems, grouping and read-across techniques. Each of these techniques is introduced below and expanded upon in subsequent chapters.

The basic tenet of quantitative structure–activity relationship (QSAR) analysis is that a given biological activity can be correlated with the physicochemical properties of a compound using a quantitative mathematical relationship. QSAR (and other *in silico* methods to predict toxicity) are powerful and highly attractive tools in science. In part this is due to the diversity of the knowledge base the techniques employ, drawing together information from several disciplines. There are the chemical and physical properties of compounds to consider in addition to knowledge of the chemical, physiological or toxicological mechanisms that give rise to the effects.

There is an ever increasing array of techniques available from which to build models employing the advances in mathematical and computational sciences. Overall, the number of models available is rapidly expanding. All models are, by definition, surrogates for real systems. The usefulness of models is determined by the extent to which they offer a greater understanding of the system and enable predictions beyond current knowledge.

In silico methods in toxicology provide a framework for combining knowledge from several disciplines, offer mechanistic insight into chemical and biological processes, help to identify anomalous observations and promote savings in time, money and animal use where estimations of toxicity are required.

The philosophy of this book is to guide readers through the processes involved in generating and using *in silico* techniques to make predictions for toxicity. The aim of this chapter is to provide an overview of how the different sections of the book link together to enable such predictions to be made. This chapter serves as an overall introduction to QSAR and *in silico* techniques, outlining how to go about generating and using the models. This general overview is supplemented by subsequent chapters which provide a more detailed analysis of each individual step in the model building process. This chapter focuses on how to develop a QSAR for a toxicological endpoint. However, the methods described are equally applicable to developing QSARs for other endpoints such as predicting drug activity or pharmacokinetic/toxicokinetic properties. The use of other *in silico* techniques (*e.g.* category formation and read-across) are also introduced in this chapter.

2.1.1 Fundamentals of QSAR

The origins of QSAR date back to the 19th century when researchers including Cros,[1] Crum Brown and Fraser,[2] and Richardson[3] all identified a relationship between the activity of a compound and its chemical properties. However, it was the pioneering work of Hansch *et al.*[4] which is most often quoted as the beginning of modern QSAR. Equation 2.1[4] encapsulates the philosophy of QSAR, *i.e.* that a given biological activity can be correlated with the physico-chemical properties of a compound using a quantitative mathematical relationship.

$$\log 1/\mathrm{C} = 4.08\pi - 2.14\pi^2 + 2.78\sigma + 3.36 \qquad (2.1)$$

[statistics not given]
where:

C is the concentration to produce a herbicidal effect
π is an indicator of hydrophobicity
σ a measure of electronic effects within the molecule (discussed below).

Many elegant descriptions concerning the chronological development of this field have been published in the literature.[5,6] Readers are referred to these articles for a comprehensive review of QSAR from a historical perspective. Here the emphasis is on the current state-of-the-art, with comment on the future of QSAR and its potential utility in predictive toxicology.

QSAR and other *in silico* techniques have been widely used by the drug industry for many years. However, new European legislation such as the REACH Regulation[7] and the Cosmetics Directive[8] have led to increased interest in these methods. This is because the legislation promotes the use of alternatives to using laboratory animals to provide estimates of toxicity for risk assessment purposes. QSAR enables the relationship between activity (toxicity) and physico-chemical properties of molecules to be determined (Figure 2.1).

Although QSAR can be applied to diverse areas of science covering a range of endpoints (drug activity, pharmacokinetics/toxicokinetics, pesticide toxicity,

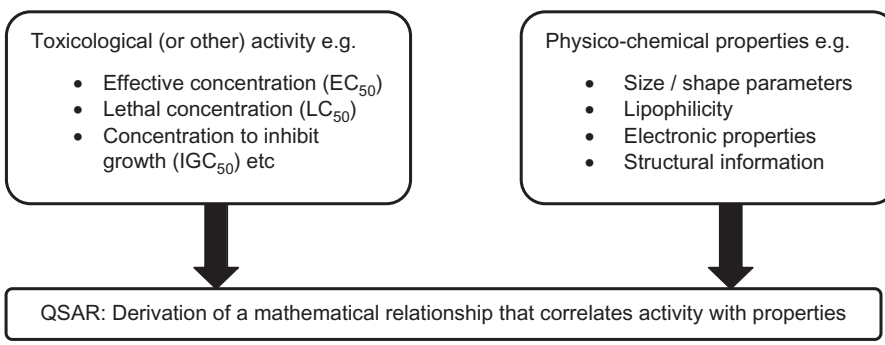


Figure 2.1 Overview of QSAR.

fragrance, *etc*.), the fundamental approach to developing, validating and using QSARs is relatively consistent. A flow diagram indicating the steps involved in this process is shown in Figure 2.2.

While this chapter is mostly concerned with the steps involved in generating a new *in silico* model, existing models such as published (Q)SARs or expert

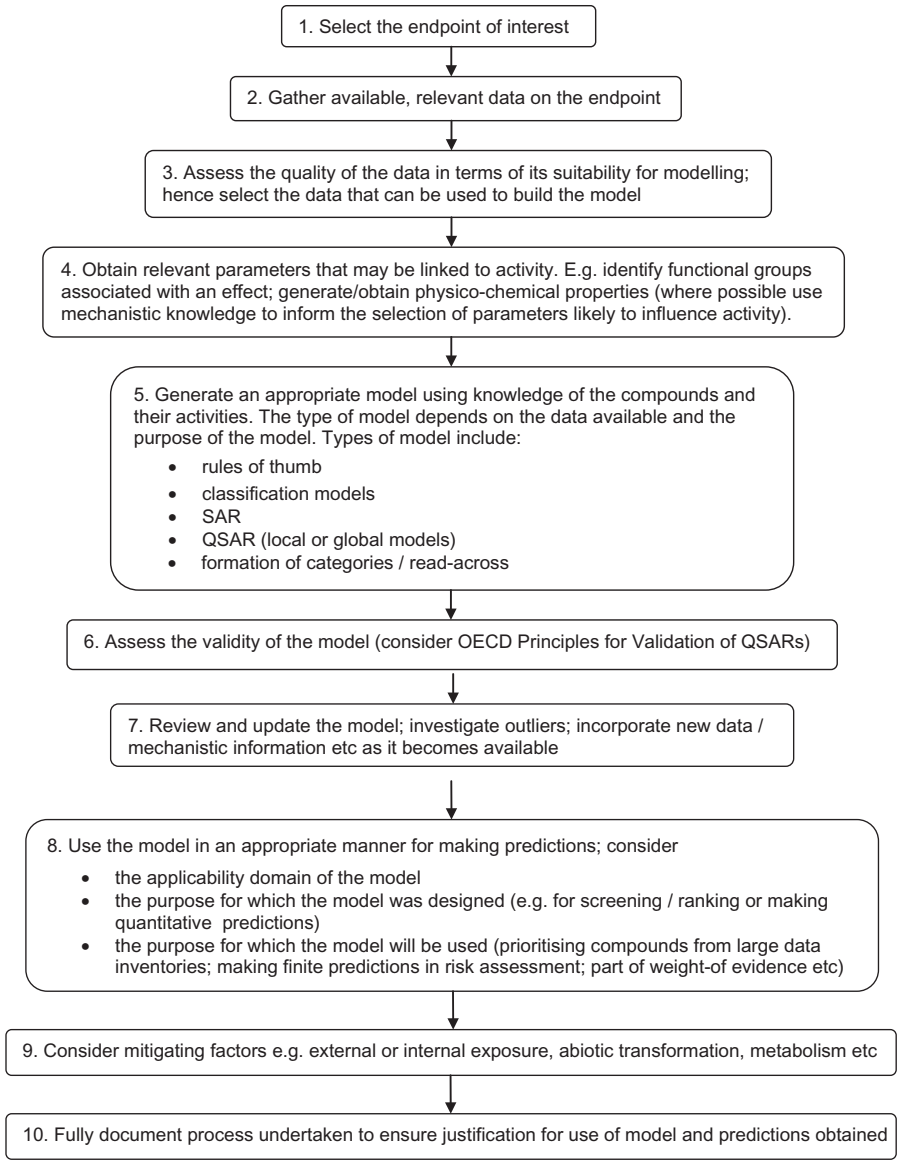


Figure 2.2 Steps involved in generating and using *in silico* predictions.

systems (see Chapter 19) can also provide useful information in terms of predicting activity, elucidating mechanisms or confirming the validity of newer models.

2.2 Building an *In Silico* Model

2.2.1 Step 1: Selecting the Endpoint to Model

Selecting the endpoint for which a model is to be built determines the nature of the data required to build the model. Although choosing the endpoint may appear trivial at first, it may not be a trivial issue in practice. In some cases the endpoint may be obvious, *e.g.* a mouse oral LD_{50} measured at a specific time point. However, consider building a model for toxicity to fish, here a 96-hour LD_{50} value measured in guppy is a different endpoint to a 96-hour LD_{50} in fathead minnow. The question is this: is the endpoint of interest (and hence the model to be built) toxicity to all fish species, an individual species or the most sensitive species? For *Daphnia* toxicity, should a model to be generated for LD_{50} at 24 hours or 48 hours, or can data from either endpoint be combined into a single endpoint for LD_{50}? In Ames testing for mutagenic effects, is the endpoint mutagenic activity in the absence or presence of metabolising enzymes? If enzymes are present, is this in the form of microsomes or the S9 fraction? Also consider the case of the human health endpoint skin sensitisation: to build a model for skin sensitisation there are several endpoints on which data may be gathered including the guinea pig maximisation test, the occluded patch (Beuhler test), Freund's complete adjuvant test, mouse ear swelling test, local lymph node assay, human repeat insult patch test, *etc*. The modeller needs to determine which is the most relevant endpoint for their purpose. In many cases, availability of data will influence the endpoint selected; however, it is important to have a clear idea from the outset which is the specific endpoint to be modelled and hence which data are relevant.

2.2.2 Step 2: Gathering the Data

Once the relevant endpoint has been identified, the next step is to gather available data for that endpoint. There are many ways in which data can be gathered for modelling purposes. Information can be retrieved based on searches for specific compounds of interest or for endpoints of interest. Sources of data include both in-house and publicly available resources. In-house data resources should be considered where available as these have several advantages:

- details of specific experimental protocols and associated metadata can be obtained and verified;
- the causes of anomalies in results are easier to investigate;
- the chemical space of in-house data sets is more likely to be representative of the chemical space of the compounds for which a prediction is sought.

Publicly available sources of data include:

- primary literature (for individual compounds and data sets);
- internet–based resources: ChemSpider;[9] ChemIDplus Advanced;[10] and the Cheminformatics and QSAR Society's web pages;[11]
- completed and ongoing efforts within EU projects such as CAESAR[12] and OSIRIS,[13] which are releasing data sets that have been highly curated;
- global (meta) portals such as ACToR[14] and the Organisation for Economic Co-operation and Development (OECD) eChem portal.[15]

These are only a few examples of the extensive resources available; further information on data resources is given in Chapter 3.

Once an endpoint has been selected and the search for data has begun, availability of data should guide the refinement of the search criteria. It may be pragmatic to expand the search criteria where data are scarce (*e.g.* to include more species within the same taxa, additional time points or related effects) or to limit the search criteria where data are plentiful (*e.g.* to a single species, study duration or effect). The more stringent the criteria for including data in the model, the more reliable the model should be as sources of variability are reduced.

As discussed in Chapter 3, there is a fine balance between making search criteria so broad as to introduce too much variability and too narrow to obtain a reasonably sized data set. Chapter 3 provides a detailed review of available resources, along with recommendations for using these and example case studies in obtaining data for different purposes. When starting to search for toxicity (or other) data, readers are referred to Chapter 3 for guidance in how to go about finding relevant data. Note, however, this is a rapidly evolving area and new sources of data with user-friendly interfaces are continually being developed.

When acquiring data it is essential to capture all the relevant information and ensure that the data storage system is reliable and portable. In many instances, a simple spreadsheet (e.g. Microsoft® Excel or Accelrys® Accord for Excel) will suffice and allows the flexibility to store required information that can be read directly into a range of statistical analysis or model development software packages.

It is essential to ensure that all transcriptions of data are correct as any errors introduced in the data gathering stage will later translate into inaccurate models being developed. It is crucial to ensure that the structure itself is correct and that the data correlate to that specific compound. The validity and accuracy of the SMILES string/CAS number should be checked rigorously before storing any associated data. Chapter 4 discusses, in more detail, issues to ensure accuracy in data collection.

Gathering data can potentially be a very long process. A pragmatic approach must therefore be taken to ensure a reasonable number of compounds and their associated data are collected for analysis without spending an excessive amount of time tracking down every last compound, as is often the case the law of diminishing returns applies!

2.2.3 Step 3: Assessing the Quality of the Data

Once the data have been acquired, the quality of the data and their suitability for the purpose of modelling needs to be assessed before their inclusion in the model building process can be justified.

Important issues in assessing the quality and/or suitability of data include:

- Can the data be unequivocally associated with the given compound? Potential problems here include:
  - incorrect nomenclature, CAS numbers or chemical structures;
  - the use of salts in place of parent compounds;
  - tautomeric and isomeric forms of the compound; and
  - questions concerning purity of the tested compound.
- Has the stated quantity or concentration of the compound been verified? Issues here include limited solubility in the test system or volatility of the compound resulting in the actual amount present being lower than the nominal (stated) concentration. Passive dosing systems or monitoring of concentration throughout the time course of the experiment may provide greater confidence in results.
- Are the data from a reliable source? For example was the result produced in a laboratory with acknowledged expertise in the area by appropriately trained staff performing work to standards commensurate with Good Laboratory Practice?

Formal scoring methods (*e.g.* Klimisch criteria)[16] for assessing data quality are available and serve as a useful guide as to whether or not the data are ideal, reasonable or should only be used with caution.

Chapter 4 deals with these issues and other aspects of assessing data quality in much greater detail. Only data whose inclusion can be justified should be incorporated into building or testing of models.

2.2.4 Step 4: Obtaining Parameters Potentially Related to the Activity

The purpose of building a model for use in *in silico* prediction of activity (toxicity) is to derive a relationship between the properties of the compounds and their biological effects. Thus, if the relevant properties of a new compound can be determined its biological effect can be predicted from these. (Note that the 'properties' of molecules may be variously referred to as 'parameters', 'variables' or 'descriptors'.)

Key to building a useful model is determining which properties of the compounds are linked causally to their effects. While it is possible to obtain properties from the literature, computationally generate or measure *de novo* thousands of properties of molecules, the majority will not have a causal relationship with effect. In some cases no correlation between the property and biological effect will be demonstrated; in other cases, a superficial relationship

may be determined but this apparent relationship will not withstand rigorous statistical investigation. Statistical analysis and the problems of spurious correlations are discussed further in Chapters 9 and 10.

In building a model it is more useful to incorporate only those properties of the compounds for which a rationale for the relationship with effect can be justified. Notwithstanding, in certain cases, so little is understood about the mechanisms involved in the processes that it may be necessary to resort to the inclusion of less readily defensible molecular properties. There are a myriad of one-, two- and three-dimensional properties that can be obtained for individual chemicals. These encompass steric, electronic, hydrophobic and topological properties in addition to composite parameters produced from combinations of such properties, e.g. the electrotopological indices which unite electronic and topological effects.[17] Properties may relate to the whole molecule or, where structural analogues are investigated, substituent parameters may be employed. Table 2.1 provides an overview of some of the key molecular properties of compounds encountered in *in silico* models for predicting activity and a rationale for their inclusion in models.

Chemical structure encodes all possible information of the nature of a chemical and how it will interact from whether it will be a solid, liquid or gas at room temperature and pressure to how it may interact with biologically relevant molecules such as enzymes or DNA. Although all of this information exists within the chemical structure, our understanding of these systems and ability to interpret the information is less than perfect.

In certain cases it is the presence of a particular structural feature or functional group that influences or determines activity; hence identifying these features may serve as useful 'parameters' in model building. Simple physicochemical properties such as the logarithm of the octanol/water partition coefficient (log *P*) or logarithm of the aqueous solubility (log *Saq*), *etc*. can be obtained from the literature, measured directly or estimated using a range of software. The ability of a molecule to elicit a biological effect may be correlated simply to such a property; for example, the narcotic effect of alcohols can be related to their hydrophobicity as it involves a non-specific interaction perturbing the cell membrane. Chapter 5 provides more detail on simple physicochemical properties and their measurement.

Two-dimensional (2-D) properties relating to the chemical's structure (*i.e*. those determined from a graphical representation of the connectivity of atoms within molecules) are discussed in Chapter 6. These allow for shape characteristics, in terms of molecular topology, to be considered which can influence the ability to reach a target site or to fit to a receptor.

Chapter 7 provides an in-depth analysis of electronic properties derived from molecular orbital calculations. These are of particular significance for more reactive compounds as they provide a quantitative indication of the reactivity of chemicals and hence determine the likelihood and degree of interaction with biologically important macromolecules such as DNA and proteins. The ability of a molecule to elicit a biological effect may be related to relatively simple physico-chemical properties. However, many biological effects are the result of specific binding to individual receptors or biological targets (*e.g*. the binding of

Table 2.1 Example descriptors used in (Q)SAR and the rationale for their inclusion in models.

Descriptor	*Definition of descriptor and rationale for inclusion*
Functional groups/ structural alerts	Certain toxicities may be associated with specific structural features, *e.g.* presence of nitroso groups associated with carcinogenic potential or Michael type acceptors associated with skin sensitisation.
Similarity indices or similarity scores	Based on the concept that 'similar' molecules will produce 'similar' effects. There are many ways in which 'similarity' may be defined (*e.g.* similarity of size, shape, spatial distribution of key atoms/functional groups, reactive potential, *etc.*). What makes one molecule 'similar' to another is a debatable issue.
Indicator variables	These indicate the presence or absence (usually denoted by 1 or 0 respectively) of specific structural features (*e.g.* hydrogen bond donating/accepting groups, presence of particular functional group, *etc.*).
Hydrophobic/hydrophilic descriptors	Indicate solubility in aqueous and/or organic medium and relative partitioning between phases. Descriptors may correlate with the ability of compounds to cross biological membranes, accumulate within biophases or relate to specific hydrophobic interactions within receptor sites.
Log P	Logarithm of the partition coefficient, *i.e.* the ratio of concentrations of a compound between an organic and an aqueous phase (usually 1-octanol/water).
Log D	Logarithm of the distribution coefficient (or apparent partition coefficient). Derived from log P but taking into account the partitioning of ionised or associated species.
Log k'	Logarithm of the high performance liquid chromatography (HPLC) capacity factor. Retention on an HPLC column is correlated with log P. It can be determined more rapidly and can be applied to compounds for which standard methods of measuring log P are not appropriate.
Log S_{aq}	Logarithm of the aqueous solubility.
Lipole	Distribution of lipophilicity within a substituent or whole molecule.
Molecular lipophilicity potential (MLP)	Geometric distribution of lipophilicity within a molecule.
π	Substituent constant indicating the influence of individual substituents on overall partitioning behaviour: $\pi = \log P_{(\text{substituted derivative})} - \log P_{(\text{parent})}$.
Electronic descriptors	These represent diverse properties which are associated with many effects. Examples include the ability to cross biological membranes or bind to macromolecules (correlated with hydrogen bond donating/accepting ability or dipole interactions) and chemical reactivity associated with (covalent) binding that may elicit DNA damage or immune

Table 2.1 (*Continued*)

Descriptor	*Definition of descriptor and rationale for inclusion*
	responses (correlated with electronegativity, E_{HOMO}, E_{LUMO}, *etc.*).
HD/HA	Hydrogen bond donating/accepting ability. This may be represented as an indicator variable for ability or lack of ability to hydrogen bond or be quantified as to strength of bonding ability.
E_{HOMO}	Energy of the highest occupied molecular orbital (negative of the ionisation potential).
E_{LUMO}	Energy of the lowest unoccupied molecular orbital, (a measure of the ability to accept electrons).
Electronegativity (x)	Ability of an atom (or group) to attract electrons, associated with reactivity.
Electrophilicity (ω)	Models the ability of a molecule to accept electron density, associated with reactivity.
Superdelocalisability	A measure of reactivity determined from: $\sum$(charges of atoms in molecular oribtial) $\div$(energy of each molecular orbital).
Atomic charge (q_n)	Charge associated with atom 'n'.
Dipole moment (δ)	Distribution of charge within a molecule.
σ	The Hammett substituent constant, indicating the electron directing effects of aromatic substituents (positive for electron attracting and negative for electron donating groups).
Steric descriptors	Associated with ability to reach target site (*e.g.* be absorbed across relevant biological membranes) and fit within specific receptors.
Molecular weight	Relative molecular mass indicating general size of the molecule.
Molecular volume	This may be calculated using the sum of van der Waals atomic volumes; indicates general size.
Molecular surface area/solvent accessible surface area	Computationally, a probe molecule can be 'rolled' over the surface of a molecule to determine the area that is accessible (to solvents or interacting molecules such as receptors).
K	The kappa index is a shape parameter based on the degree of branching of the molecular graph.
Sterimol (L_1, B_1-B_5)	Shape descriptors that indicate the length (L) of a substituent and its widths in different directions (B_1-B_5).
Es	The Taft steric constant indicates the size contribution of substituents on a parent molecule.
Topological descriptors	These are based on graph theory and relate to overall topology, dictated by the way in which atoms are connected to each other. Whilst many studies relate topology to molecular properties, their use remains controversial due to the difficulty in interpreting some of these parameters.
${}^{\mathrm{n}}\chi$	N^{th} order connectivity index. Zero order connectivity is obtained by counting the number of non-hydrogen links to each atom and taking the reciprocal square

Table 2.1 (*Continued*)

Descriptor	*Definition of descriptor and rationale for inclusion*
	root. Higher order is obtained by summing across different numbers of bonds.
${}^{n}\chi^{v}$	Valence corrected connectivity indices are used to distinguish between heteroatoms.
3-D descriptors	Three-dimensional representation of molecules provides a more accurate description of molecular dimensionality. Such parameters are important in terms of receptor fit and binding.
Composite parameters	These represent combination effects and can provide additional information reflective of more than one feature.
Polar surface area; hydrophobic surface area	Dividing the surface area of a molecule into regions of polarity or hydrophobicity can provide useful information for example in terms of specific receptor binding interactions.
Electrotopological state indices (*S*)	A combination of electronic features and topological environment for given atoms.

drugs to enzymes inhibiting their activity or the binding of an oestrogen-mimic to the oestrogen receptor, associated with endocrine disrupting effects). These processes require specific interaction between a chemical and its target macromolecule. The correct orientation of the atoms and the distribution of their electronic and hydrophobic features in space can only be accurately determined by consideration of three-dimensional (3-D) descriptors. Three-dimensional interactions between chemicals and biological targets are considered further in Chapter 8.

Clearly, it is beyond the scope of a single chapter to describe all potential molecular properties of interest. Thus for further information on the use and generation of one-, two- and three-dimensional properties as well as comparisons of the software that may be employed to generate such descriptors, readers are referred to Chapters 5–8.

2.2.5 Step 5: Generating the Model

As Figure 2.2 shows, once the data on the endpoint have been gathered and assessed and the appropriate parameters generated, the next stage is to generate the model itself. This is the key step in the process and careful consideration needs to be given as to how to approach model generation.

There are many different modelling approaches and statistical methods of analysis from which to choose. Selection of the most appropriate method must take into account:

- the nature of the endpoint to be modelled and the accuracy of the data for the endpoint (*i.e.* accurately measured continuous data or categorical classifications based on positive/negative indicators or rank ordering);

- the amount of data available (small or large data sets);
- the information available on the compounds (*i.e.* which parameters can be reasonably obtained and what they can tell you about the molecule); and
- the purpose for which the model is being built (*i.e.* for general screening/rank ordering purposes or for quantitative prediction).

The simplest models are those based on 'rules of thumb', which are broadly applicable and useful for general screening purposes. A notable example of this is Lipinski's 'rule-of-fives'[18] which has gained broad acceptance in drug development because of its simplicity and interpretability. The rule states that for any given compound: log $P > 5$; molecular weight > 500; number of hydrogen bond donors > 5; number of hydrogen bond acceptors > 10, are all factors associated with poor intestinal absorption. Similarly, cut-off values for certain properties can be useful for classifying compounds into broad categories. For example in toxicological assessment, compounds with log P values > 4.5 have been classified as having the potential to bioaccumulate.[19]

Where specific structural features can be identified as being associated with a particular activity, structure–activity relationship (SAR) models may be generated. For example the presence of nitroso groups has been associated with carcinogenicity,[20] glycol ethers have been associated with developmental toxicity,[21] and Ashby and Tenant indicated a number of structural alerts associated with carcinogenicity and mutagenicity endpoints.[22] The identification of structural alerts such as these can be formalised into knowledge-based prediction methods; these form the basis (or rule-base) of many 'expert systems' to predict toxicity. The methods employed by these systems, their use and application, are discussed further in Chapter 19.

Moving beyond these simple and intuitive relationships, more mathematical models can be derived providing quantitative estimates of potency, *i.e.* quantitative structure–activity relationships (QSARs). These models may be global (*i.e.* covering large data sets of diverse molecules) or may be more local models applicable to fewer compounds representing a narrower area of chemical space. Global models have the advantage of being generally applicable (although detailed information on interactions may be lost), whereas local models may provide more insight into the mechanisms involved in the process (at the cost of being applicable to fewer chemicals).

A QSAR may be developed using simple linear regression, *i.e.* where a single parameter is correlated with biological activity. More commonly, more than one parameter determines activity; hence multiple linear regression is necessary to generate the model.

There are many statistical methods that can be used to generate QSAR models; the most appropriate depends on the nature of the data set, the descriptors available and the ratio between the numbers of compounds and descriptors. Chapters 9 and 10 discuss in detail the most appropriate choice of statistical methods for analysing continuous and categorical data, respectively, and how to interpret the statistics generated for the models.

Recently there has been a great deal of interest in category formation and read-across methods as tools for *in silico* prediction of toxicity. These are based on the premise that 'similar' molecules will possess 'similar' activity. Similarity itself can be defined in many ways. For example compounds may be similar in terms of parameter values (*e.g.* log P, log S_{aq}), size, molecular shape, reactivity, presence of structural features, *etc.*

Similarity in one respect does not mean that they will be similar in others (*e.g.* compounds of the same molecular weight may have vastly different log P values). Hence it is important to determine which property is related to the activity and look for molecules that are similar in respect of that property. This allows for 'groups' of chemicals to be formed; the common feature of the group should be associated with the activity. If the activity of one (or preferably more) of the members of the group is known, then the activity of other members of the same group can be predicted. This can be done in either a qualitative or quantitative manner.

One useful way to group chemicals together is by mechanism of action. For example, chemicals acting as Michael acceptors are known to react with skin proteins resulting in skin sensitisation. This means that a category can be built for chemicals that may act as Michael acceptors. If several members of the category are known to be skin sensitisers, then the prediction could be made that other category members are also likely to be skin sensitisers (note this is a simplified example, further details of this example can be found in Enoch *et al.*[23]).

Chapter 13 discusses the role of the underlying mechanisms in toxicity, information which could be useful in forming toxicologically meaningful categories; this topic is developed further in Chapter 14.

An increasing number of tools are becoming available for category formation and read across. Chapter 15 provides a brief introduction to the principles of read-across and the theme is developed further in Chapter 16, which discusses the use of the OECD (Q)SAR Application Toolbox for making predictions using read-across. Chapter 17 considers other tools that may be used to investigate similarity of molecules which can be used to form categories.

Whilst there is a vast array of techniques available for generating models, the data available and the purpose for which the model is being developed should provide guidance on which is the most appropriate method to select for a given query.

2.3 Assessing the Validity of the Model

Given sufficient data and a high number of descriptors, it is possible to generate a large number of models; however, in order to be useful, the model needs to be valid. Whilst there has been a great deal of interest lately in what constitutes a 'valid' model, key principles of developing a good model were formally proposed more than 35 years ago. Unger and Hansch[24] proposed five criteria for selecting the 'best equation' to correlate activity, these were:

(i) selection of independent variables, *i.e.* those that are not inter-correlated with other variables;

(ii) justification of the choice of independent variables, *i.e.* these should be validated by an appropriate statistical procedure;
(iii) principle of parsimony, *i.e.* to use the simplest model;
(iv) number of terms, *i.e.* to have sufficient data points available per descriptor to avoid chance correlations. (The Topliss and Costello[25] rule recommends at least five data points to be incorporated per descriptor added.);
(v) qualitative model, *i.e.* one which is consistent with the known physical-organic and biomedicinal chemistry of the process involved (now generally referred to as 'mechanistically interpretable').

Moving forward to 2004, the OECD Principles for the Validation for Regulatory Purposes of (Q)SARs (which can be used also to guide assessment of the validity of other *in silico* models) were proposed.[26] These state that the model should be associated with:

(i) a defined endpoint;
(ii) an unambiguous algorithm;
(iii) a defined domain of applicability;
(iv) appropriate measures of goodness-of-fit, robustness and predictivity;
(v) a mechanistic interpretation, if possible.

Sections 2.2.1 and 2.2.5 have dealt with selecting an appropriate endpoint and method to generate the model (the algorithm). In terms of selecting an unambiguous algorithm, more transparent models such as multiple linear regression (MLR) are preferred to non-transparent models such as certain neural network methods.

The domain of applicability is used to determine the chemical space for which the model is applicable. This may be defined in terms of structural features of the compounds, physico-chemical properties or mechanism of action (or a combination of these). For example, if a model relating log P to toxicity was built entirely using compounds with log P values between 2 and 5, then it would be questionable to use the same model to predict the toxicity of a compound with a log P value of 8, as this is beyond the scope of the parameters used to generate the model (*i.e.* the compound falls outside of the applicability domain of the model). There are many ways in which the applicability domain of the model can be defined and these are presented in detail in Chapter 12.

In terms of assessing 'appropriate measures of goodness-of-fit, robustness and predictivity', the method by which the performance of a model is judged is dependent upon the type of model generated. For a classification model, a measure of concordance (*i.e.* the percentage of compounds placed into the correct category) gives a useful indication of overall model performance. However, in terms of toxicity prediction, sensitivity and specificity may be more important. Sensitivity is the proportion of compounds correctly classified as 'toxic' (or active) relative to the total actual number of toxic (active) compounds; specificity is the proportion of compounds correctly classified as

'non-toxic' (or inactive) compared to the total number of actual non-toxic (inactive) compounds. These are important in toxicity as potentially the consequences of a false negative prediction (*i.e.* predicting a toxic compound to be non-toxic) are much more serious than a false positive prediction (*i.e.* predicting a non-toxic compound to be toxic). In the case of the former, inadequate measures may be put in place to protect humans and the environment resulting in harm; in the latter case, excessive protective measures may be put in place which could be costly to industry.

For linear and multi-linear regression based models, the proportion of variability in the data accounted for by the model (r^2 values) are often used to determine model performance. Statistical measures should also include an assessment of the predictivity of the model, *i.e.* how well it performs in predicting compounds that were not present as part of the training set. Appropriate statistical measures on which to judge model performance are given in Chapters 9 and 10.

Considering point (v) of the Principles above, the model should be interpretable, *i.e.* the parameters included in the model should 'make sense' in terms of what is known about the process.

If the process is well understood, then it is generally easier to rationalise the presence of readily interpretable parameters; where little is known of the mechanisms of the process this may not be readily achievable. Chapter 11 discusses the validation of *in silico* models in more detail.

2.4 Reviewing and Updating the Model

Once the model has been generated and the validity assessed, this should not be considered as the end of the process in terms of using *in silico* models for toxicity prediction. The presence of outliers (*i.e.* compounds for which the model poorly predicts toxicity/activity) can provide useful information which can be used to revise the current model or generate new models. In some cases the outlier may have the potential to act *via* a different mechanism to other compounds used to generate the model. This can provide useful insight into mechanisms of action involved in toxicity.

An outlier may be due to an erroneous experimental measurement (in which case it should be omitted) or may provide information on the limitations of the experimental procedure. For example, highly volatile compounds may be lost from the experimental media before a true measure of their toxicity (at a given concentration) could be measured. Similarly a compound may be metabolised or abiotically transformed to another compound before it reaches its site of action. Such compounds may also appear as outliers as the parent compound is no longer present (or is present in much lower concentration) and the measured effect may in fact be due to metabolite or other transformation product.

Acquisition of new information of this type or of further data for a given endpoint should be used to re-develop models in an iterative manner. In this way models are continually updated and improved. Within a given industry,

the chemical space of compounds of interest may gradually evolve over time; hence the applicability of a model developed in the 'old' chemical space may not be predictive for compounds in the 'new' chemical space. Model development should be a dynamic, ongoing concern utilising the most up-to-date information and techniques.

2.5 Using the Model

Many *in silico* models already exist and there is the potential to generate infinitely more. When electing to use a model whether obtained from the literature, formalised in an expert system or generated in-house, very careful consideration must be given to its appropriate use. Models developed have often been highly criticised because they fail to perform well when tested with a given set of compounds. However, such apparent 'failures' of models can often be traced back to inappropriate use.[27] To make a prediction for any compound it must be determined as to whether or not it falls within the applicability domain. Compounds that do not fall within the domain may be poorly predicted.

Models may be developed for very general purposes and be useful for these. For example, models to predict intestinal absorption of drug candidates can be used to screen large in-house virtual libraries in drug companies; however, these may not be appropriate for predicting absorption of pollutants that enter into the food chain. In this case more specific models giving a more accurate prediction for compounds of a different chemistry are needed.

The confidence needed for a prediction also depends on the use to which it will be put. A global model may be useful for prioritising testing of compounds from a large inventory, where the *in silico* model is used to select which of the compounds are more likely to be associated with a toxic effect. In Integrated Testing Strategies (ITS), which are discussed further in Chapter 23, *in silico* models may be used to select which of a group of compounds should be selected, based on their predicted activities, for further testing. For example, category formation can be used to assign chemicals to appropriate groups, one or more members from each group can then be selected for testing to help fill data gaps. Another use is in weight-of-evidence approaches in cases where several models exist for the same endpoint. Whilst use of each individual model may not result in a high confidence estimate, the use of several models in combination may increase confidence, particularly where results from all models are concordant. Hewitt *et al*. demonstrated an example of this for reproductive toxicity endpoints.[28] The weight-of-evidence concept is discussed further in Chapter 22.

2.6 Consideration of Mitigating Factors

Although a good *in silico* model may be able to provide an accurate prediction of inherent toxicity of a given compound, the ability of the compound to actually elicit such an effect can be significantly modulated by other factors.