NAFTA

TECHNICAL WORKING GROUP ON PESTICIDES (TWG)

(Q)UANTITATIVE STRUCTURE ACTIVITY RELATIONSHIP [(Q)SAR] GUIDANCE DOCUMENT

North American Free Trade Agreement (NAFTA)
Technical Working Group on Pesticides (TWG)

(Quantitative) Structure Activity Relationship [(Q)SAR]
Guidance Document

November, 2012

Contributors

Mary Manibusan, US EPA
Joel Paterson, PMRA
Dr. Ray Kent, US EPA
Dr. Jonathan Chen, US EPA

Dr. Jenny Tao, US EPA
Dr. Edward Scollon, US EPA
Christine Olinger, US EPA
Dr. Patricia Schmieder, US EPA
Dr. Chris Russom, US EPA
Dr. Kelly Mayo, US EPA
Dr. Yin-tak Woo, US EPA
Dr. Thomas Steeger, US EPA
Dr. Edwin Matthews, US FDA
Dr. Sunil Kulkarni, Health Canada

External Peer Reviewers

Kirk Arvidson, US FDA
Mark Bonnell, Environment Canada
Bob Diderich, OECD
Terry Schultz, OECD
Andrew Worth, European Commission – Joint Research Centre

PREFACE

Integrated Approaches to Testing and Assessment (IATA) and (Q)SAR

Pesticide regulatory agencies have traditionally relied on extensive *in vivo* and *in vitro* testing to support regulatory decisions on human health and environmental risks. While this approach has provided strong support for risk management decisions, there is a clear recognition that it can often require a large number of laboratory animal studies which can consume significant amounts of resources in terms of time for testing and evaluation. Even with the significant amounts of information from standard *in vivo* and *in vitro* testing, pesticide regulators are often faced with questions and issues relating to modes of action for toxicity, novel toxicities, susceptible populations, and other factors that can be challenging to address using traditional approaches.

Recognizing the limitations of current testing approaches and the rapid development of new biochemical and cellular assay systems and computational predictive methods, pesticide and other regulatory agencies have initiated the long-term investigation of Integrated Approaches to Testing and Assessment (IATA). IATA integrate existing knowledge bases on classes of chemicals with the results of biochemical and cellular assays, computational predictive methods, exposure studies, and other sources of information to identify requirements for targeted testing or develop assessment conclusions. In some cases, the application of IATA could lead to the refinement, reduction, and/or replacement of selected conventional tests (*e.g.,* animal toxicity tests). IATA also have the potential to further enhance the understanding of mode/mechanism of action[1] including the consideration of relevant adverse outcome pathways (AOPs) that provide biological linkages between molecular initiating events to adverse

[1] In this context, mode of action for toxicity is the description of key events and processes, starting with interaction of an agent with the cell through functional and anatomical changes, resulting in cancer or other health endpoints. Mechanism of action for toxicity is the detailed molecular description of key events in the induction of cancer or other health endpoints and represents a more detailed understanding and description of events than is meant by mode of action. Mode of action for toxicity can also be differentiated from the pesticidal mode of action which is the specific biochemical or physical effect(s) by which the pesticide kills, inactivates or otherwise controls pests. Mechanism and mode of action for toxicity are important components of adverse outcome pathways (AOPs).

outcomes in individual organisms and populations that are the bases for risk assessments.

The subject of this guidance document, (Quantitative) Structure Activity Relationships [(Q)SAR], is an important set of predictive tools that can be considered when applying IATA to pesticide assessments. (Q)SAR represents a variety of techniques for predicting activities and properties of untested chemicals based on their structural similarity to chemicals with known activities and properties.[2] (Q)SAR methods have a long history of use both for the industrial design and regulatory assessment of pharmaceuticals, pesticides, and other chemicals. While historical and current applications of (Q)SAR methods have focused on the prediction of physical-chemical properties and apical endpoints (*e.g.,* toxicity, ecotoxicity), as IATA are developed and applied to pesticides, a greater emphasis will be placed on using (Q)SAR to predict key events along the cascade of obligatory steps toward the adverse outcome in modes of toxicological action and AOPs (*e.g.,* receptor binding potential, enzyme activation/inhibition, DNA/protein binding).

The development and application of IATA and (Q)SAR methods to pesticide assessments is consistent with the United States Environmental Protection Agency (US EPA) commissioned National Research Council (NRC) report, *Toxicity Testing in the 21st Century: A Vision and a Strategy* (NRC, 2007). The NRC's vision emphasizes moving away from checklists of conventional toxicity studies towards integrated approaches using existing knowledge of chemicals and the results of alternative testing methods, including computational tools such as (Q)SAR, to identify toxicity pathways and streamline data requirements for more efficient, and effective, targeted toxicity testing. (Q)SAR has also been highlighted as an important IATA tool in the report, *Integrating Emerging Technologies into Chemical Safety Assessment*, sponsored by Health Canada and prepared by the Expert Panel on the Integrated Testing of Pesticides of the Canadian Council of Academies (CCA) (CCA, 2012). The CCA report provides an update on the status of IATA and IATA tools, and a vision for the evolution of IATA in the regulatory context.

---

[2] (Q)SAR is the study of the correlation between chemical structure and associated biological activity, with the ultimate goal of predicting the activity of untested chemicals based on structurally related compounds with known activity. The parentheses around the "Q" in (Q)SAR indicates that the term refers to both qualitative predictive tools (i.e., structure-activity relationships (SARs)) and quantitative predictive methods (quantitative structure-activity relationships (QSARs)). Although the term (Q)SAR is often used to refer to predictive models, especially computer-based models, it should be noted that (Q)SAR is actually inclusive of a wide variety of computerized and non-computerized tools and approaches.

Moving IATA from a long-term vision into mainstream practice for pesticide assessments will require the development and application of biochemical and cellular assays, along with the further development and broader application of existing tools such as (Q)SAR. Towards that end, the United States Environmental Protection Agency Office of Pesticide Programs (US EPA OPP) has partnered with the Pest Management Regulatory Agency (PMRA) of Health Canada to develop common approaches to IATA for the human health and ecological risk assessment of pesticides. The formalized framework for this partnership is a North American Free Trade Agreement (NAFTA) Joint Project on "21st Century Toxicology: Integrated Approaches to Testing and Assessment". While this project is intended to cover a broad array of computational toxicity tools, a key current activity is the development of this NAFTA (Q)SAR guidance document for pesticide risk assessors.

The primary purpose of this guidance document is to help pesticide evaluators to evaluate (Q)SAR predictions and to identify the important issues that may be involved when incorporating predictions in the risk assessment process. The document is not intended to reproduce or replace the ever-expanding volume of journal articles, reports, documents, and textbooks on the development and application of (Q)SAR, but to provide an introduction to the evaluation of (Q)SAR tools and their application to pesticide regulatory risk assessments. While the focus of this document is on the application of (Q)SAR to pesticide risk assessments, the principles and issues described in this document are general and may also be used for other types of chemical assessments. Regardless of the scenario to which (Q)SAR is being applied, the peer review process is critical and relevant to the consistent application of this tool. To that end, appropriate (Q)SAR experts should be consulted and peer review procedures used to ensure scientific excellence and rigor.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

GLOSSARY

This Glossary section is intended to provide additional explanation for common scientific terms which are presented in order to enhance communication between (Q)SAR experts and users of (Q)SAR models, particularly in the field of pesticides.

There are two portions in the glossary — abbreviations (acronyms) and terms with more detailed explanations.

ABBREVIATIONS

A/I ratio	Ratio of active to inactive chemicals
ACC	American Chemistry Council
ADME	Absorption, distribution, metabolism, and elimination
AOP	Adverse Outcome Pathway
BCF	Bioconcentration Factor
CAS	Chemical Abstract Service
CCA	Council of Canadian Academies
CEPA	Canadian Environmental Protection Act
CFR	Code of Federal Regulations
DER	Data evaluation record
DSL	Domestic Substances List (Canada)
EC	European Commission
ECHA	European Chemicals Agency
EDSP	Endocrine Disruptor Screening Program (US EPA)
EEC	European Economic Community
EFSA	European Food Safety Authority
ER	Estrogen Receptor
EU	European Union
FAO	Food and Agriculture Organization (of the United Nations)
FDA	(US) Food and Drug Administration
FQPA	Food Quality Protection Act

HPV	High Production Volume Chemicals Program (US EPA)
IATA	Integrated Approaches to Testing and Assessment
ILSI	International Life Sciences Institute
InChI™	IUPAC International Chemical Identifier
IPCS	International Program on Chemical Safety
IUPAC	International Union of Pure and Applied Chemistry
JRC	Joint Research Centre (European Commission)
kNN	k Nearest Neighbor
K_{ow}	Octanol-water partition coefficient
K_p	Permeability coefficient through the skin for a chemical in water
LMO	Leave many out
LOAEL	Lowest Observed Adverse Effect Level
Log P	Logarithm to the base 10 of the 1-octanol/water partition coefficient, also Log K_{ow}
LOO	Leave one out
MED	Mid-Continent Ecology Division (US EPA ORD)
MIE	Molecular initiating event
MOA	Mode of (toxicological) Action
NAFTA	North American Free Trade Agreement
NAS	(US) National Academy of Sciences
NHEERL	National Health and Environmental Effects Research Laboratory (US EPA ORD)
NOAEL	No Observed Adverse Effect Level
NRC	(US) National Research Council
OECD	Organization for Economic Co-operation and Development
OFAS	Office of Food Additive Safety (US FDA)
OPP	US EPA Office of Pesticide Programs
OPPT	US EPA Office of Pollution Prevention and Toxics
ORD	US EPA Office of Research and Development
PBT	Persistent, bioaccumulative and toxic
PCKOC	Organic Carbon Partition Coefficient model components within the EPI Suite (US EPA)
PMN	Premanufacturing notification

PMRA	Pest Management Regulatory Agency (Health Canada)
Q^2	Cross-validated correlation coefficient
Q^2_{ext}	External correlation coefficient
(Q)SAR	Quantitative structure-activity relationship or structure-activity relationship
QSAR	Quantitative structure-activity relationship
QMRF	QSAR Model Reporting Format (European Commission)
QPRF	QSAR Prediction Reporting Format (European Commission)
QSPR	Quantitative structure-property relationship
R^2	Coefficient of determination
REACH	Registration, Evaluation, Authorization (and Restriction) of Chemicals legislation (European Union)
SAR	Structure-activity relationship
SDF	Structure Data Format
SEE	Standard error of the estimate
Spress	Cross-validated standard error of prediction
TSCA	US Toxic Substances Control Act
TTC	Threshold of Toxicological Concern
WHO	World Health Organization

TERMS

Adverse outcome pathway (AOP)	A conceptual construct that portrays existing knowledge concerning the linkage between a direct molecular initiating event and an adverse outcome at a biological level of organization relevant to risk assessment.
Algorithm	A sequence of instructions for carrying out a defined task. Typically the instructions are mathematical equations or computer code.
Analog	A chemical compound that has a similar structure and similar chemical properties to those of another compound, but differs from it by one or a few atoms or functional groups.
Apical endpoint	Observable effects of exposure to a toxic chemical in a test animal. The effects reflect relatively gross changes in animals after substantial durations of exposure.
Chemical category	A group of chemicals with similar physicochemical, human health, or ecotoxicological properties usually resulting from structural similarity.
Congeneric series	A group of chemicals with a common base structure (*e.g.* aliphatic alcohols) but differing in the arrangement of common substituents. The polychlorinated biphenyls are considered a congeneric series.

Cross-validation	A statistical technique for assessing the predictive ability of a QSAR by the removal of different proportions of the chemicals from the training set, developing a QSAR on the remaining chemicals and using that QSAR to predict the activity of those removed. This procedure is repeated a number of times, so that a number of statistics can be derived from the comparison of predicted data with the known data.
Data mining	A collective term that refers to all procedures (informatic and statistical) that are applied to large heterogeneous data sets, in order to develop a data matrix amenable to statistical methods.
Descriptor	A quantifiable physical, chemical, or structural property specific to a chemical that can be correlated with an endpoint under investigation. There are three main categories of descriptors: hydrophobic, steric, and electronic. Steric descriptors are those relating to molecular size or shape. Electronic descriptors are those concerning molecular interactions such as hydrogen bonding and dipole forces and they include quantum mechanical and quantum chemical descriptors such as atomic charge. Hydrophobic descriptors such as Log P are those relating to the tendencies of chemical to partition between hydrophilic (aqueous) and hydrophobic/lipophilic (lipid) phases.
Domain of Applicability	The domain of applicability of a (Q)SAR model is the chemical structure and response space in which the model makes predictions with a given reliability. It can be thought of as a theoretical region in multi-dimensional space in which the model is expected to make reliable predictions. It depends on the nature of the chemicals in the training set, and the method used to develop the model and helps the user of the model to judge whether the prediction for a new chemical is reliable or not.
EC_{50}	Half Maximal Effective Concentration. Statistically derived concentration of a substance expected to induce a response halfway between baseline and maximum effect.
Endpoint	The measure of a biological effect, e.g., LC_{50} or EC_{50}. A large number of endpoints are used in regulatory assessments of chemicals. These include lethality, carcinogenicity, immunological responses, organ effects, developmental and reproductive effects, etc. In (Q)SAR analysis, it is important to develop models for individual toxic endpoints.
Expert system	A formalized system (often computer based) that utilizes a knowledgebase of structure-activity relationships accumulated from human experts. The knowledgebase is applied using a set of expert rules to derive predictions of biological activity for chemicals of interest based on the presence of specific chemical structures.

External validation	A validation exercise in which the chemical structures selected for inclusion in the test set are different from those included in the training set, but which should be representative of the same chemical domain. The QSAR model developed by using the training set chemicals is then applied to the test set chemicals in order to assess the predictive ability of the model.
Functional group	A molecular moiety that imparts certain characteristics to a molecule, e.g., hydroxyl (OH^-), amino (NH_2^-), or nitro (NO_2^-). When only a limited number of functional groups are present, they may be the primary basis for the specific chemical, physical or biological characteristics of a chemical. However, for complex chemicals with many functional groups, the simple interactions associated with individual functional groups may not be reliable predictors of chemical behavior unless one functional group predominates for the particular activity.
Genetic algorithm	A statistical method that selects the best combination of descriptors to describe a given property, modeled on the principle of the survival of the fittest (best) in the breeding of organisms.
In silico	An expression that means “performed on computer or via computer simulation.”
LC_{50}	Median Lethal Concentration. Statistically derived concentration of a substance expected to cause death in 50% of test animals, usually expressed as the weight of substance per weight or volume of water, air or feed, e.g., mg/l, mg/kg or ppm.
LD_{50}	Median Lethal Dose. Statistically derived single dose causing death in 50% of test animals when administered by the route indicated (oral, dermal, inhalation), expressed as a weight of substance per unit weight of animal, e.g., mg/kg.
Lipinski’s rule of 5	A rule of thumb developed by Christopher Lipinski for evaluating whether the properties of a chemical are likely to make it an orally active drug in humans. The rule states that, in general, an orally active drug has no more than one violation of the following criteria: not more than 5 hydrogen bond donors, not more than 10 hydrogen bond acceptors, a molecular weight under 500, and an octanol-water partition coefficient less than 5.
Mechanism of Action (Toxicity)	The detailed molecular description of key events in the induction of cancer or other health endpoints. Mechanism of action for toxicity represents a more detailed understanding and description of events than is meant by mode of action. Mechanism of action of toxicity is an important component of an adverse outcome pathway (AOP).

Mode of Action (Pesticide)	The mode of action of a pesticide refers to the specific biochemical or physical effect(s) by which the pesticide kills, inactivates, or otherwise controls pests.
Mode of Action (Toxicity)	The description of key events and processes, starting with interaction of an agent with the cell through functional and anatomical changes, resulting in cancer or other health endpoints. Mode of action for toxicity is an important component of an adverse outcome pathway (AOP).
OECD QSAR Toolbox	The OECD QSAR Toolbox is a software application intended to be used by governments, chemical industry, and other stakeholders in filling gaps in (eco)toxicity data needed for assessing the hazards of chemicals. The Toolbox incorporates information and tools from various sources in a logical workflow. Crucial to this workflow is the grouping of chemicals into chemical categories (http://www.oecd.org/document/54/ 0,3746,en_2649_34379_42923638_1_1_1_1,00.html).
Outlier	A data point that is far removed from other members of the dataset. Typically, the outlier of a (Q)SAR model has a cross-validated standardized residual greater than three standard deviation units.
Partition coefficient	The ratio of equilibrium concentrations of a chemical distributed between two immiscible solvents. Frequently octanol and water are used to mimic a chemical distributing between lipid and aqueous phases in an organism, normally expressed as a logarithm to base 10, *i.e.*, Log K_{ow}, or Log P, a descriptor of hydrophobicity.
Point of Departure	Commonly abbreviated POD, the point of departure is the dose-response point that marks the beginning of a low dose extrapolation. This point is often the lower bound on an observed incidence or on an estimated incidence from a dose-response model.
Predictivity	A measure of a model's ability to make reliable predictions for chemical structures not included in the training set of the model.
Read-across	Endpoint information for one or more chemicals (the source chemical(s)) is used to predict the same endpoint for another chemical (the target chemical), which is considered to be "similar" in some way (usually on the basis of structural similarity or similar mode or mechanisms of action). Sometimes, it is also referred to as "data bridging." In principle, read-across can be used to estimate physicochemical properties, toxicity, environmental fate, and ecotoxicity. For any of these endpoints, it may be performed in a qualitative or quantitative manner.
QSAR	Quantitative structure-activity relationship — a quantitative relationship between an endpoint (biological activity, *e.g.*, toxicity) and one or more descriptors associated with the endpoint/activity.

SAR	Structure-activity relationship — a qualitative relationship (*i.e.,* an association) between a molecular (sub)structure and the presence or absence of a biological activity, or the capacity to modulate a biological activity imparted by another substructure.
SMILES	Simplified Molecular Input Line Entry System — a computer-compatible, standardized, two-dimensional description of chemical structure. The SMILES string is written by following a small number of rules. In brief, each non-hydrogen atom (hydrogen is only explicitly included in special circumstances) is denoted by its symbol; double and triple bonds are shown by "=" and "#" symbols, respectively; branches are shown in parentheses; and rings are opened and closed by the use of numbers. For example, CCO represents ethanol, and c1ccccc1N represents aniline (the digits indicate the beginning and ending of ring, and lower case "c" indicates aromatic carbon).
Structural alert	A molecular (sub)structure associated with the presence of a specific (usually adverse) biological activity.
Substructure	A portion of the overall structure of a chemical that may be associated or correlated with a biological activity or property of the chemical.
TD_{50}	The statistically derived median toxic dose of a drug or toxin at which toxicity occurs in 50% of the test population.
Test set	A set of chemicals, not included in the training set used to develop a QSAR, that is used to validate (assess the predictive ability of) the QSAR. It is sometimes called an "independent" or "external" test set or validation set. For the purpose of (Q)SAR validation, it is important that the test set has the same domain of applicability as the training set, and contains a sufficient number of chemical structures.
Toxicity pathway	A cellular response pathway that, when sufficiently perturbed, is expected to result in adverse health effects (NRC, 2007). Toxicity pathways are important components of adverse outcome pathways (AOPs).
Training set	A set of chemicals used to derive a QSAR. The data in a training set are typically organized in the form of a matrix of chemicals and their measured properties or effects observed in a toxicity test. A homogeneous training set is a set of chemicals which belong to a common chemical class or share a common chemical functionality or a common mechanism of action. A heterogeneous training set is a set of chemicals which belong to multiple chemical classes, or which do not share a common chemical functionality or common mechanism of action.
Validation	The testing of a (Q)SAR tool to assess its reliability and relevance. The OECD Guidance Document on the Validation of (Quantitative)Structure-Activity Relationship (Q)SAR Models (OECD Series on Testing and Assessment No. 69) defines validation as the process by which the reliability and relevance of a particular approach, method, process or assessment is established for a defined purpose.

1. EXECUTIVE SUMMARY

While pesticide regulatory agencies have traditionally relied on extensive *in vivo* and *in vitro* testing to support regulatory decisions on human health and environmental risks, these and other agencies have initiated the long-term investigation of Integrated Approaches to Testing and Assessment (IATA). The application of IATA could lead to the refinements, reduction, and/or replacement of conventional tests through the integration of existing knowledge bases on chemicals, biochemical and cellular assays, computational predictive methods, exposure studies, and other sources of information to identify targeted testing requirements or develop assessment conclusions. (Quantitative) Structure Activity Relationships represent an important set of predictive tools to be considered when applying IATA to pesticide risk assessments.

Moving IATA from a long-term vision into mainstream practice for pesticide assessments will require the development and application of new predictive tools and the further development and broader application of existing tools such as (Q)SAR. In recognition of these requirements and the need to develop common approaches to IATA for the risk assessment of pesticides, the United States Environmental Protection Agency (US EPA) Office of Pesticide Programs (OPP) and the Pest Management Regulatory Agency (PMRA) of Health Canada have established a North American Free Trade Agreement (NAFTA) Joint Project on "21st Century Toxicology: Integrated Approaches to Testing and Assessment".

(Q)SAR is the study of the correlation between chemical structure and associated (biological) activity, with the ultimate goal of predicting the activity of untested chemicals based on structurally related compounds with known activity. The parentheses around the "Q" in (Q)SAR indicates that the term refers to both qualitative predictive tools *(i.e.,* structure-activity relationships (SARs)) and quantitative predictive methods (*i.e.,* quantitative structure-activity relationships (QSARs)). Although the term (Q)SAR is often used to refer to predictive models, especially computer-based models, (Q)SAR is actually inclusive of a wide variety of computerized and non-computerized tools and approaches.

The development of this NAFTA (Q)SAR Guidance Document is a key activity under the NAFTA Joint Project. The purpose of this guidance document is to help pesticide evaluators to evaluate (Q)SAR related information and to identify the important issues that may be involved when incorporating (Q)SAR information into the risk assessment process. This document does not reproduce or replace the ever-expanding volume of journal articles, reports, documents, and textbooks

that provide guidance on the development and application of (Q)SAR, but provides an introduction to the evaluation of (Q)SAR tools and their application to pesticide regulatory risk assessments. While the focus of this document is on the application of (Q)SAR to pesticide risk assessments, the principles and issues are general enough to be applied to other types of chemicals. Regardless of the type of risk assessment scenario, (Q)SAR experts should be consulted and peer review procedures used to ensure scientific excellence and rigor.

The document is organized into eight sections including this executive summary.

Section 2 provides an introduction to some current applications of (Q)SAR to pesticide risk assessments with an emphasis on the use of (Q)SAR by the US EPA OPP and the PMRA. It also includes a brief discussion of other regulatory applications of (Q)SAR at the US EPA, the US FDA, Health Canada and Environment Canada, the OECD, and the European Commission. Finally the overall purpose of the guidance document is discussed and a schematic is provided as guide to the contents of the document.

The purpose of section 3 is to provide some brief background information on the definition of (Q)SAR, types of (Q)SAR tools and approaches, and some key issues associated with the development of (Q)SAR tools. In particular, the importance of data quality and mode/mechanism of action for toxicity information in the development of (Q)SAR models is highlighted. Also, while (computerized) (Q)SAR models are frequently cited in examples in this document, section 3 illustrates that (Q)SAR actually consists of a range of tools and approaches.

Section 4 focuses on the preliminary analysis of (Q)SAR predictions as one of the several potential sources of information to be integrated at the problem formulation stage of a pesticide assessment. Problem formulation for (Q)SAR essentially involves answering questions on the assessment context for (Q)SAR, the characteristics of the pesticide, the characteristics of the (Q)SAR tool and prediction, and what empirical data are available including any information on mode/mechanism of action for toxicity.

The topic of section 5, evaluating whether a (Q)SAR prediction is adequate or “fit for purpose”, is an important component of applying a prediction to a pesticide assessment. Four key factors originally outlined by the European Commission: the scientific validity of the model, the applicability of the model to the query chemical, the reliability of the (Q)SAR result, and the relevance of the (Q)SAR model for the regulatory purpose are used to guide pesticide evaluators through the information to be considered when evaluating whether predictions from

(Q)SAR tools are adequate for consideration in pesticide assessments. Section 5 also includes a discussion of the documentation of (Q)SAR tools and predictions.

Section 6 briefly discusses approaches to combining information from multiple (Q)SAR predictions, advantages and disadvantages of combining predictions, selecting (Q)SAR tools for multiple predictions, and the evaluation of multiple predictions. Because different (Q)SAR tools may have different prediction paradigms and different strengths and limitations, combining predictions has the potential to increase the confidence in the overall prediction. It is also noted that combining predictions from multiple (Q)SAR tools does not eliminate the need to ensure that each prediction is adequate or fit for purpose and it is not always necessary to combine predictions.

The National Academy of Sciences risk assessment paradigm (i.e., hazard identification, dose response assessment, exposure assessment, and risk characterization) provides the context for section 7 which emphasizes guidance on the integration of (Q)SAR tools into the hazard identification component of the risk assessment process for pesticides. Section 7 builds upon previous sections and includes a consideration of the findings at the problem formulation stage, evaluating empirical data versus (Q)SAR predictions, a consideration of mode of action data, the overall weight of evidence, and hazard identification, and risk communication.

Section 8 provides conclusions and perspectives on the future vision for (Q)SAR and pesticides. It is noted that the conclusions of the *NAS report on Toxicity Testing in the 21st Century: A Vision and a Strategy* with respect to increased reliance on existing knowledge-bases for chemical classes and alternative testing methods is especially relevant for pesticide regulatory authorities and will require research on new testing technologies and integrated approaches to testing and assessment (IATA) for more efficient and effective reviews that don't compromise public health and the environment. (Q)SAR tools are one example of an alternative method that could be applicable to IATA and the increasing use of these tools by pesticide authorities make it important to communicate a systematic and transparent approach to using (Q)SAR in pesticide assessments. This guidance document is consistent with the current hazard/risk assessment paradigm with an overall emphasis on not using (Q)SAR in isolation. In addition to the validity and relevance of the individual (Q)SAR tools and predictions, the defensibility of predictions depends on biological consistency and plausibility across all scientific lines of evidence in a holistic weight of evidence approach. Future applications will involve anchoring (Q)SAR predictions with what is known about chemical classes/categories, biological mode of action, toxicity pathways

and population effects. Eventually, (Q)SAR predictions will be built into larger conceptual frameworks or Adverse Outcome Pathways (AOPs) that delineate the documented, biologically plausible, measurable, and testable processes by which a chemical induces molecular perturbations and subsequent biological responses that are relevant for risk assessment.

In addition to the eight sections discussed above, the document also includes an appendix of the web pages for a number of national and international organizations that may be useful to evaluators seeking additional information on general (Q)SAR concepts, and the development, validation, and evaluation of (Q)SAR tools and predictions (Appendix I), an appendix summarizing the content of the European Commission's (Q)SAR model and prediction reporting formats, examples of detailed information templates that could be considered when (Q)SAR predictions are used as critical sources of data in pesticide assessments (Appendix II), and an appendix of several examples of the application of (Q)SAR tools and methods to pesticides and other chemicals (Appendix III).

2. INTRODUCTION

INTRODUCTION

Topics Discussed in this Section:

- **Current applications of (Q)SAR in pesticide risk assessments**
- **Other regulatory applications of (Q)SAR**
- **Purpose of the NAFTA (Q)SAR Guidance Document**

2.0 Current Applications of (Q)SAR in Pesticide Risk Assessments

In general, pesticide regulatory programs have extensive testing requirements as part of the registration process and as a result, they have not had to rely heavily on predictive methods such as (Q)SAR. However, this is changing over time as pesticide agencies have begun to investigate alternative testing methods such as (Q)SAR to help enhance the efficiency of their assessment processes. This is particularly the case for the investigation and application of Integrated Approaches to Testing and Assessment (IATA) to pesticide risk assessments. IATA have the potential to integrate existing data on pesticides with the results of alternative methods (*e.g.,* biochemical/cellular assays, (Q)SAR) leading to the refinement, reduction, and/or replacement of conventional test requirements.

Provided below is a brief overview of some of the current applications of (Q)SAR by pesticide regulatory agencies.

2.0.1 United States Environmental Protection Agency, Office of Pesticide Programs (US EPA OPP)

2.0.1.1 Application of (Q)SAR to Pesticide Metabolites and Degradates

The United States Environmental Protection Agency, Office of Pesticide Programs (US EPA OPP) generally considers that the toxicity and ecotoxicity studies required to support the evaluation of pesticides adequately address the mammalian and environmental hazard profiles. The agency does not typically require separate toxicity testing of pesticide plant or livestock metabolites or environmental degradates even though there may be much greater human and non-human exposure to the metabolites and degradates than the parent pesticides. Also, historically, the US EPA OPP has typically included only major

(>10%) pesticide metabolites and degradates in human dietary and environmental risk assessments.

The advent of the Food Quality Protection Act (FQPA) has necessitated an increased refinement of pesticide risk assessments including a closer scrutiny of all metabolites and degradates. In order to determine whether these metabolites and degradates should be included in human dietary and environmental risk assessments in the absence of detailed toxicity data, the US EPA OPP has relied upon various types of structural similarity evaluations. Also, more recently, the agency has explored the use of (Q)SAR models to predict the potential toxicity of pesticide metabolites/ degradates in order to provide scientific rationales and support for requiring additional toxicity testing, to substantiate the use of metabolites/degradates in estimates of total toxic residues, or to exclude metabolites/degradates from further testing based on a lack of toxicity concerns. Similarly, the US EPA OPP has made sporadic use of bridging techniques and structure activity relationships to identify whether additional ecotoxicity testing of environmental degradates should be required and whether these residues should be included in environmental exposure estimates for pesticides.

Since empirical data are typically available on the parent pesticide, one of the key factors considered when determining whether (Q)SAR model predictions for the toxicity or ecotoxicity of metabolites and degradates are reliable enough to be used is how well predictions from the same model for the parent pesticide compare to the empirical data for the parent pesticide.

2.0.1.2 Application of (Q)SAR to Antimicrobial Agents

The US EPA OPP has several on-going initiatives and projects related to the application of (Q)SAR to antimicrobial pesticide agents and one of the most important initiatives is the proposed revised testing requirements for antimicrobial agents (40CFR158 subpart W). In this proposed rule, EPA has indicated that it will consider any submission using appropriate SAR analyses and QSAR modeling to supplement or fulfill data requirements for antimicrobial pesticide chemicals.

Approaches with concepts similar to structure activity relationships (SARs) are also being utilized in a pilot project on non-animal eye irritation tests for antimicrobial products with cleaning claims. The purpose of this project is to assess the predictive performance of registrant submitted non-animal eye irritation studies for antimicrobial agents by having registrants include any available Draize rabbit test results for structurally related compounds in their submissions.

There is also an ongoing threshold of toxicological concern (TTC) project that is designed to determine a level of concern for various chemical classes of antimicrobial pesticides. Human exposures below the TTC would not be considered to be of concern and no additional toxicological data would be required. SAR will be used to characterize the toxicity of all chemicals within specific classes of antimicrobial chemicals. This is an American Chemistry Council (ACC) Biocide Panel sponsored project conducted through the International Life Sciences Institute (ILSI) with the US EPA participating on the ILSI Steering Committee and Expert Working Group.

2.0.1.3 Application of (Q)SAR to Ecological Risks from Pesticides

The US EPA OPP estimates chemical properties, environmental fate parameters, and ecological toxicity values for pesticides, inert compounds, and degradates using the EPI Suite and the Assessment Tools for Ecological Risk (ASTER) software on a case-by-case basis when measured values are not available from studies submitted to the Agency or from the open literature. In EPI Suite, the organic carbon partition coefficient (PCKOC) model is used to estimate soil mobility, the KOWWIN model is used to estimate the octanol-water partition coefficients, and the BCFWIN model is used to estimate bioconcentration factors. The Ecological Structure Activity Relationships (ECOSAR) component of EPI Suite and ASTER are used to estimate pesticide ecotoxicity values. These estimates may be used to support human dietary and ecological risk assessments although their use, at this time, is not uniform across the US EPA OPP since formal guidance has not yet been developed. Also, in evaluations against measured values, the organic carbon partition coefficient (PCKOC) model component of EPI Suite has acceptable predictive performance for organic pesticides, but does not perform well for ionic compounds, organometallics, and highly fluorinated pesticides. In addition, while ASTER contains models for five aquatic species (*i.e*., fathead minnow, bluegill sunfish, water fleas, rainbow trout, and channel catfish), it does not support models for terrestrial species. Similar to ASTER, ECOSAR only predicts toxicity for aquatic species and cannot be used to profile inorganic or organometallic chemicals. In the case of environmental degradates, since empirical data are typically available on the parent compound, (Q)SAR-generated toxicity estimates for the parent compound are compared to the available empirical data in order to decide on whether it is appropriate to use (Q)SAR models to estimate the potential ecological toxicity of the environmental degradates. When determining which (Q)SAR model to use, consideration is also given as to which model(s) has the best predictive performance. (Q)SARs have been used by the US EPA OPP to address data gaps in ecological risk assessments on an *ad hoc* basis.

For ecological risk assessments, OPP has made increasing use of bridging techniques and structure activity relationships (SARs) to identify whether additional testing of degradates/transformation products should be required and whether these residues should be included in modeling exposure estimates.

2.0.2 Pest Management Regulatory Agency (PMRA), Health Canada

The Pest Management Regulatory Agency (PMRA) of Health Canada takes into account the same kinds of considerations as the US EPA OPP when addressing the potential toxicity of pesticide metabolites/degradates of chemical pesticides. If metabolites or degradates of a pesticide are identified in plants or soil, but not in rat metabolism studies, the agency will require the submission of available toxicity data on those metabolites. Also, toxicity data on metabolites/degradates are sometimes voluntarily submitted to the PMRA by applicants. In terms of the application of (Q)SAR, the PMRA can include a request for (Q)SAR predictions on metabolites/degradates when requiring the submission of existing data and can also generate (Q)SAR predictions to help identify potential concerns.

2.1 Other Regulatory Applications of (Q)SAR

Unlike pesticide regulatory agencies, programs involved in the regulatory assessment of industrial chemicals, food additives, and other chemicals often only have a limited amount of data available to support their assessments. Consequently, many of these programs have a longer history with the development and use of (Q)SAR tools and approaches.

Several examples of non-pesticidal regulatory applications of (Q)SAR at the US EPA, the US FDA, Health Canada and Environment Canada, the OECD, and the European Commission are summarized below. This is not intended as an exhaustive listing of the uses of (Q)SAR by regulatory agencies, but it should give the reader some context on the development and application of (Q)SAR tools and approaches by a number of prominent national and international agencies. For further information on current developments and applications of (Q)SAR by various national and international agencies, the reader is directed to the various websites listed in Appendix I.

2.1.1 US EPA, Office of Pollution Prevention and Toxics (OPPT)

(Q)SAR methods have been used for identification of potential mutagenic, carcinogenic and other potential health and ecotoxicological hazards and subsequent regulation of new industrial chemicals (premanufacturing notification, or PMN, chemicals) for more than two decades by the US EPA's Office of Pollution Prevention and Toxics (OPPT) under the Toxic Substances Control Act (TSCA) which regulates all industrial chemicals in US commerce. Under TSCA, OPPT is charged with assessing, and if necessary, regulating all phases of the life cycle of industrial chemicals including manufacturing, processing, use and disposal (OECD 2007b).

OPPT has also developed a number of publicly available (Q)SAR tools that are used in regulating substances under TSCA. Examples include the EPI Suite program which includes several models for estimating physical-chemical properties and environmental fate parameters. EPI Suite also contains the ECOSAR model for predicting ecotoxicity. Other tools developed by OPPT include Oncologic, an expert system for predicting carcinogenicity, and an analog identification tool (AIM) for identifying structural analogs.

2.1.2 US EPA, Office of Research and Development

The US EPA Office of Research and Development (ORD), National Health and Environmental Effects Laboratory (NHEERL), Mid-Continent Ecology Division (MED) has been developing (Q)SAR models and related databases since the 1980s. Examples include a database of ecotoxicity information (ECOTOX) as well as ASTER[3], a collection of databases and (Q)SAR models for toxicity to aquatic species. ASTER also includes models to estimate physical-chemical properties, bioconcentration, and environmental fate.

Research at ORD on receptor based toxicity mechanisms in aquatic species has led to the development of a QSAR based expert system for predicting the estrogen receptor binding potential of data poor pesticidal inerts and antimicrobial pesticide active ingredients. The system is designed to prioritize chemicals for further testing in the US EPA Endocrine Disruptor Screening Program (EDSP) and it has been incorporated into the OECD QSAR Toolbox.

[3] ASTER is a US EPA intranet application only accessible to US EPA staff and contractors. http://www.epa.gov/med/Prods_Pubs/aster.htm

2.1.3 US FDA, Office of Food Additive Safety

The US Food and Drug Administration's (FDA) Office of Food Additive Safety (OFAS) has utilized (Q)SAR analysis in the pre-market review of food contact substances for many years and has recently implemented the use of multiple commercial and publicly available (Q)SAR software models (Lo Piparo *et al.,* 2011; Arvidson *et al*., 2010; Bailey *et al*., 2005). OFAS is also investigating the potential application of metabolism prediction software to the review of food contact substances. OFAS uses (Q)SAR analysis as a decision support tool in conjunction with open literature data and submitted test results, and (Q)SAR may also be used to identify the need for additional toxicity testing during pre-submission consultations for food contact substances.

2.1.4 Health Canada and Environment Canada

Health Canada and Environment Canada have extensive experience with the use of (Q)SAR to address selected data requirements for new substances under the Canadian Environmental Protection Act (CEPA). Adequately validated (Q)SAR predictions may be submitted by notifiers or in some cases generated by government evaluators to address physical-chemical properties, persistence/bioaccumulation, human health effects, ecotoxicity endpoints and other endpoints included in the New Substances notification requirements under CEPA. For instance, predictions are sometimes used for assessing substances with low production volumes and in cases where the substance cannot be isolated in pure enough form to provide meaningful test results. (Q)SAR data are generally combined with empirical data and expert judgment in a weight of evidence approach. (Q)SAR was also utilized by both departments for the categorization (prioritization) of existing substances on the Domestic Substances List (DSL) for further assessment. Environment Canada used (Q)SAR predictions to assist with determinations of persistence, bioaccumulation and inherent toxicity to non-human organisms from existing substances while Health Canada used (Q)SAR to generate physical-chemical data to support determinations of greatest potential for human exposure and as part of the hazard tools used to prioritize chemicals for inherent toxicity to humans when data for specific endpoints were not available. (Q)SAR can also be used by both departments as supporting information in screening level risk assessments for DSL substances when experimental data are not available.

2.1.5 Organization for Economic Cooperation and Development (OECD)

Starting in the 1990s, the Organization for Economic Cooperation and Development (OECD) began the investigation of various (Q)SAR methodologies

with the aim of facilitating the application of (Q)SAR approaches in regulatory settings and their regulatory acceptance. One of the most important products from the OECD (Q)SAR project has been the principles for the validation of (Q)SAR models (OECD, 2004). Comprehensive guidance has also been produced on the development and application of grouping methods for chemicals including chemical categories, read-across and trend-analysis (OECD, 2007a).

The OECD has also done extensive work on software for identifying structural characteristics and mode/mechanism of action data on chemicals, systematically grouping them into chemical categories and applying read-across, trend analysis and (Q)SARs to fill data gaps. The end result of these efforts, the OECD QSAR Toolbox, is intended for use by government agencies and stakeholders for addressing gaps in the toxicity and ecotoxicity databases used in the hazard and risk assessment of chemicals and is freely available (OECD, 2011a).

2.1.6 European Commission Joint Research Centre

The European Union's Registration, Evaluation, and Authorization of Chemicals (REACH) legislation is designed to improve the protection of human health and the environment while maintaining competitiveness and increasing innovation in the European chemicals industry. Under the REACH legislation there is a strong emphasis on the use of alternative testing methods to refine, reduce or replace conventional animal testing. The European Commission Joint Research Centre (JRC) Computational Toxicology Group is involved in projects to promote the availability for regulatory application of validated computational methods for assessing environmental distribution and fate, and the effects on human health and the environment in support of the REACH legislation, the European Cosmetics Directive and the assessment of food safety (Mostrag-Szlichtyng *et al*., 2010). The group conducts research on the development of freely available (Q)SAR tools (*e.g.,* Toxtree, DART, Toxmatch), regulatory applications of (Q)SARs and grouping approaches, the use of computational methods to assess the properties of nanomaterials, and the consideration of molecular interactions in the assessment of toxicity. The JRC has also developed templates for documenting the application of the OECD (Q)SAR validation principles to (Q)SAR models including the (Q)SAR Model Reporting Format (QMRF) (EC 2008a) and (Q)SAR Prediction Reporting Format (QPRF) (EC 2008b), and is leading the development of a reporting format for describing key events/intermediate effects in AOPs (OHT 201) in collaboration with the OECD and the European Chemicals Agency (ECHA). The Joint Research Centre has also established a database of (Q)SAR models.

2.2 Purpose of the NAFTA (Q)SAR Guidance Document

The purpose of this guidance document is to help pesticide evaluators to evaluate the (Q)SAR-related information and to identify the important issues that may be involved when incorporating (Q)SAR information into the risk assessment process. It is recognized that there is an ever-expanding volume of journal articles, national and international reports and guidance documents, and academic textbooks on the subject of (Q)SAR. This document does not reproduce or replace these journal articles, reports, guidance documents, and textbooks on (Q)SAR, but provides an introduction to the evaluation of (Q)SAR tools and their application to pesticide regulatory risk assessments.

(Q)SAR predictions can be considered as one of the many potential sources of data for the weight of evidence approaches used in the risk assessment of pesticides. Similar to other sources of data considered, the defensibility of the use of (Q)SAR predictions can be related to the consistency of the predictions generated from the various (Q)SAR tools used as well as the consistency between the predictions and the results of other lines of evidence considered in the weight of evidence approaches.

While many of the illustrative examples in this document involve the application of (Q)SAR to the prediction of toxicity in pesticide hazard assessments, the general principles discussed can also be applied to (Q)SAR predictions for ecotoxicity, physical chemical parameters, and other activities and properties of relevance to pesticide assessments. Similarly, although many issues are raised in the context of the prediction of apical endpoints, pesticide evaluators should recognize that most of these issues will also apply when (Q)SAR is eventually used in IATA to predict key events related to mechanism/mode of action for toxicity and AOPs such as receptor binding, gene activation, enzyme inhibition/activation, etc.

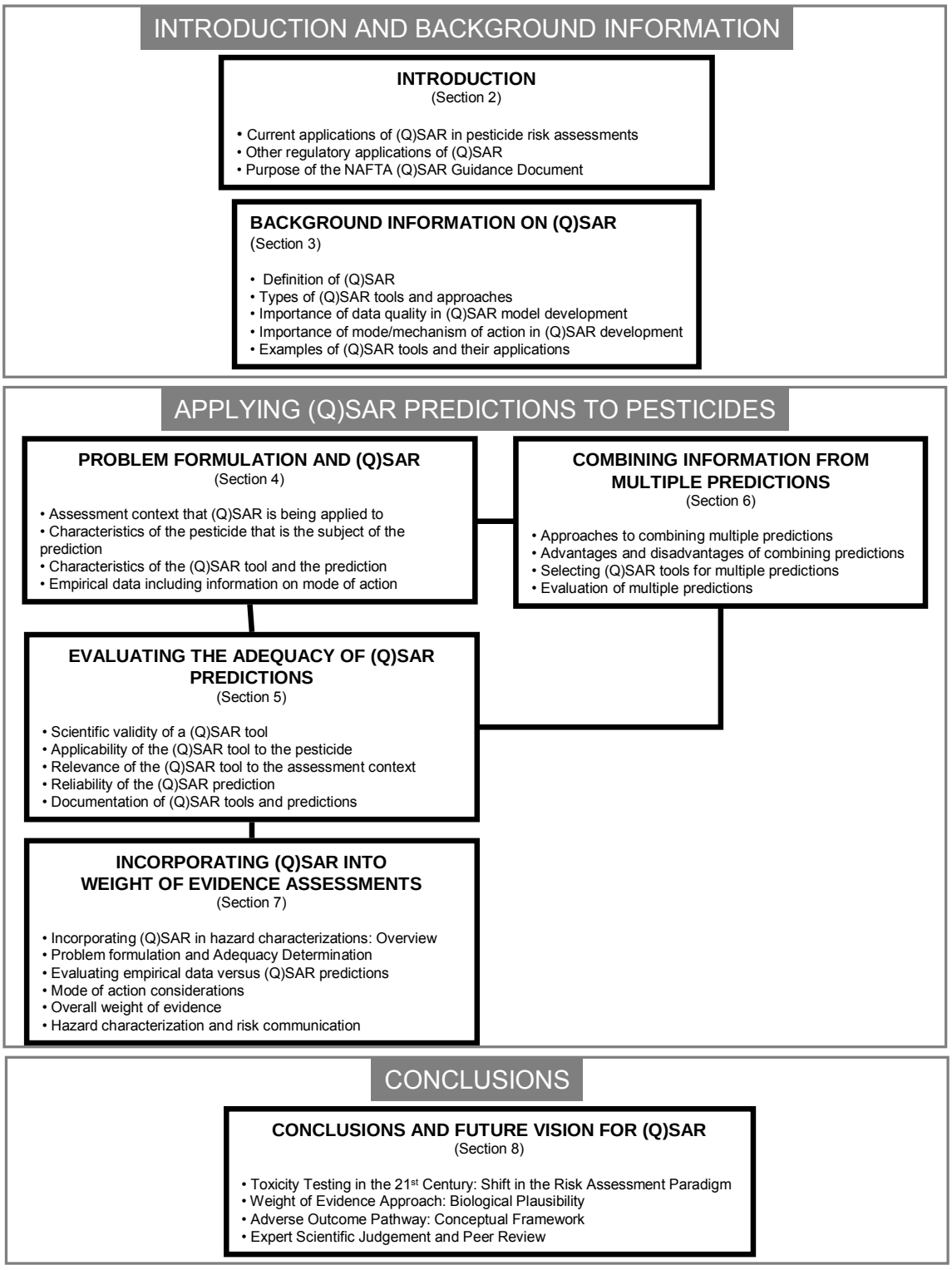
Although this document focuses primarily on (Q)SAR in the context of pesticide risk assessments, the principles and issues discussed are general enough to also be broadly applicable to the use of (Q)SAR in risk assessments for other types of chemicals. Regardless of the type of assessment that (Q)SAR is being applied to, it is recommended that experts in the (Q)SAR field be consulted and that adequate peer review procedures be in place to ensure overall scientific excellence and rigor.

The overall structure of this guidance document is presented in schematic form in Figure 2–0. The document is organized to navigate the pesticide evaluator through sections that provide an introduction to (Q)SAR and (Q)SAR tools, and

information on problem formulation and (Q)SAR, evaluating the adequacy of (Q)SAR predictions, combining information from multiple predictions, and incorporating predictions into weight of evidence assessments. Each section can also be considered as stand-alone guidance on its particular subject area.

As mentioned previously, Appendix I provides a listing of the websites of a number of national and international agencies involved in the development and application of (Q)SAR tools and approaches. These websites could be a useful starting point for those who are interested in learning more about (Q)SAR and obtaining more guidance on its use beyond what is presented in this document. Appendix II summarizes the key features of the European Commission's (Q)SAR model and reporting formats, and Appendix III provides several case study examples of the application of (Q)SAR to pesticides and other chemicals.

Figure 2–0: (Q)SAR Guidance Document Schematic



3. BACKGROUND INFORMATION ON (Q)SAR

BACKGROUND INFORMATION ON (Q)SAR

Topics Discussed in this Section:

- **Definition of (Q)SAR**
- **Types of (Q)SAR tools and approaches**
- **Importance of data quality in (Q)SAR model development**
- **Importance of mode/mechanism of action in (Q)SAR model development**
- **Examples of (Q)SAR tools and their applications**

3.0 Introduction

The purpose of this section is to provide some brief background information on the definition of (Q)SAR, types of (Q)SAR tools and approaches and some key issues associated with the development of (Q)SAR tools. In particular, the importance of data quality and mode/mechanism of action in the development of (Q)SAR models is highlighted. Also, while (computerized) (Q)SAR models are frequently cited in examples elsewhere in this document, this section illustrates that (Q)SAR actually consists of a range of tools and approaches.

3.1 Definition of (Q)SAR

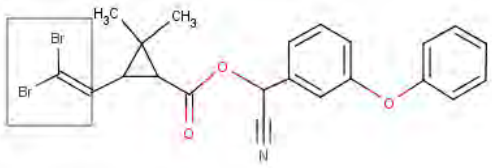
(Q)SAR is the study of the correlation between chemical structure and associated biological activity, with the ultimate goal of predicting the activity of untested chemicals based on structurally related compounds with known activity (Cronin, 2010). Structure-activity relationships (SARs) are qualitative relationships, often in the form of structural alerts that incorporate molecular substructures or fragments related to the presence or absence of activity (Dearden *et al*., 2009). Quantitative structure-activity relationships (QSARs) attempt to quantify the relationship between an aspect of chemical structure and an activity or property imparted by that structure. Chemical structure is often described by descriptors (*e.g.,* electrophilicity, hydrogen bonding, molecular fragments) or physical-chemical properties (*e.g.,* Log P) which are then used to develop a mathematical correlation between a group of structures and a defined activity or endpoint. The mathematical correlations usually take the form of

statistical algorithms developed through a variety of techniques (*e.g.,* univariate regression, multiple linear regression, partial least squares analysis).

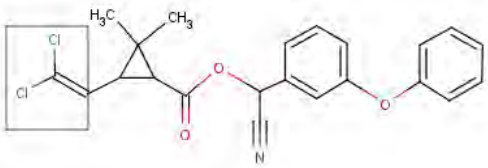
3.1.1 Defining Similarity

Structurally similar chemicals or structural analogs usually have similar chemical structures but with one or more atoms or groups of atoms replaced with other atoms or groups of atoms. Figure 3-1 lists the chemical structures of two pyrethroid insecticides, deltamethrin and cypermethrin. These two structural analogs share a cyclopropane carboxylic acid substructure that is common to most pyrethoid structures.

Figure 3–1: Example of Structural Analogs



Deltamethrin



Cypermethrin

Listed below are some common criteria used to identify structurally similar substances. Many of these have been proposed by the OECD and the US EPA as a basis for building chemical categories (OECD, 2007a; US EPA, 1999).

- a common functional group or sub-structure (*e.g*., phenols, aldehydes)
- a common precursor or break-down product, which can result from structurally-similar chemicals; this approach can be used to examine related chemicals such as acids/esters/salts. (*e.g*., short-chained alkyl-methacrylate esters which are metabolized to methacrylic acid)
- an incremental or constant change in a chemical structure (*e.g*., increased carbon chain lengths; typically used for physicochemical properties such as boiling point)
- common constituents or chemical classes, such as similar carbon range numbers, often used with “substances of unknown or variable composition, complex reaction products or biological material” (UVCBs)
- functionally similar chemicals or functional analogs that have similar biological activities (*e.g.,* toxicity endpoints, pesticidal mode of action) or physical-chemical properties (*e.g.,* Log P, solubility, vapour pressure). Note that

functional analogs are not necessarily structural analogs and vice versa (Saliner *et al*., 2005; Russom *et al*., 1997).

Table 3–1 lists examples of pesticidal modes of action for several examples of insecticides. Most pesticidal modes of action include more than one chemical class. Consequently, 'similarity' can be based on at least three aspects of a pesticide, (i) pesticidal mode of action (*e.g.,* acetylcholinesterase inhibition), (ii) pesticide classification (*e.g*., insecticide) and (iii) chemical classification/common functional group (*e.g*., carbamate, organophosphate, etc.). Therefore, a weight of evidence approach can be important when defining similarity for the purpose of developing (Q)SAR tools and approaches. As discussed in example 5, Appendix III, information on the pesticidal mode of action and structural similarity can be combined with pharmacokinetic and empirical animal study results in a weight of evidence approach in pesticide assessments.

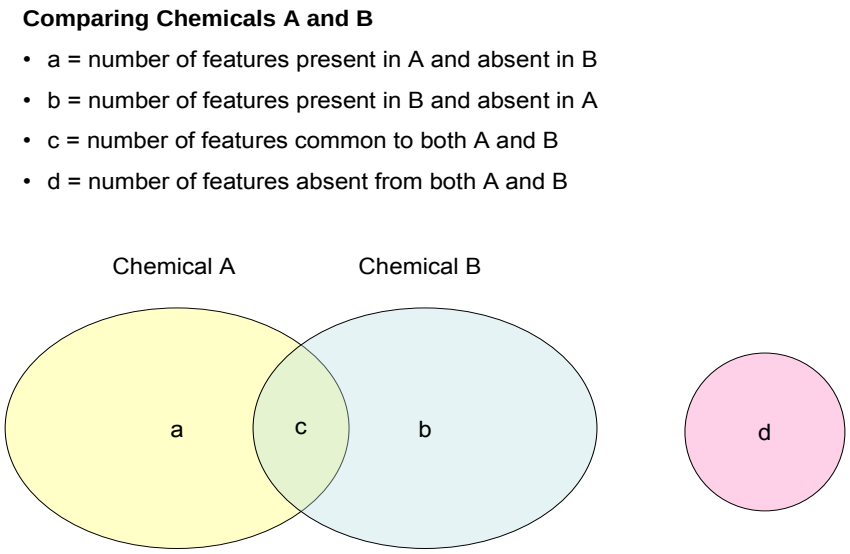
Table 3–1: Pesticidal Mode of Action and Associated Chemical Class for a Select Group of insecticides (adapted from Insecticide Resistance Action Committee (IRAC); http://eclassification.irac-online.org/)

Pesticidal Mode of Action	Chemical Class
Acetylcholine esterase inhibitor	Carbamates
	Organophosphates
GABA-gated chloride channel antagonists	Cyclodiene organochlorines
	Phenylpyrazoles (Fiproles)
Sodium channel modulators	Pyrethroids
	Organochlorines
Nicotinic acetylcholine receptor agonists	Neonicotinoids
Juvenile hormone mimics	Juvenile hormone analog
	Carbamates
	Pyridine insect growth regulator

A common mathematical approach to defining structural similarity is the use of algorithms or similarity indices that calculate similarity based on pattern

matching. These estimation tools rank chemicals based on (structural) characteristics or features of each chemical that are similar (match/overlap), and features that are dissimilar (mismatch/difference) (Saliner *et al*., 2005; Monev, 2004.). Figure 3–2 provides a schematic of the measures that can be described in similarity indices. Similarity indices can utilize two- or three-dimensional structural information and examples include correlation-type indices (*e.g*., Tanimoto Index (also known as Jaccard coefficient), Hodgkin Ricards Index, Cosine-similarity index), dissimilarity measures (*e.g*., Euclidean distance index, Hamming distance), and composite measures of similarity and dissimilarity (*e.g*., Hamann measure, Yule measure). For an overview of these approaches see Saliner *et al*., 2005; Monev, 2004; and Urbano-Cuadrado *et al*., 2008. It is also important to know that a high degree of similarity based on mathematical similarity indices does not necessary indicate there are similarities in the mode of action (MOA) for the concerned effects.

Figure 3-2: Measures that can be Described in Similarity Indices



3.2 Types of (Q)SAR Approaches

Although the term (Q)SAR is often used to refer to predictive models, especially computer-based models, it should be noted that (Q)SAR is actually inclusive of a wide variety of tools and approaches such as analogs, chemical categories and computer-based or non-computer based SAR/QSAR models. A brief overview of these tools and approaches is provided below.

3.2.1 Analogs

Analog approaches have traditionally involved predicting an endpoint or property of one chemical based on the available data for the same endpoint or property of a similarly structured chemical (OECD, 2007a). An example of an analog technique is bridging or extrapolating the results of toxicological studies on a parent pesticide compound to a metabolite or transformation product of that same parent pesticide. When using an analog approach for bridging from a parent pesticide to metabolites or transformation products, it is important to have sufficient evidence to link a particular substructure or substructures to the toxicity endpoint of interest, and that the substructure is conserved from the parent pesticide to the metabolite or transformation product.

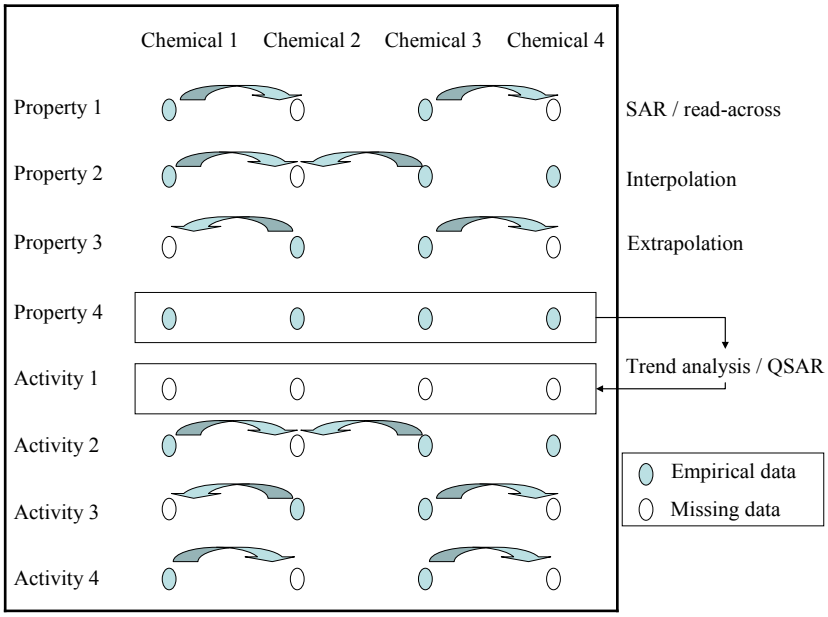
3.2.2 Chemical Categories

A chemical category is defined as a group of substances with physical-chemical, human health, or ecotoxicological attributes that are similar or follow a pattern as a result of structural similarity (OECD, 2007a). As discussed in section 3.1.1, the US EPA and the OECD have identified a number of ways of identifying similar chemicals for the purpose of building categories. Both agencies have also developed a consistent approach for defining chemical categories (OECD, 2007a; OECD, 2009; US EPA, 1999). Chemicals within a category are not required to be similar with respect to all properties, and a substance can belong to more than one chemical category. In most instances, chemical category approaches are based on a weight of evidence, considering multiple lines of information from many tested chemicals and inferring information for an untested substance.

When using the chemical category approach, it is common to construct a matrix table as depicted in Figure 3–3. The matrix consists of chemical category members in each of the columns and corresponding sets of properties and/or activities in each of the rows. The solid dots are properties/activities for which reliable data exist and the hollow dots are data gaps. Data gap filling in categories can be done using techniques such as read-across, interpolation, extrapolation, and trend analysis (see examples in Figure 3–3). Read-across is estimating the activity/property for one untested chemical from a tested chemical or chemicals. Read-across can be qualitative or quantitative. Interpolation is the estimation of a property/activity for a data poor category member based on existing data from other category members on both sides of the data poor chemical in the matrix. Extrapolation is estimating an activity/property for a chemical that is near or at the boundary of the category based on data for other category members. Extrapolation is more prone to error than interpolation,

especially when the boundary of the category is difficult to define. The observation of a quantitative trend (increasing, decreasing, or constant) in the experimental data for a given endpoint across chemicals in a category can also be used as the basis for interpolation or extrapolation (*i.e*., trend analysis). In addition, it is possible to develop a QSAR within a chemical category by plotting the activities versus the properties of chemicals with empirical data. By using a combination of tools, *i.e*., read-across, trend analysis and (Q)SAR, the matrix of properties/activities for chemicals under consideration can be rendered less uncertain through the greater use of existing data (OECD, 2007a).

Figure 3–3: A Schematic of a Chemical Category Matrix Table (modified from van Leeuwen *et al*., 2009)



Chemical category approaches have been used for assessing chemicals with data gaps by the US EPA's OPPT, in the US EPA HPV Challenge Program, under the REACH legislation, and in OECD SIDS program (van Leeuwen *et al*., 2009). Additional examples of categories can be found in Enoch, 2010; Enoch *et al.*, 2009; US EPA, 1999; and Worth and Patlewicz, 2007.

3.2.3 (Q)SAR Models

(Q)SAR models generally refer to computerized systems developed to predict activities or properties of chemicals using SAR or (Q)SAR methods. There are numerous commercial (*i.e.,* available for a fee) and non-commercial (*i.e.,*

freeware) models available for predicting human health related and environmental activities, physical-chemical properties, and other parameters.

SAR models generally follow a process of identifying active and inactive chemicals based on the presence or absence of specific structural features. For example, SAR/expert systems use decision logic to categorize potential activity of untested chemicals based on expert knowledge gathered from the analysis of data on tested chemicals. Some systems use a series of questions that the user responds to or the system automatically responds to. The questions may be based on databases of structural alerts or chemical parameters known to be associated with biological activity and can capture multiple types of interactions within a specific biological system.

(Q)SAR models usually consist of computerized mathematical correlations (*i.e.,* algorithms) that relate descriptors of chemical structure or physical-chemical properties to an activity or property to be predicted. The descriptors or physical-chemical properties for a chemical of interest may be input by the user or generated by the model and then used in the algorithm to make a prediction. For example, in a QSAR model developed to use the octanol and water partition coefficient (K_{ow}) to estimate the permeability coefficient through the skin for a chemical in water (K_p), the K_{ow} is the descriptor which can be a measured value or estimated by the model (US EPA, 2007). QSAR models can produce qualitative predictions of activity/inactivity or quantitative (continuous) values related to biological activity (*e.g.,* receptor binding affinity, acute oral LD_{50} in rats, etc.) or other parameters (*e.g.,* bioconcentration factor). QSAR models generally rely on data for many chemicals (*i.e.,* training sets) for the development of the algorithms used to predict the activity of a single chemical lacking data.

In addition to classifying (Q)SAR models as relying on SAR versus QSAR approaches, they can also be considered in terms of statistical versus mechanistic approaches and global versus local approaches. In general, statistically-based (Q)SAR models rely on a statistical association between structure and activity, can be developed objectively with little mechanism of action expertise, are useful for detecting structural features/molecular descriptors predictive of toxicity, but may be noisier and tend to perform poorer for endpoints with multiple mechanisms. Mechanistically-based models can focus on key features that provide more clear-cut relationships and mechanistic backing but generally require considerable expert knowledge of the relationship between mechanism of action and descriptors of chemical structure, may be subjective, and could have high levels of uncertainty if the mechanism is unclear or

presumptive. Ideally, (Q)SAR models should strive to achieve statistical association but have a mechanistic foundation.

Local QSAR models are generally developed for individual classes of chemicals. Their training sets usually consist of highly structurally homogeneous or congeneric chemicals or classes of chemicals with similar known biological activity/function (*e.g*., peroxisome proliferators). Local models require fewer training set chemicals and tend to perform better, presumably because they are more likely to focus on a single mechanism of action. However, they are often limited in scope to a small subset of narrowly defined chemicals. Global models are generally derived using training sets of structurally heterogeneous or non-congeneric chemicals. Due to the diversity of the training set chemicals, these models often cover a range of different mechanisms of action, usually resulting in poorer predictive performance than local models, unless the training sets are subdivided based on mechanism of action. Global predictive models tend to be more adept in discovering new insights, but may be more likely to yield incorrect results if the predicted chemical structure is not well-represented in the training set. Several publications have investigated the ability of global and local (Q)SAR approaches to fulfill regulatory requirements (*e.g*., EC , 2010; Yuan *et al*., 2007; and Worth *et al*., 2011).

3.3 Importance of Data Quality in (Q)SAR Model Development

Developing (Q)SAR models depends on experimental data, molecular representation (2-D or 3-D structures), availability of chemical descriptor or parameter data (measured or calculated) associated with structure, and fitting relationships (*e.g*., algorithms) to the data (Bradbury *et al*., 2003; Perkins *et al*., 2003; Tong *et al*., 2003; Walker *et al*., 2003). Among these factors, experimental data are generally the most important determinants of the accuracy of the predictions from (Q)SAR models as the confidence in a model can be no greater than the understanding of, and confidence in, the underlying data.

In (Q)SAR model development, usually a set of chemicals with reliable data are collected for a particular biological/chemical activity. Typically the original test data are randomly separated into a training set and a validation set, with the training data set used to develop a model and the validation data set used to test the assumptions that the model works for chemicals not involved in the development of the original model (Leonard and Roy, 2006).

Model training sets can be assembled prospectively or retrospectively. In the prospective approach bioassays are developed and optimized for testing the type of chemicals for which the (Q)SAR predictions are needed. Mechanistic

information on the training set chemicals can also be obtained prospectively. In the more commonly used retrospective approach data are collected from readily available sources (*e.g*., the open literature). This often results in noisier predictions because of the lack of control or consistency in study protocols, interpretation criteria, etc., although this can be compensated for by evaluating the available data and selecting only consistent higher quality studies (*i.e.,* chemical identity/form confirmed, concentrations/purity measured, standard test protocols, assays optimized for the type of chemicals, etc.). Also, it is important to verify that the identities of the chemicals in the training set correspond to their structural representations used in the predictions. Sometimes information on the metabolism of the chemicals tested and mechanisms of action can also be obtained retrospectively to help enhance the interpretability of the predictions.

The importance of data quality in (Q)SAR model development is also discussed further in section 5.1.1 of this document.

3.4 Importance of Mode/Mechanism of Action in (Q)SAR Model Development

An understanding of a chemical's mode/mechanism of action is highly sought when developing (Q)SARs. Mode/mechanism of action considerations can help in the selection of appropriate molecular descriptors or physicochemical properties that are associated with activity, determination of whether the training set is applicable to the chemicals to be predicted, separation of the training set into more mechanistically homogenous groups to help improve predictive performance, and the interpretation of model outliers. An understanding of mode/mechanism of action can also provide support for predictions, help in the assessment of the human significance of predictions of toxicity in laboratory animals, and help to identify and prioritize additional testing to fill data gaps.

One example of the utility of mode/mechanism of action data in (Q)SAR model development is a study by Russom et al. (1997). A diverse dataset of more than 600 chemicals was divided into mechanistic groups prior to developing (Q)SARs for fathead minnow acute toxicity. Combining all of the chemicals into a single training set would have resulted in a much poorer correlation of the LC_{50} values to the chemical parameter, log $K_{ow,}$ which in turn, would have resulted in much poorer predictive performance.

An understanding of mechanistic considerations is especially important for complex biological systems which may have metabolism and chemical kinetics adding to the complexity. In general, the less complex the biological system, the greater the confidence that the structure of the chemical is directly related to the

observed activity and that the relationship can be reliably modeled. For example, *in vitro* systems (*e.g.,* Ames mutagenicity) are often less complex and more reliably modeled than many *in vivo* systems (*e.g.,* carcinogenicity, teratogenicity). However, this is not always the case as *in vivo* fish acute toxicity LC_{50} values are well predicted for several modes of action because chemical concentration in water is a good surrogate for chemical activity in the blood (MacKay *et al*., 1983). Also, if *in vitro* systems include metabolic components, their complexity for (Q)SAR development will increase.

The importance of mode/mechanism of action information in (Q)SAR model development is also discussed further in section 5.1.1.5 of this document.

3.5 Examples of (Q)SAR Tools and their Applications

This section is not intended to provide an exhaustive overview of computational tools available via government, open access, or commercial sources, but rather an overview of the types of tools that currently exist.

Several reviews have been written on the types of tools available (EC, 1995a,b; Hulzebos *et al*., 1999; Jensen *et al*., 2008; Pavan *et al*., 2005a,b; Rorije and Hulzebos, 2005; Tsakovska *et al*., 2005, 2008), but it should be kept in mind that the inventories of available tools is constantly changing with emerging research in this area.

With the development of Simplified Molecular Input Line Entry System (SMILES) notation (Weininger, 1988) as a means to identify structure information in a computer readable format, and the advancement of desk top computing in the 1970's, (Q)SAR tools have become more readily accessible to risk assessors (Benfenati, 2007). Although initially (Q)SAR approaches were primarily used in the drug and pesticide discovery and development fields, these techniques became especially important to regulatory risk assessment after the promulgation of the Toxic Substances Control Act (TSCA) (Zeeman *et al*., 1995). The use of QSARs in assessing potential toxic effects of organic chemicals on ecologically relevant species and humans evolved as computational efficiency and toxicological understanding advanced, and in many cases has proved to be scientifically-credible for use in estimating toxicity for substances with little or no available empirical data (OECD, 2007b).

(Q)SAR models also exist for specific endpoints such as skin sensitization (Patlewicz *et al*., 2008), eye irritation (Tsakovska *et al*., 2005), acute toxicity and repeated-dose endpoints for mammalian species (Tsakovska *et al*., 2008), bioaccumulation (Arnot and Gobas, 2004), mutagenicity and carcinogenicity

(Benfenati *et al*., 2009; Benigni *et al*., 2007a,b), estimating physical chemical properties (EC, 1995a,b; Deardon and Worth, 2007), toxicity to aquatic species (EC, 1995a,b; Netzeva *et al.*, 2007; Pavan *et al*., 2005a,b), and reproductive toxicity (Jensen *et al*., 2008).

Software applications are available for assisting in the identification of chemical similarity (Gallegos-Saliner *et al*., 2008; Patlewicz *et al*., 2005). In addition, the availability of state-of-the-art software programs for use in the development of QSAR models from any data set allows one to generate models at their desk top for endpoints/applicability domains not covered by existing models (see the series of FDA papers as an example: Matthews *et al*., 2009 a, b; Matthews *et al*., 2007 a,b; Ursem *et al*., 2009). Key to this type of analysis is a high quality, structurally-annotated data set for use in the development of models (Judson *et al*., 2009; Richard *et al*. 2006, 2008; Williams *et al*., 2009). Another important aspect of many risk assessments is metabolism/degradation products, and (Q)SAR tools to simulate metabolism have been developed to assist in identifying these products (Dimitrov *et al*., 2005a,b; Mekenyan *et al*., 2006; Ringeissen, *et al*., 2010).

(Q)SAR approaches can be used to better inform testing strategies via screening, prioritization, and ranking of large chemical inventories based on receptor binding (Jensen *et al*., 2008; Klopman and Chakravarti, 2003 a,b; Schmieder *et al*., 2004), human health endpoints (Demchuk *et al*., 2008; Klopman *et al*., 2003; Ruggeri, 2009), and environmental toxicity, fate, and persistence (Brown and Wania, 2008; Daginnus *et al*., 2009; Walker *et al*., 2004). These rankings can be used for a variety of risk assessment purposes including developing chemical categories, identification of PBT (persistent, bioaccumulative and toxic) substances, and risk characterization (Pavan and Worth, 2008). Similarly, (Q)SAR tools have been investigated in combination with physical-chemical data and read-across to improve the application of TTC methods (Bassan *et al*., 2011; Worth *et al*., 2011).

Under REACH, information on models that meet the OECD validation principles and are proposed for use in filling data gaps are currently being gathered. A searchable catalog of all models including background information required to validate the models, authors/source of model, related publications, endpoint estimated and related experimental protocol, algorithm with training set and validation set, including all input variables for the models can be found at the following website: http://qsardb.jrc.ec.europa.eu/qmrf/index.jsp. Some actual example cases are listed in Appendix III.

3.6 Summary

(Q)SAR tools and approaches involve the study of correlations between chemical structure and associated biological activity, physical-chemical properties or other properties, with the ultimate goal of predicting the activity or properties of untested chemicals using available data from structurally-related compounds. While frequently associated with computerized models, (Q)SAR tools actually encompass a wide range of approaches such as analogs, chemical categories, and computer or non-computer based SAR and (Q)SAR. The development of reliable (Q)SAR models depends upon a number of factors, among which, experimental data are probably the most important. In particular, data quality and a good understanding of the available information on mode/mechanism of action can contribute to the confidence in (Q)SAR model predictions. Types of endpoints or properties from a pesticide context that can be predicted using (Q)SAR and related methods include *in vivo* ecotoxicity and human health-related toxicity endpoints, specialized *in vitro* endpoints, metabolism, physical-chemical parameters, and environmental fate parameters. While this document is not intended to recommend or endorse individual (Q)SAR tools, it is recognized that there are currently a variety of computerized and non-computerized, commercial and non-commercial (Q)SAR tools for predicting the endpoints or properties described above. Sections 2 and 3 of this document were designed to provide a brief introduction and background information on (Q)SAR tools and approaches. The subsequent sections of this document (4, 5, 6, and 7) focus on issues associated with applying (Q)SAR predictions to pesticides including problem formulation and (Q)SAR (section 4), evaluating the adequacy of (Q)SAR predictions (section 5), combining information from multiple predictions (section 6) and incorporating (Q)SAR into weight of evidence assessments (section 7).

4. Problem Formulation and (Q)SAR

PROBLEM FORMULATION AND (Q)SAR

Topics Discussed in this Section:

- **Assessment context that (Q)SAR is being applied to**
- **Characteristics of the pesticide that is the subject of the prediction**
- **Characteristics of the (Q)SAR tool and the prediction**
- **Available empirical data including information on mode of action**

4.0 Introduction

Problem formulation is an important initial step for framing the specific question(s) to be addressed in assessments of human health and environmental risks from pesticides. In its *Guidelines for Ecological Risk Assessment*, the US EPA has indicated that problem formulation involves the on-going integration of the available information that eventually leads to three products: assessment endpoints, a conceptual model of the risk to be investigated, and an analysis plan (US EPA, 1998).

Since guidance on the general problem formulation process for the risk assessment of chemicals such as pesticides has been outlined in other published documents (*e.g.*, US EPA, 1998; Doull *et al*., 2007), the details of that guidance will not be discussed here. Instead, this section will focus on the preliminary analysis of (Q)SAR predictions as one of the several potential sources of information to be integrated at the problem formulation stage. Preliminary analysis of (Q)SAR prediction for a pesticide at the problem formulation essentially involves answering the following questions:

- What is the assessment context that the (Q)SAR prediction is being applied to?
- What are the characteristics of the pesticide that is the subject of the prediction?
- What are the characteristics of the (Q)SAR tool and the prediction?
- What empirical data are available including any information on mode of action?

Answering these questions at the problem formulation stage may enable an evaluator to immediately determine that a prediction is not suitable or relevant for addressing the specific pesticide risk assessment question. Alternatively, these questions may lead to a more in-depth evaluation of whether the (Q)SAR prediction is adequate or “fit for purpose” (see section 5) and eventually to the consideration of how the results of a fit for purpose prediction could be incorporated into an overall weight of evidence decision (see section 7).

4.1 Assessment Context that (Q)SAR is being Applied to

Identifying the assessment context for a (Q)SAR prediction involves understanding why the prediction is being considered for the assessment of a pesticide and the specific endpoint or property that the prediction is intended to address. Both of these points will assist the evaluator in determining whether a (Q)SAR prediction should be considered in a pesticide assessment and if yes, what will be an acceptable level of reliability and uncertainty associated with the use of (Q)SAR.

(Q)SAR predictions are generally used to try to gain some insights into the toxicity, ecotoxicity, behavior in the environment or other aspects of a pesticide in the absence of empirical data. Consideration of a (Q)SAR prediction for the premarket assessment of a pesticide would likely involve one of the following scenarios: 1) submission of a (Q)SAR prediction by a registrant to address a data requirement or as supporting evidence for a data requirement for pesticide, a metabolite or a transformation product, or 2) use of a prediction by an evaluator to identify or support a data requirement for a pesticide, metabolite or transformation product.

In the first scenario, an applicant would likely submit a (Q)SAR prediction or predictions as a replacement for or as supporting evidence to waive a requirement for a specific type of empirical data (*e.g.*, to address a requirement for acute irritation toxicity data). In most cases, using a (Q)SAR prediction as a stand alone replacement for a data requirement is not likely to be acceptable depending on the nature of the endpoint and the specific policies of the pesticide regulatory agency. Combining a (Q)SAR prediction with other types of data to support a waiver request may be more acceptable depending on what other types of data are available, the reliability and level of uncertainty for the (Q)SAR prediction, the overall scientific defensibility of the rationale, and regulatory agency policies.

The second scenario could involve using a (Q)SAR prediction to justify a requirement for a study not normally included in regulatory data requirements for

pesticides or to justify a requirement for a study on a metabolite or transformation product for which no data have been submitted. This would also include cases in which (Q)SAR predictions are used as supporting information when questioning the reliability of experimental data, leading to a requirement for the submission of more reliable studies. Criteria for what constitutes a reliable (Q)SAR prediction and acceptable levels of uncertainty would likely be less stringent for scenarios in which (Q)SAR predictions are used to drive data requirements compared to cases where (Q)SAR is used to support waiving data requirements.

While these premarket scenarios are likely to be the most frequent applications of (Q)SAR, there may also be instances where (Q)SAR tools could be used post-market such as the toxicity characterization of a novel impurity (*e.g.*, leachable/extractable) not originally characterized during the pre-market approval process.

Endpoints or properties that can be predicted by (Q)SAR and could be relevant to pesticide assessments include toxicity (*e.g.*, carcinogenicity, developmental toxicity), metabolism, ecotoxicity (*e.g.*, fat head minnow LC_{50}, longer-term toxicity in terrestrial species), other biological activities (*e.g.*, estrogen receptor binding), and physical-chemical properties (*e.g.*, Log K_{ow}, partition coefficients, bioaccumulation factor). The type of endpoint and whether it is a critical data point for a pesticide assessment (*e.g.*, used for a point of departure analysis) will have an influence on how reliable a (Q)SAR prediction should be (see section 5.4). For example, it may be possible to accept a less reliable prediction for an acute toxicity endpoint used as supporting information for labeling requirements compared to a predicted NOAEL for chronic toxicity that is to be considered in a point of departure analysis. Furthermore, the use (Q)SAR predictions to address critical endpoints in pesticide risk assessments would likely require much more detailed analyses of whether the predictions are fit for purpose compared to predictions generated for non-critical endpoints.

4.2 Characteristics of the Pesticide that is the Subject of the Prediction

Understanding the characteristics of the pesticide that is the subject of the (Q)SAR prediction is critical to determining whether the pesticide is correctly identified; whether the prediction is to be made for an active ingredient, a metabolite or a transformation product; whether an accurate structural representation is available; or whether it is even possible to generate a prediction for the pesticide in question.

4.2.1 Chemical Identifiers and Mixtures

Examples of the types of pesticides for which a (Q)SAR prediction may be required include discrete substances; individual isomers or mixtures of isomers; crystalline structures (*e.g.,* minerals); substances with unknown or variable composition, complex reaction products and biological materials; polymers; other mixtures or formulations; and complex salts and metal-containing compounds. Therefore, accurate information on the identity, composition and structure of a pesticide is critical to determining whether a prediction was based on a correct structure. Confusion can result when common or trade names are applied to multiple isomers of a pesticide, salt forms, acid/base forms or polymeric and monomeric forms. The use of more precise chemical nomenclature (*e.g.*, International Union of Pure and Applied Chemistry (IUPAC)) can assist with more accurate identification (IUPAC, 2010). While Chemical Abstract Service (CAS) numbers (American Chemical Society, 2010) are frequently used as unique identifiers for pesticides, in some cases they may actually represent isomer mixtures, polymers, and unknown or variable composition substances rather than discrete, single chemicals, so it may be necessary to review the CAS number to clearly determine which structure(s) it actually represents.

In general, mixtures cannot be run through (Q)SAR models, nor can synergistic or antagonistic effects of chemicals in mixtures be accounted for because models typically use single, discrete chemical structures as input. For mixtures of discrete organic chemicals, one option may be to make separate predictions for each chemical and compare and contrast the results. Alternatively, if one component of a mixture is predominant, in some cases that component may be used to represent the entire mixture. However, for pesticides with variable compositions, (*i.e.*, oligomers, natural fats, or mixtures that change composition depending on reaction conditions) evaluators should be aware that (Q)SAR predictions generated using a representative structure may not accurately reflect the true nature of the material used in the pesticide application.

4.2.2 Transformation, Degradation, and Metabolism

A number of pesticides are reactive chemicals that can be readily transformed in the environment or in the body (*e.g.*, hydrolyzable acid halides, isocyanates, etc.). Transformation products may have dramatically different toxicity profiles than the original pesticides and need to be considered when identifying the correct structures for (Q)SAR predictions. Similarly, information on potential environmental degradates and metabolic by-products of pesticides in livestock, food plants or the body should also be considered as the toxicity may not reflect the parent pesticide, but rather a reactive intermediate, degradate or metabolite.

Information on degradates and metabolites may be available from empirical pharmaco/toxicokinetic studies or from *in silico* models of potential metabolic or degradation pathways. If degradates and metabolites are identified, it will also be important to consider their stability and in the case of (Q)SAR predictions of metabolites, the likelihood that they could occur in vivo. Individual pesticide regulatory agencies have specific criteria they use to identify probable, stable metabolites.

When (Q)SAR models are developed, available information on the metabolism of the training set compounds should be taken into consideration whenever possible. Training set compounds that require metabolic transformation prior to inducing a specific endpoint are likely to generate models that are unreliable if those models are constructed based on the parent structures alone. Failure to consider the structure of a metabolite could lead to an inaccurate assessment of the chemical features or properties associated with the predicted endpoint, errors in analog selection, problems with characterization of similarity based on mode of action, errors in inter-species extrapolation when metabolic differences exist between species, and ultimately poor predictive performance. Another aspect of metabolism that may need to be considered during (Q)SAR model development is differences in metabolism from different routes of exposure. While the industrial chemical, bis-(chloromethyl) ether is one of the most potent human and animal respiratory carcinogens, it is not expected to be carcinogenic via the oral route because it hydrolyses in seconds upon contact with aqueous solution (Woo and Lai, 2010; ATSDR, 1989). Consequently, (Q)SAR predictions for direct acting reactive pesticides that are used to support data requirements for inhalation toxicity should be treated with caution if they come from (Q)SAR tools whose training sets only include analog substances tested via the oral route. Finally, the results of *in vitro* tests can also be impacted by metabolic transformations, so information on the degree to which metabolic capability is incorporated into *in vitro* toxicity tests should be considered when constructing models for those tests.

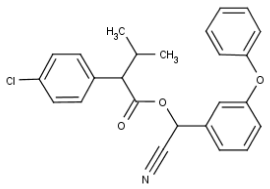
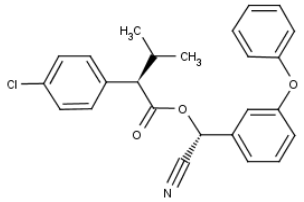
4.2.3 Isomers and Structural Representations for (Q)SAR

A pesticide's three-dimensional molecular structure or shape and its molecular conformation can influence properties such as absorption, distribution, and excretion, as well as enzyme or receptor binding, and the resulting differences can readily impact toxicity profiles. Consequently, the isomeric form of a pesticide is another important piece of information to consider when determining whether a prediction is based on the correct structure and whether the (Q)SAR tool is applicable to the structure in question.

Examples of isomeric forms to consider include stereoisomers that differ in their spatial orientation of atoms. The pyrethroid insecticide fenvalerate is a racemic mixture of stereoisomers (*i.e.*, R/S enantiomers) of a chiral active ingredient, although the S-isomer in the mixture (esfenvalerate) has the greatest insecticidal activity (WHO/FAO, 1996) (see Figure 4–1). Because these different isomers have the same molecular formula, molecular weight, and physical-chemical properties, it can be difficult for some (Q)SAR models to distinguish them, especially models that do not take stereoisomerism into account.

Many regulatory agencies make the conservative assumption that stereoisomers will have similar mammalian toxicity and ecotoxicity, unless data are available to demonstrate the contrary. In addition to isomeric forms, the position of flexible groups in a molecule can also be important as relatively free rotation of attached groups about single bonds can influence the conformation of a molecule and determine the overall molecular size, especially in complex molecules with multiple rotation points. More advanced (Q)SAR techniques may employ three-dimensional molecular descriptors to account for rotation of flexible groups and other characteristics, but calculating these descriptors can be complex and time consuming.

Figure 4–1: Fenvalerate Racemic Mixture

	
Fenvalerate (RS)-alpha-Cyano-3-phenoxybenzyl (RS)-2-(4-chlorophenyl)-3-methylbutyrate CAS No. 51630-58-1	Esfenvalerate (S)-alpha-Cyano-3-phenoxybenzyl (S)-2-(4-chlorophenyl)-3-methylbutyrate CAS No. 66230-04-4

In addition to understanding a pesticide's three dimensional structure and conformation, the method of entering structures into a (Q)SAR model should also be taken into consideration. Some of the more common structural entry options that have been historically employed for single structure entries include the SMILES (simplified molecular input line entry system), International Chemical Identifier (InChI™) codes, the MDL Mol file (MOL), and various drawing applets

and molecular editors (Daylight Chemical Information Systems, 2008; IUPAC, 2010b; Dalby *et al.,* 1992). For multiple (batch) chemical entries, the Structure Data Format (SDF) file and SMILES (SMI) file formats are commonly used (Dalby *et al.,* 1992). These structural entry methods have strengths and limitations, and in some cases, it may be necessary to verify the accuracy of the structural representations from these methods to ensure that correct structures are used for predictions.

4.3 Characteristics of the (Q)SAR Tool and the Prediction

Prior to considering and weighting the results of an empirical study in the assessment of a pesticide, it is necessary to obtain and evaluate the details of the study protocol and how the study was conducted, as well as the results of the study and how they were interpreted. Similar concepts apply to the use of (Q)SAR predictions in pesticide assessments, as the characteristics of the tools used to make the predictions and the predictions themselves need to be understood and evaluated before weighting the predictions in an assessment.

Many of the concepts discussed in this section overlap with the evaluation of the scientific validity of a (Q)SAR tool as discussed in section 5.1. However, at the problem formulation stage, it is intended that the evaluator will gain a basic understanding of these issues. This may enable an immediate decision that the (Q)SAR prediction is not adequate for the assessment context or it could lead to a more detailed evaluation as discussed in section 5.1, especially with respect to the application of the OECD (Q)SAR validation principles (section 5.1.1).

A starting point for characterizing a (Q)SAR tool at the problem formulation stage is a sufficient understanding of the general methodology behind the tool. Is the tool based on simple analog extrapolations, read-across or trend-analysis approaches using chemical categories, a structural alert/rule based SAR/expert system, a statistical (*e.g.*, regression based) QSAR derived from a specific database of chemicals and their descriptors or some other method? Each of these methods has strengths and limitations that can influence how they should be interpreted and the reliability of predictions from them. For instance, SAR/expert systems based on structural alerts may be supported by expert reviews of relevant research, and can include a mechanistic rationale to support predictions. However, in some cases these systems do not include structural alerts associated with inactivity, may have limited databases of alerts, and may not have a clearly defined domain of applicability. Statistical QSAR models based on training sets of active/inactive chemicals and descriptors of chemical structure may provide insights into associations between specific structures and activity that were not previously investigated, help to identify structures that modify or

eliminate specific activities, and may be capable of generating quantitative predictions (*e.g.*, probabilities or specific numerical values) rather than dichotomous active/inactive (yes/no) predictions. However, in some cases QSAR models may overemphasize statistical associations in the absence of mechanistic rationales, their domains may be restricted by the structural diversity in their training sets, and their training sets may include chemicals with a variety of different mechanisms which can result in poor predictive performance and/or considerable uncertainty in their predictions. A number of reviews of the strengths and limitations of (Q)SAR models are available in the scientific literature (*e.g.*, Hulzebos *et al.*, 2001; Greene, 2002).

Gaining an understanding of the empirical data from which the (Q)SAR tool was derived is another important starting point for determining whether a (Q)SAR prediction is likely to be relevant to a pesticide assessment. It may be possible to quickly discount (Q)SAR tools derived from studies based on outdated protocols not conducted according to GLP standards, based on endpoints that are vague or inconsistent, interpreted according to non-standard criteria, involving chemicals significantly structurally dissimilar to the pesticide of interest, and/or obtained from non-peer reviewed sources. On the other hand, (Q)SAR tools based on higher quality empirical data may be subjected to a more detailed evaluation and potentially included in a weight of evidence assessment.

As discussed in section 4.1, information on the endpoint on which a (Q)SAR tool is based can be one of the important factors to consider at the problem formulation stage for determining whether a (Q)SAR prediction will be relevant to the specific pesticide assessment context. In particular, because many pesticide assessment questions involve quantitative toxicity values (*e.g.*, $LD_{50,}$ $EC_{50,}$ NOAEL, etc.) for identifying labeling requirements, and calculating margins of exposure, reference doses, etc., it is important to determine whether a (Q)SAR tool can generate quantitative or qualitative predictions, and if quantitative, the type of value predicted. A qualitative yes/no prediction for chronic toxicity will not be particularly useful if a prediction of a NOAEL is required to derive a regulatory point of departure for a pesticide. Alternatively, a (Q)SAR model that only predicts quantitative LOAELs for short-term endpoints may also have limited applicability. As mentioned above, predicted endpoints that are somewhat vague such as general developmental toxicity potential may not be specific enough to address questions about endpoints such as post implantation loss, developmental delays, fetal dysmorphogenesis, etc. An overall point to consider is whether there is likely to be sufficient, high quality empirical data available on an endpoint of interest so that (Q)SAR tools could be developed that are relevant to a particular assessment context.

Investigating other details of the (Q)SAR tool used may also assist in determining the relevance of a (Q)SAR prediction to a pesticide assessment during problem formulation. For example, details on a (Q)SAR model such as the specific name of the model, version number, date it was developed, and contact information for the developer can be important for determining the relevance of a (Q)SAR prediction. Model developers can make significant changes from one version to another such as increasing the number and diversity of the chemicals in the training set, modifying the library of descriptors or structural alerts, and modifying model algorithms. As a result, predictions from a newer version of a model may not be comparable to predictions from previous versions. Model developers can even discontinue support for older versions making it difficult to obtain additional information on training sets, interpretation criteria, etc.

Information on the prediction output should also be considered including the actual prediction and information on the structural or other features of the test pesticide that influenced the prediction. For dichotomous endpoints, predictions may take the form of a positive/negative or active/inactive result, but often a dichotomous result will be expressed as a numerical probability (*i.e.*, 0 – 1) by QSAR models or as a semi-quantitative probability (*e.g.*, probable, likely, not likely) by SAR/expert systems which can then be interpreted as positive or negative according to various interpretation criteria. Information on the predicted probability, and the interpretation criteria and the rationale for their use may assist an evaluator to determine whether a prediction is relevant at the problem formulation stage and/or can be considered when determining the reliability of a prediction during a more detailed evaluation (see section 5.4). Numerical endpoints (*e.g.*, NOAEL, $LD_{50,}$ BCF) predicted from QSAR models may be taken at face value, but in some cases specific criteria may be recommended by the model developer or by the regulatory agency if the predictions are used to support labeling requirements or hazard classifications (*e.g.*, specific classification/label statement when predicted value is within an order of magnitude of value X). It should also be remembered that some numeric predictions will need to be converted before application to a weight of evidence assessment (*e.g.*, conversion of a predicted LOAEL in units of mmoles/kg bw/day to mg/kg bw/day).

Other information that can be important to consider at the problem formulation stage includes the structural or other features of the test pesticide that have influenced the (Q)SAR prediction including structural fragments, specialized descriptors of structure (*e.g.*, molecular size, shape and electronic parameters), and physical-chemical properties (*e.g.*, molecular weight, Log K_{ow}, boiling point) that are used as variables in QSAR model algorithms, and structural alerts that

are used by SAR/expert systems to identify potentially active compounds. Information on how these features influenced the overall prediction either quantitatively or qualitatively can impact on the level of reliability assigned to a (Q)SAR prediction (see section 5.4). In particular, it can be important to investigate whether the structural fragments, descriptors, and/or physical-chemical properties that drive a prediction are consistent with available information on mechanism of action or not.

When using QSAR models, it can be important to review the identities of the compounds similar to the test pesticide that influenced the prediction. This would likely be obtained from an analysis of the training set compounds that formed the basis for the model algorithm. Similarly, for SAR/expert systems, the compounds that were utilized to support the development of any structural alerts identified in the test pesticide could be reviewed. The compounds that make up a category or group used in a read-across or trend analysis approach can also be considered as compounds that are similar to the test pesticide and that directly influenced the (Q)SAR prediction from that approach. Regardless of the type of tool used, the identities of the compounds that influenced the prediction, how their similarity to the test pesticide was assessed and the degree of similarity, how they influenced the prediction, the nature of the empirical data for them that is related to the predicted endpoint, how well they are predicted by the (Q)SAR tool (*i.e.*, internal validation), and whether a mode and/or mechanism of action has been established are all important considerations when determining the reliability of the (Q)SAR prediction for a test pesticide (see section 5.4).

4.4 Empirical Data Including Information on Mode of Action

Although this section is intended to focus on the preliminary analysis of (Q)SAR predictions as one of the sources of information in a problem formulation for pesticide risk assessment, it must be remembered that the empirical database for a pesticide can impact on the determination of relevance of a (Q)SAR tool for a particular regulatory application (see section 5.3) and the reliability of predictions obtained from that tool (see section 5.4). Empirical data that may influence the use of (Q)SAR predictions includes not only the results of conventional toxicity tests, but also information on mode (and mechanism) of action.

As indicated previously, it is likely that in most pesticide assessment scenarios, (Q)SAR will not be used in a stand alone manner, but will represent only one of multiple lines of evidence considered. Therefore, understanding what relevant empirical data are available, the strengths and limitations of these data, and any gaps that need to be addressed will facilitate the determination of whether there are any (Q)SAR tools relevant for those gaps. Integrating existing empirical data

on a pesticide with relevant and reliable (Q)SAR predictions could also help build defensible rationales for requiring additional empirical studies on specific endpoints, mode of action, etc. (*e.g.*, targeted testing).

The integration of the empirical database with (Q)SAR predictions at the problem formulation stage could also be important for more detailed evaluations of the reliability of the (Q)SAR predictions at a later stage in the assessment (see section 5.4). Questions to consider include whether a predicted endpoint for a pesticide is consistent with and supported by empirical data for related endpoints for the same pesticide or whether the prediction contradicts these data. Empirical data for similar compounds, metabolites and degradation products can be particularly important to consider when assessing the reliability of a (Q)SAR prediction. Knowledge of the toxicity database for a parent pesticide compound could impact on the level of confidence assigned to a (Q)SAR prediction for a metabolite. In some cases, the consistency of the results of (Q)SAR predictions for a parent pesticide versus a metabolite may be useful in determining the confidence in the prediction for the metabolite, especially if the parent compound contains structural alerts known to be associated with specific mechanisms of toxicity and those alerts are preserved or activated following metabolic transformation (*e.g.,* substructures associated with DNA/protein binding). Also, an evaluation of the existing empirical data for a pesticide may provide justification for using more or less conservative criteria to interpret a (Q)SAR prediction for that same pesticide.

As mentioned previously, information on mode of action for toxicity is one type of empirical data that could impact on the consideration of (Q)SAR predictions at the problem formulation stage of a pesticide assessment. A consideration of mode of action can also include the pesticidal mode of action against the target species, and any postulated modes of action of toxicity in non-target species (*e.g.,* humans) which could be used to support the results of existing (Q)SAR predictions or rationales for generating additional (Q)SAR predictions and/or obtaining additional empirical data.

If a pesticidal mode of action is not species specific (*e.g.*, acetylcholinesterase inhibition), information on this mode of action may support the need to investigate related endpoints (*e.g.*, neurotoxicity, developmental neurotoxicity) using (Q)SAR predictions for various taxa, which could, in turn, lead to requirements for additional *in vivo* studies of those endpoints in the relevant organisms. A common pesticidal mode of action may also be a means of identifying groups of similar pesticides from which to build categories and support data bridging through read-across or other types of predictions.

For a postulated (eco)toxicological mode of action, the extent to which the initial chemical-biological system relationship is understood and how well the cascade of key events leading to the adverse outcome is understood (*i.e.*, mode of action, mechanism of action, adverse outcome pathway) in taxa under consideration could directly influence the level of confidence in (Q)SAR predictions for endpoints associated with this mode of action. For instance, when a (eco)toxicological mode of action has already been established for a structurally similar compound, or for a chemical class in which the pesticide in question resides, this mode of action could be used at the problem formulation stage to focus (Q)SAR predictions on particular endpoints and taxa, bridge from the structurally similar compound to inform dose selection for any study required for the pesticide in question, provide support for waiving the need for specific studies based on the current pesticide dataset, and/or help to rule out the relevance of the observed or predicted effect to humans or other species, so that additional studies are unlikely to be required.

Although information on postulated modes of toxicological action can provide support for (Q)SAR predictions at the problem formulation stage and during weight of evidence assessments, it should be noted that, mode of action determinations are generally data rich decisions that must be made on a case-by-case basis. The International Program on Chemical Safety (IPCS) has developed an extensive framework for mode of action analysis based on the Bradford Hill criteria which can be used for cancer and non-cancer endpoints in the context of human health, and for ecological endpoints (Boobis *et al.*, 2008). In most situations, information on toxicological mode of action will not be readily available for a majority of pesticides.

In some cases, comprehensive toxicological mode of action data for pesticide may not be available, but it may be possible to use information on the chemical structure of a pesticide and/or selected (Q)SAR tools (*e.g.*, OECD QSAR Toolbox) to identify potential (chemical) mechanisms of action of pesticides (*e.g.*, mechanisms of protein or DNA binding) to assist in identifying analogs, grouping chemicals into categories and supporting read-across extrapolations (OECD, 2011b; 2009; 2007a).

4.5 Summary

The initial step in framing the questions to be addressed in the human health or environmental assessment of a pesticide is problem formulation. Although the questions to be addressed in pesticide risk assessments have traditionally been framed in terms of the available empirical data, (Q)SAR predictions are another source of information that can be considered during the problem formulation process. The assessment context in which (Q)SAR is being applied, the characteristics of the pesticide that is the subject of the prediction, the characteristics of the (Q)SAR tool and the prediction, and the available empirical data including mode of action data that could impact on the application of (Q)SAR are all important factors to consider when integrating (Q)SAR into a problem formulation. This type of preliminary analysis of the (Q)SAR information on a pesticide could lead to an immediate conclusion that (Q)SAR is not suitable for the particular pesticide assessment question or it could set the stage for a more in depth evaluation of whether a (Q)SAR prediction is fit for purpose for integrating into a weight of evidence decision (see section 5).

5. Evaluating the Adequacy of (Q)SAR Predictions

EVALUATING THE ADEQUACY OF (Q)SAR PREDICTIONS

Topics Discussed in this Section:

- **Scientific validity of a (Q)SAR tool**
- **Applicability of the (Q)SAR tool to the pesticide**
- **Relevance of the (Q)SAR tool to the assessment context**
- **Reliability of the (Q)SAR prediction**
- **Documentation of (Q)SAR tools and predictions**

5.0 Introduction

Evaluating whether a (Q)SAR prediction is adequate or "fit for purpose" is an important component of applying the prediction to a pesticide assessment. The European Commission Joint Research Centre (JRC) has noted that whether a prediction from a (Q)SAR model is adequate or not depends upon four key factors: the scientific validity of the model, the applicability of the model to the query chemical, the reliability of the (Q)SAR result, and the relevance of the (Q)SAR model for the regulatory purpose. The validity of the model was to be established through the application of the OECD QSAR validation principles (OECD, 2004), the applicability of the model relates to whether the chemical of interest lies within the model domain of applicability, reliability is based on the application of a valid (Q)SAR to a chemical within its domain of applicability, and relevance involves considering whether a predicted endpoint can be directly applied to a particular regulatory purpose (EC, 2008b). Similarly, the REACH guidance for applying (Q)SARs provides a flexible framework for using (Q)SAR models in lieu of experimental data that is based on four main conditions: the scientific validity of the model used, the applicability of the model to the chemical of interest, the relevancy of the prediction for the regulatory purpose, and whether appropriate documentation on the (Q)SAR and the prediction is provided (ECHA, 2008; ECHA, 2010; Worth *et al*., 2011).

In this section of the NAFTA (Q)SAR Guidance Document, the key factors noted by the JRC for assessing the adequacy of (Q)SAR models and the REACH

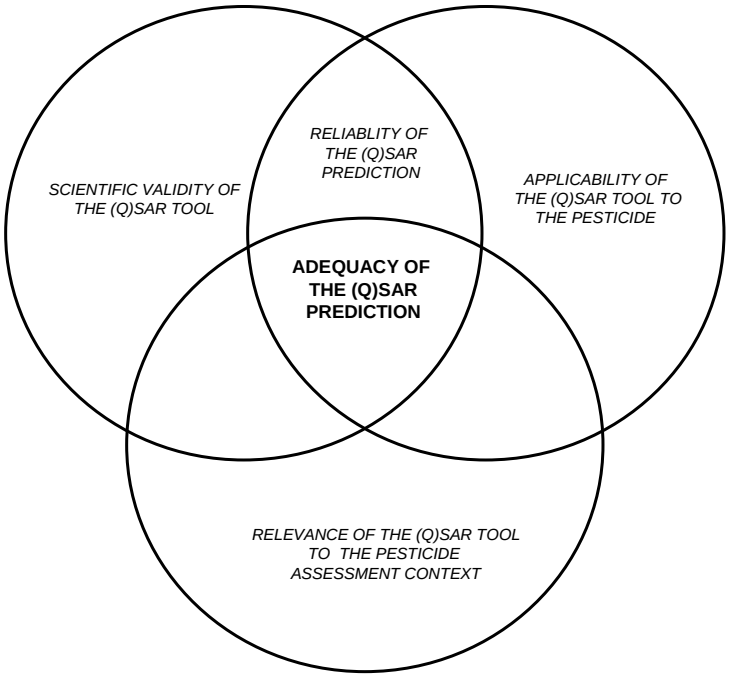
framework have have been adapted to guide pesticide evaluators through the information to be considered when evaluating whether predictions from (Q)SAR tools are adequate for use in pesticide assessments. A schematic for the resulting modified framework is shown in Figure 5–1. Evaluating the adequacy of (Q)SAR predictions relies on a lot of the same information initially considered at the problem formulation stage (see section 4), but with a more focussed consideration of validity, applicability, relevance, and reliability. This type of evaluation can be done in advance of or at least independently of the process of combining the prediction with other information in a weight of evidence assessment (see section 7). Since clear and complete documentation of (Q)SAR tools and predictions is important both to the evaluation of the adequacy of predictions and their incorporation into weight of evidence assessments, this section also includes a discussion of documentation.

The guidance provided here is not meant to be prescriptive, but is intended to allow for case-by-case flexibility and the incorporation of expert scientific judgment. As such, it is recognized that the level of detail and effort employed in evaluating the adequacy of predictions and documenting them will vary depending on a number of factors including the assessment context in which (Q)SAR is being applied.

Although the evaluation of the adequacy of a (Q)SAR prediction may be a new concept to many pesticide evaluators, the process can be thought of as parallel to evaluating the adequacy of empirical studies. When evaluating traditional animal toxicity studies, evaluators can generally rely on the existence of validated test guidelines that are applicable to most pesticides, whereas for (Q)SAR predictions, additional effort needs to be invested to assess the validity of the (Q)SAR tool and its applicability to the pesticide in question.

Much of the discussion in this section focuses on (Q)SAR models. However, it should be recognized that the key issues to consider when evaluating the adequacy of (Q)SAR predictions are applicable to all types of (Q)SAR tools.

Figure 5–1: Evaluating the Adequacy of a (Q)SAR Prediction for a Pesticide (modified from ECHA, 2008 and Worth *et al*., 2011)



5.1 Scientific Validity of the (Q)SAR Tool

The OECD has defined (Q)SAR validation as "the process by which the reliability and relevance of a particular approach, method, process or assessment is established for a defined purpose" (OECD, 2007c). In the context of (Q)SAR model validation, the OECD considers that reliability focuses on the predictive accuracy of the (Q)SAR tool for a range of different chemicals and relevance refers to specific toxicological pathways and mechanisms that culminate in the test endpoint. In particular, it is assumed that a (Q)SAR tool that has a mechanistic basis for the predicted endpoint tends to be more relevant and reliable for groups of chemicals acting via the mechanism in question (OECD, 2007c).

5.1.1 OECD (Q)SAR Validation Principles

The OECD previously noted that one of the critical challenges to the regulatory acceptance of (Q)SAR predictions was the lack of an internationally harmonized framework for assessing (Q)SARs. In particular, there was a need for an internationally-agreed-upon set of principles for (Q)SAR validation to provide a

scientific basis for making decisions on the acceptability of (Q)SAR predictions, and to improve the transparency and consistency of (Q)SAR reporting leading to a greater mutual acceptance of predictions (OECD, 2007c).

In response, the OECD developed the *Principles for the Validation, for Regulatory Purposes, of (Q)SAR Models* which can be used as guidance for the types of information to review when determining if a (Q)SAR model is acceptable or not for use in a regulatory or decision-making framework. The principles include "1) a defined endpoint, 2) an unambiguous algorithm, 3) a defined domain of applicability, 4) appropriate measures of goodness-of-fit, robustness and predictivity, and 5) a mechanistic interpretation, if possible." (OECD, 2004). The OECD also drafted and finalized a separate guidance document (*Guidance Document on the Validation of (Quantitative) Structure-Activity Relationships [(Q)SAR] Models*) that includes a discussion of the principles and information on how to validate (Q)SARs for different applications (OECD, 2007c).

The five OECD (Q)SAR validation principles are presented in sections 5.1.1.1–5.1.1.5 along with a summary of some of the key issues identified in the OECD guidance document and other sources that should be considered in the context of evaluating (Q)SAR tools for application to specific purposes in pesticide risk assessments. For further details on the principles and their application, evaluators should consult the OECD guidance document (OECD, 2007c). Evaluators may also be interested in consulting a recent paper by Dearden *et al*. (2009) which outlined 21 types of errors related to the OECD (Q)SAR validation principles which were identified in various (Q)SAR analyses published in the scientific literature.

Application of the principles is an important step in determining the adequacy of (Q)SAR predictions for use in pesticide assessments. However, the OECD has noted that because of the designs of many of the currently available (Q)SAR models, it may not be possible to completely address all of the principles in every case. Consequently, evaluators will need to be flexible and take into account the available information on (Q)SAR tools and predictions, and individual regulatory program requirements when applying the principles (OECD, 2007c). Also, because of the range of (Q)SAR tools that could be used to make predictions for pesticides and the varying levels of complexity of these tools, use of the OECD (Q)SAR validation principles will require the application of expert scientific judgment, in some cases from a multidisciplinary team.

Example No. 1 in Appendix III provides a summarized version of a case study of reliability and validation testing of a set of (Q)SAR models for predicting acute toxicity to fish species.

5.1.1.1 Principle 1 — Defined Endpoint

The purpose of this principle is to make sure that the endpoint being predicted by a given (Q)SAR tool is transparent. According to the OECD a “defined endpoint” can be considered as “any physicochemical, biological or environmental effect that can be measured and therefore modeled.” (OECD 2007c).

Unlike empirical data derived from standardized guideline based studies designed to meet regulatory requirements for pesticides, studies for chemicals in (Q)SAR model training sets may be based on non-standardized, non-uniform, experimental protocols and conditions. The variability induced by these differences can affect predictive performance and may be a limitation for some (Q)SAR models. However, this variability does not necessarily invalidate the data or models derived from them, but the characteristics of the data and their potential impacts on model predictions must be taken into account.

No (Q)SAR model can be better than the data upon which it is based. Optimally, all of the training set data for a particular (Q)SAR model should correspond to the specific regulatory endpoint of interest, have been generated using the same experimental protocol (ideally a standardized guideline type protocol), and be interpreted using evaluation criteria that correspond to those of the specific pesticide regulatory program. While this type of approach would help to ensure the reliability and relevance of (Q)SAR predictions, (Q)SAR model developers often have to rely on studies conducted under different protocols and conditions in order to ensure sufficient numbers and diversity of chemical structures in the model training sets (OECD, 2007c).

Variability can also be induced by the nature of the regulatory endpoint. Regulatory test guideline type endpoints such as developmental toxicity may actually encompass a range of subendpoints (*e.g.*, teratogenicity, fetal growth retardation, fetal death). Attempting to model poorly defined endpoints may result in the use of model training sets containing a variety of chemical structures producing different subendpoints via different mechanisms of action in a variety of study types. Failing to take this variability into account can result in poor correlations between model parameters and predicted endpoints resulting in poor predictive performance. Alternatively, building a model for a more defined endpoint such as a 96-hour LC_{50} in fish using a more mechanistically homogeneous training set will likely produce better correlations and predictive performance. Finally, in some cases, there may be uncertainties associated with the model endpoint and training set data because information on study protocols and evaluation criteria may not be readily available for some (Q)SAR models, particularly certain commercial models. Pesticide evaluators should take these

potential sources of variability and uncertainty into account when evaluating the validity of a (Q)SAR tool.

5.1.1.2 Principle 2 — Unambiguous Algorithm

The second principle states that a (Q)SAR model should be associated with an unambiguous algorithm. This means that the specifics of the relationship between the chemical structures and the predicted endpoint or property (*e.g.*, an equation) should be clear and transparent. For a mathematically-based QSAR model, the algorithm may take the form of a regression equation that relates descriptors of the chemical structures to the predicted endpoint. Although it is recognized that the unambiguous algorithm principle may be best applied to statistical QSAR models, the OECD has extended the principles to other model types, such as structural alert based SAR/expert system, where the algorithm would take the form of expert-derived rules (OECD, 2007c).

Ideally, a (Q)SAR algorithm should be clear enough that an independent (Q)SAR analyst should be able to explain how predictions were generated and reproduce the results, if required. Although some (Q)SAR models that do not have transparent algorithms may have equal or better predictive performance than more transparent models, the lack of transparency of the former may negatively impact their regulatory acceptance. While transparency is critical, the OECD has stated that there is a difference between having a transparent algorithm and being able to interpret the algorithm as a cause-and-effect relationship. The descriptor values and equation for a QSAR model may be readily available, but a mechanistic/causal link between the descriptors and the predicted endpoint may not have been identified (OECD, 2007c).

In practice, the degree of transparency varies depending on the type of (Q)SAR tool considered, as some non-commercial models have fully transparent algorithms, while most commercial model developers consider specific algorithms and how they were derived to be proprietary information. Also, what constitutes a sufficient level of transparency for regulatory purposes will likely depend upon the assessment context such that only limited information may suffice when (Q)SAR predictions are used to prioritize inventories of chemicals for further testing/assessment, whereas much more detail would likely be required for the algorithm of a (Q)SAR model used to derive a quantitative prediction for a regulatory point of departure estimate. In some cases, it may be possible to compensate for less than complete transparency by analyzing a model’s predictive performance for a set of chemicals similar to the test chemical, but not used in the model training set, and for which empirical data are available (ECHA, 2008).

5.1.1.3 Principle 3 — Defined Domain of Applicability

Netzeva *et al*. (2005) have defined the applicability domain of a (Q)SAR model as "the response and chemical structure space in which the model makes predictions with a given reliability." This means the range of chemical structures, physicochemical properties, mechanisms, and responses over which the (Q)SAR tool can generate reliable predictions for the intended regulatory purpose. The domain of applicability is dependent upon the set of chemicals on which the tool is based (*e.g.*, (Q)SAR model training set).

While it is possible to make predictions for chemicals outside of the applicability domain of a (Q)SAR model, such predictions are extrapolations that are assumed to be less reliable than predictions for chemicals within the domain of applicability (*i.e.*, interpolations). Also, because there are multiple ways of defining domain of applicability (*e.g.*, structures, physico-chemical properties, mechanisms), there may be variations in the reliability of predictions even for chemicals within the domain of applicability of a (Q)SAR tool. For instance, a prediction for a test chemical that is structurally similar to chemicals in the training set of a (Q)SAR model may still be unreliable if the test chemical has a different mechanism of action compared to the chemicals in the training set (OECD, 2007c).

There is a balance between the overall range of the domain of applicability and the predictivity of a (Q)SAR tool. Models with large training sets and diverse domains of applicability may be capable of generating predictions for a wider variety of chemical structures than smaller more structurally and mechanistically homogeneous models, but there is a greater chance that many of those predictions will be unreliable (OECD, 2007c; ECHA, 2008). Using information on mechanisms, mode of action, and/or adverse outcome pathways to group chemicals can improve predictive performance for large heterogeneous training sets.

A number of existing commercial and non-commercial QSAR models have built-in methods for determining whether a compound lies within the domain of applicability. Examples include the univariate analysis (whether the training set substructures include the substructures in the query chemical) and multivariate analysis (whether the query chemical's descriptors are in the optimum prediction space) in the TOPKAT program, the CAESAR models' warnings for descriptor values outside the range of the training set compounds and classes or groups of compounds known to be less than optimally predicted, and the ASTER system's notification when a chemical is outside the predictive capability of a model (Accelrys Inc., 2004; Benfenati, 2010; US EPA, 2011).

The OECD guidance document on the validation of (Q)SAR models summarizes a variety of different methods for defining domain of applicability including the use of structural features that enhance (toxicophores) or modulate toxicity to define the mechanistic domain, characterizing the descriptor or interpolation space by graphing and distance (geometric) analysis, using Williams plots to visualize outliers in descriptor and response space, comparing the structural and physical-chemical similarity of the test chemical to the training set by fragment based approaches, and other methods (OECD, 2007c). A number of reviews of different methods for defining domain of applicability have also been published (Nikolova and Jaworska, 2003; Dimitrov *et al.*, 2005a; Jaworska *et al.*, 2005; Netzeva *et al.*, 2005).

It should be noted that there is no single approach, or set of accepted approaches, to assessing domain applicability. Consequently, whatever approach is adopted should be transparently presented and documented.

In the context of (Q)SAR predictions for pesticide active ingredients, the need to assess the domain of applicability cannot be over-emphasized. A long-standing limitation of many commercial and non-commercial (Q)SAR models has been domains of applicability that are not sufficiently representative of the structures and mechanisms of action associated with pesticide active ingredients. This is in part related to the nature of pesticide data (*i.e.*, confidential unpublished studies accessible only by regulatory agencies). Fortunately, this is changing over time as resources such as the US EPA ToxRef database should it make it possible to build models and other tools with domains of applicability that are more encompassing of pesticide active ingredients.

5.1.1.4 Principle 4 — Appropriate Measures of Goodness-of-fit, Robustness, and Predictivity

According to principle 4, a (Q)SAR should be associated with “appropriate measures of goodness-of-fit, robustness and predictivity” which are obtained through statistical validation of a (Q)SAR tool. For a QSAR model, goodness-of-fit refers to how well the model accounts for the variability in the endpoint or property measured for the training set chemicals. Robustness is a measure of how much change will be induced in the coefficients, etc. in the model algorithm if the training set chemicals are changed. Predictivity involves determining how well the model can make predictions, generally for an external test set of data not included in the training set (Eriksson *et al*., 2003).

Goodness-of-fit for regression-based QSAR models is usually expressed as a multiple correlation coefficient (R^2 value; range: 0 – 1) which is the amount of

variation in the predicted values that can be explained by the regression equation, and the standard error of the estimate (*s*) which measures the dispersion of the predicted values around the regression line. Well-fitted models have R^2 values close to 1 and low *s* values. Poorly-fitted models are not likely to be too useful for regulatory applications. However, it should be noted that deceptively high R^2 and low *s* values can be obtained by including a large number of variables or descriptors in the regression equation (*i.e.*, over-fitting a model). Generally, better predictive performance can be obtained when the ratio of the number of chemicals in the training set to the number of descriptors in the regression equation (*i.e.*, the Topliss ratio) is 5:1 or more. Note that R^2 and *s* values alone are not enough to assess model validity as they do not provide information on the predictive performance for chemicals not included in the training set of a (Q)SAR model (OECD, 2007c; ECHA, 2008).

For (Q)SAR tools that make dichotomous classifications (*i.e.*, active/inactive, positive/negative), goodness-of-fit is usually expressed as Cooper statistics such as sensitivity (fraction of true positive chemicals predicted as positive), specificity (fraction of true negative chemicals predicted as negative), accuracy (fraction of true positive and negative chemicals correctly predicted as positive and negative, respectively), and positive and negative predictivities (probabilities that chemicals predicted as positive and negative are actually positive and negative, respectively). Some (Q)SAR models can be biased towards high specificity or sensitivity depending on the specific application they are designed for. Because the Cooper statistics are interrelated, designing a model for high specificity can result in decreased sensitivity (*i.e.*, high false negative prediction rate) and vice versa. This can be an important consideration as there is generally a greater emphasis on correctly predicting positive chemicals (*i.e.*, high sensitivity) for pesticides and other environmental chemicals. Cooper statistics can also be influenced by the distribution of positive and negative chemicals in the test set such that the predictive performance for the largest class of chemicals in the test set (*i.e.*, positive or negative) will impact on the accuracy of the model and the proportion of positive chemicals in the test set will influence the positive and negative predictivities. The OECD has recommended that the Cooper statistics be significantly greater than 50% for classification models used in a stand-alone manner, but there is no absolute value for differentiating good from poor predictive performance and a (Q)SAR tool with poor performance for one Cooper statistic may still be useful depending upon the application (OECD, 2007c; ECHA, 2008). Also, as discussed in section 6, it may be possible to combine predictions from multiple (Q)SAR tools to enhance overall predictive performance.

Predictivity can be assessed by external validation, either through the use of a test set of chemicals separate from the (Q)SAR model training set or by separating a set of chemicals into a training set and a test set at the design stage (Gramatica, 2007). External validation is usually measured by an external correlation coefficient (Q^2_{ext}). External test sets should be of sufficient size and representative of the types of chemicals to be predicted using the (Q)SAR model.

In some cases, model developers may also present the results of internal validation techniques such as leave-one-out (LOO) and leave-many-out (LMO) methods. For these methods, one or more chemicals is removed from the training set, the model is re-built, the removed chemicals are predicted, the process is repeated, and the average predictivity across the various versions of the model is estimated as a cross-validated regression coefficient (Q^2). One of the reasons internal validation statistics are presented is that there may be limited data from which to construct an independent external test set because (Q)SAR model developers generally want to maximize the number of training set chemicals, leaving few chemicals for external validation testing.

Q^2 or Q^2_{ext} values of >0.5 and >0.9 are considered to represent good and excellent performance, respectively, but it should be noted that predictivity is dependent on the statistical method used and the composition of the test set.

Also, as stated previously, predictions outside the domain of the training set are likely to be less reliable than predictions within the domain of applicability, so that validation principle 4 is closely linked to validation principle 3 (OECD, 2007c; ECHA, 2008).

It should be noted that not all elements of principle 4 are applicable to all (Q)SAR tools, so the assessment of goodness-of-fit, robustness, and predictivity may have to be made on a case by case basis. Rule-based SAR/Expert Systems that use databases of structural alerts are one example in which there is generally no training set and as such LOO, LMO, and other methods will not be applicable. Also, when considering Cooper statistics for external validation testing, the determination of specificity and negative predictivity may be difficult if the expert system is only based on structural alerts for activity.

5.1.1.5 Principle 5 — Defined Mechanism of Action, if Possible

The fifth validation principle states that a (Q)SAR model should be associated with a mechanistic interpretation wherever possible. Although it is recognized that mechanistic information is not always available for (Q)SAR models, whenever it is available, it should be investigated and reported. A transparent mechanistic interpretation can assist in the determination of whether the domain of applicability of a model is suitable for predictions for the chemical of interest, help with the interpretation of outliers, guide hypothesis testing, and provide support for the biological plausibility (*i.e.*, toxicological interpretation) and reliability of the predictions from a model. However, the absence of a clearly identified mechanistic basis for a model does not necessarily mean that the model is not potentially useful for a given regulatory application (OECD, 2007c).

For QSAR models, a mechanistic interpretation represents the physical, chemical and/or biological basis for the model descriptors and their relationship with the endpoint or property to be predicted (ECHA, 2008). A mechanistic interpretation can be associated with a QSAR model through the selection of mechanistically relevant descriptors at the time of model development (*i.e.*, *a priori*) or through the investigation and delineation of the mechanistic basis for the descriptors in an existing model (*i.e.*, *a posteriori*) (OECD, 2007c).

As indicated previously using mechanistic similarity to group chemicals in the training set of a (Q)SAR tool can provide a solid basis for QSAR model development and interpretation. Consequently, when evaluating the mechanistic basis of a (Q)SAR tool, it is important that the rationales for grouping training set chemicals be presented, particularly with respect to any mechanistic hypotheses that were applied (*e.g.*, skin sensitization associated with protein binding potential) and how the mechanistic hypotheses were translated into structural inclusion/exclusion rules (*e.g.*, grouping thiol compounds with potential for protein binding via disulfide formation).

For knowledge-based SAR/expert systems and other related tools, the mechanistic interpretation can be related to observed empirical data, expert knowledge, and expert derived rules on the chemical reactivity and/or biological activity of various chemical substructures (OECD, 2007c).

5.2 Applicability of the (Q)SAR Tool to the Pesticide

Whether a (Q)SAR tool can be considered as applicable to a pesticide depends upon the characteristics of the pesticide (see section 4.2) and the domain of applicability of the (Q)SAR tool (see section 5.1.1.3).

In terms of the characteristics of a pesticide, accurate information on identity, composition, and structure is necessary when determining whether a (Q)SAR tool could be applicable. Many (Q)SAR models are limited to making predictions for discrete organic chemicals and are incompatible with pesticides that are mixtures, salts, or polymers. These incompatibilities may necessitate the use of surrogate compounds such as monomers, uncharged acid forms, and single mixture components to make predictions. Although surrogates may in some cases be a useful approach to making predictions for pesticides that are incompatible with available models, their use should be supported by rationales that account for the potential impacts of molecular size and weight, ionization state, variations in mixture composition, synergism/ antagonism between mixture components, and other factors.

Similarly, if the isomeric form of a pesticide could have an impact on the endpoint or property to be predicted, the (Q)SAR tool will need to be capable of differentiating between isomers to be applicable. A QSAR model that uses 2-D structural descriptors and only accepts 2-D structural representations of chemicals to be predicted will not be very useful for predicting differences in toxicity between stereoisomeric forms of a pesticide. A better approach would be to use a QSAR model capable of recognizing structural representations of isomers, that includes isomer specific descriptors, and whose training set is sufficiently diversified with respect to data on different isomeric forms.

As discussed in section 4.2, for pesticides that can be transformed in the environment or through metabolism in the body, the toxicity, ecotoxicity, physical-chemical properties and other properties of the transformation products, degradates or metabolites may differ from those of the parent pesticide. Whether or not a (Q)SAR tool could be applicable to a pesticide that can be metabolized or transformed would involve identifying metabolites or transformation products (*i.e.*, from empirical data or model predictions) and determining whether predictions can be generated for them or not.

Section 5.1.1.3 outlines the concept of defining domain of applicability during the evaluation of the validity of a (Q)SAR tool. While it is possible for some (Q)SAR tools to make predictions for pesticides outside their domains of applicability, those predictions are likely to be less reliable at best or in some cases the

pesticides will be so far outside the domain of applicability that the (Q)SAR tools should not be considered as applicable. As discussed, there are a number of commercial and non-commercial (Q)SAR models that include automated methods for assessing whether a chemical lies within their domain of applicability based on limits on descriptor values, the presence of unrecognized structural features, and other parameters. Also, a variety of different methods of defining domain of applicability have been published (OECD 2007c; Nikolova and Jaworska, 2003; Dimitrov *et al*., 2005a; Jaworska *et al*., 2005; Netzeva *et al*., 2005).

5.3 Relevance of the (Q)SAR Tool to the Assessment Context

As noted by the JRC, the relevance of a (Q)SAR model involves considering whether a predicted endpoint can be directly applied to a particular regulatory purpose (EC, 2008b). This is based on the endpoint or property that the tool is capable of predicting and the specific type of prediction information that the tool can generate for a particular assessment context. The information obtained at the problem formulation stage on the assessment context that (Q)SAR is being applied to (section 4.1) and the characteristics of the (Q)SAR tool and the prediction (section 4.3) can provide a useful starting point for assessing the relevance of the (Q)SAR tool.

In order for a (Q)SAR tool to be relevant, the endpoint or property that it predicts must correspond to the endpoint or property for which a data requirement exists in a given pesticide assessment context. A (Q)SAR model, capable of generating reliable predictions for the mutagenicity of chemicals in *Salmonella typhimurium* TA1538 may provide useful information on the *in vitro* mutagenicity of a pesticide, but it will not provide specific information to address a data requirement for an *in vivo* clastogenicity study. Similarly, a positive prediction for general pre-natal developmental toxicity for a pesticide may not be sufficient to address a question about whether a pesticide can induce specific skeletal malformations.

Whether the endpoint or property predicted by a (Q)SAR tool could address a specific pesticide data requirement involves a clear understanding of the data from which the tool was derived. This corresponds to the OECD validation principle of defined endpoint which was discussed in section 5.1.1 — *i.e.*, understanding study protocols, data interpretation criteria, and other study elements.

The type of information that a (Q)SAR tool can generate can also impact on its relevance to a pesticide assessment question. In particular, (Q)SAR models are usually designed to generate qualitative or quantitative predictions for particular

endpoints. A model that can provide a qualitative (*e.g.*, yes/no, positive/negative) estimate of the toxicity of a pesticide to freshwater fish may provide some useful information, but will be of limited relevance if a prediction of an acute LC_{50} in trout is required for a particular assessment context.

5.4 Reliability of the (Q)SAR Prediction

In addition to considering the validity of a (Q)SAR tool for a particular pesticide assessment context, the applicability of the tool to the pesticide, and the relevance of the tool to the assessment context, it is also necessary to evaluate the level of reliability (or confidence) in the individual prediction itself. Evaluating the reliability of a prediction takes into account information gleaned from the problem formulation process (see section 4) and information obtained when the (Q)SAR tool is evaluated using the OECD validation principles (see section 5.1).

The pesticide assessment context is an overlying consideration when evaluating the reliability of a (Q)SAR prediction. Moving from a less comprehensive to a more comprehensive assessment context will likely require a higher level of reliability from any (Q)SAR predictions used in the assessment. When rapidly prioritizing chemicals for further assessment, it may be possible to take predictions from validated (Q)SAR tools almost on face value. However, prior to relying on a (Q)SAR prediction as a critical piece of information in a human health or environmental risk assessment for a pesticide, the relationship of the pesticide to the domain of applicability of the (Q)SAR tool, the strengths and limitations of the tool, the prediction results and how they are interpreted, the predictive performance of the tool for similar chemicals, and the potential impact of other available information all have to be evaluated in more detail in order to judge the reliability of the (Q)SAR prediction.

5.4.1 Relationship of the Pesticide to the Domain of Applicability of the (Q)SAR Tool

The importance of considering the domain of applicability of the (Q)SAR tool has already been mentioned with respect to the OECD validation principles (see section 5.1.1.3) and the applicability of a (Q)SAR tool to a pesticide (see section 5.2), respectively. Section 5.1.1.3 also references a number of methods for assessing domain of applicability.

Evaluating the relationship of the pesticide to the domain of applicability of the (Q)SAR tool essentially involves determining whether the pesticide lies within the domain of applicability or outside of it. As indicated previously, predictions for pesticides outside of the domain of applicability of a (Q)SAR tool are not

necessarily inaccurate, but are generally considered less reliable than predictions for compounds falling with the domain of applicability.

As mentioned in section 5.1.1.3, domain of applicability may be defined in different ways (*e.g.*, descriptor, structural fragment, mechanistic, and metabolic domains). Whether a pesticide is within the domain of a descriptor based QSAR model is usually based on comparing the pesticide descriptor values to the range of values for the chemicals in the training set. Structural fragment domain analyses would involve ensuring that the pesticide doesn't contain fragments that are not present in the training set of the model. For the mechanism of action or metabolic domain, the key question is whether the pesticide is likely to act via the same mode/mechanism of action and/or be metabolized in the same manner as other chemicals for which the (Q)SAR tool is applicable (EC, 2010). The OECD has noted that because there are different ways of defining domain of applicability, a prediction for a pesticide that is within the domain of applicability of a (Q)SAR tool based on structural and physicochemical parameters may still not be reliable if it has a unique mechanism of action not covered by the mechanistic domain(s) of applicability of the (Q)SAR tool (OECD, 2007c).

The age of the QSAR model and its training set may also have impacts on the consideration of the domain of applicability of the model and the reliability of the prediction. An older, global type QSAR model may make a negative prediction for a pesticide because its training set is populated with a limited number of chemicals that contain the key structural elements in the pesticide and that all tested negative in historical empirical studies. However, a more up-to-date model, whose training set has been tested in more modern empirical studies, has been segregated into groups according to mechanism of action, and contains a larger number of compounds from the same chemical class as the pesticide of interest, many of which have positive empirical test results, may generate a positive prediction that is more reliable even though the pesticide falls within the domains of applicability of both models. Consequently the use of the most up-to-date versions of models and training sets is recommended and could be particularly important when combining information from multiple predictions (see section 6).

Finally, as discussed above, assessing the domain of applicability may be particularly important for pesticides as (Q)SAR tool developers have not always had access to proprietary pesticide empirical studies for incorporation into training sets.

5.4.2 Strengths and Limitations of the (Q)SAR Tool

The strengths and limitations of a (Q)SAR tool can impact on the evaluation of the reliability of the predictions from that tool (Hulzebos *et al.*, 2001; Greene, 2002). One source of strengths and limitations is the general methodologies on which various (Q)SAR tools are based (*e.g.*, analog approaches, chemical categories, SAR and QSAR models, etc.) (see section 4.3). An example already cited in this document is the lack of structural alerts linked to inactivity or negative test results in some SAR/expert systems. If no structural alerts are identified for a pesticide using this type of system and this is considered as equivalent to a prediction of inactivity (negative), the prediction may be less reliable than a positive prediction from the same system or a negative prediction from another type of (Q)SAR tool that uses descriptors, alerts or other parameters directly related to inactivity, depending on the assessment context. Similarly, the overemphasis on statistical associations and lack of a mechanistic basis for predictions may make some statistical QSAR models less reliable.

Built-in biases are another source of strengths and limitations of (Q)SAR tools that could influence the reliability of predictions. For instance, some QSAR models for pharmaceutical applications have training sets with distributions of positive and negative compounds designed to generate higher specificity versus sensitivity scores (Section 5.1.1.4). This type of bias needs to be taken into account when models of this type are applied to pesticides as they may generate a higher proportion of false negative predictions. The European Chemicals Agency noted a potential source of bias for biodegradation models in their guidance for the implementation of the REACH legislation. Because QSAR models for biodegradation are often biased towards non-ready biodegradability, predictions of biodegradability may be less reliable than predictions of non-ready biodegradability (ECHA, 2008).

The sources of data for training set compounds, and the sources of data or methods of calculation for descriptors (see section 5.1.1.1) can be another type of strength or limitation of (Q)SAR model that could impact on the reliability of predictions. While empirical datasets for registered pesticides usually consist of peer reviewed guideline type studies, many model training sets are based on open literature studies of varying quality. Also, as noted by Doull *et al.* (2007), for some chemical classes, potential training set data may not be available from the published literature. Similarly, the sources of the descriptor values and/or methods used to estimate them may need to be scrutinized when evaluating the reliability of a QSAR model prediction. Whether calculated descriptors, especially obscure types, are reproducible or whether methods used to estimate descriptors

for older versions of QSAR models have been supplanted by newer methods could impact on the acceptability of predictions. These considerations also apply to chemical category/read-across approaches. The methods used to identify similar compounds, and the sources used for the endpoint related, physical-chemical property, mechanistic and other data used to support chemical category development and read-across predictions may need to be carefully considered when determining the reliability of those predictions (OECD, 2007a).

5.4.3 Prediction Results and How They are Interpreted

Along with the basic qualitative (*e.g.*, positive, negative, marginal) or quantitative (*e.g.*, LC_{50}, LOAEL, TD_{50}, etc.) prediction results, additional information is available from most (Q)SAR tools which can be used to assist in evaluating the reliability of predictions. Many QSAR models provide information on the structural fragments, descriptors or physical-chemical parameters used as variables in their algorithms. Examination of the values of these variables and their coefficients in the model algorithm can indicate whether they positively or negatively influenced a (Q)SAR prediction and the magnitude of their impact. Combining information on the influence of structural fragments or descriptors on a model prediction with knowledge of their relationship to the mechanism of action for the predicted endpoint can provide powerful evidence to support or question the reliability of the prediction. For instance, an increased level of reliability could be assigned to a QSAR model prediction for a toxicity endpoint directly related to a receptor binding process if the model algorithm contains descriptors of molecular size and shape known to be related to receptor binding affinity and the values of those descriptors for the pesticide in question are similar to those for chemicals known to bind to the receptor and produce the effect in question.

Other information generated by some (Q)SAR models includes calculated values for molecular weight of the test chemical, and properties such as Log K_{ow} and bioavailability (*e.g.*, based on Lipinski's rule of 5) which can help with the consideration of whether a prediction for an endpoint could reliably represent what might occur following an *in vivo* exposure to a chemical. Similarly, some SAR/expert systems have expert rules that can take into account physical-chemical factors which can impact on absorption/bioavailability by discounting the presence of a structural alert associated with toxicity in a pesticide if the physical-chemical parameters of the pesticide are outside the range normally associated with the endpoint in question.

Understanding how the results of (Q)SAR predictions have been interpreted is another consideration in the evaluation of their reliability. Algorithms for QSAR models are generally derived from training sets of empirical study results for

chemicals. How those empirical study results are interpreted can influence the nature of the algorithm, the predictive performance of the model, and ultimately the reliability of predictions from that model. For a carcinogenicity (Q)SAR model developed from a training set of rodent bioassays, the bioassay results may have been interpreted as positive based on a specific percentage increase in tumor incidence over controls, a statistically significant increase in incidence over controls, a statistically significant trend over several dose groups, and/or other criteria. The reliability of a prediction from such a model could be influenced by whether the study interpretation criteria were consistent among the training set chemicals and whether the criteria correspond to regulatory agency specific interpretation criteria.

The criteria for interpreting predictions that have been developed by the originator of the (Q)SAR tool and the rationale for them should also be taken into account when evaluating the reliability of predictions. Statistical-based QSAR models often generate probabilities (*i.e.*, 0 – 1.0) for dichotomous (*e.g.,* positive/negative) endpoints and the model developers recommend specific criteria for interpreting the predicted probabilities (*e.g.*, TOPKAT criteria: ≥0.7 and ≤1.0 = positive; ≥0.0 and <0.3 = negative; ≥0.3 and <0.7 = inconclusive; Accelrys Inc., 2004). Criteria of this nature are usually developed based on internal and/or external validation testing to optimize the predictive performance of the model.

Although originators of (Q)SAR tools may recommend prediction interpretation criteria, users of the tools may make modifications to those criteria. A pesticide applicant or regulatory evaluator may decide that interpretation criteria put forth by a (Q)SAR model developer are too conservative or not conservative enough based on previous experience with the model, the results of validation studies or other information. In some cases, comparisons of predictions for parent compounds versus metabolites, data from related empirical studies, kinetic and/or mechanism of action data, and other information may be used to modify, override or contradict the interpretation criteria recommended by a model developer. Also, regulatory agencies may develop standing policies on how selected (Q)SAR tools should be interpreted that may differ from those of the tool developers. Regardless of which criteria are used, they should be transparent so that they can be considered in the evaluation of the reliability of the predictions.

5.4.4 Predictive Performance of the (Q)SAR Tool for Similar Chemicals

Testing the predictive performance of a (Q)SAR tool on chemicals that are similar to the pesticide in question and have empirical data available for them can provide another source of information for evaluating the reliability of predictions. Chemicals from the same chemical class as the pesticide in question, as well as isomers, salts, and other forms could be considered for testing the predictive performance of the (Q)SAR tool. For example, a starting point for testing the predictive performance of a (Q)SAR tool for a sodium salt of an organic acid would be to generate a prediction for a de-salted acid form of the compound for which empirical data are available. Which chemicals to use would depend on the type and quality of empirical data available for them, the parameter used to assess similarity (*e.g.*, physical-chemical parameters, structure, metabolism) and the degree of similarity.

As discussed in other sections of this document, one of the advantages of applying (Q)SAR to pesticide metabolites, transformation products or manufacturing impurities can be the abundance of high quality guideline study derived data on a structurally related parent pesticide. In some cases, parent pesticides with existing empirical databases can be used as external validation sets or “positive and negative controls” for the associated metabolites, transformation products or manufacturing impurities. When an endpoint for a parent pesticide is well predicted by a (Q)SAR tool and the structures or descriptors with the greatest influence on predictions are similar for the parent pesticide and its metabolite, it may be possible to assign a greater reliability to the prediction for the metabolite from the same tool.

5.4.5 Other Available Information

Although there may be some instances where (Q)SAR predictions will be used in a stand-alone manner, in most pesticide assessments, predictions will be only one of many lines of evidence to be considered. Therefore, when evaluating a (Q)SAR prediction for a pesticide, it is important to consider the impact of other available information on the evaluation of the reliability of the (Q)SAR prediction. Available empirical data on a pesticide could contradict the (Q)SAR prediction or support it. Also, even if the prediction is supported by the empirical data, the degree of support may justify a more or less conservative interpretation of the prediction. For example, a prediction of high acute inhalation toxicity for a pesticide may be generated by a QSAR model, but contradictory empirical data on the physical-chemical properties of the pesticide may indicate low volatility or low potential for aerosolization which may lead the evaluator to question the

reliability of the prediction and seek additional data or predictions for structurally similar chemicals. On the other hand, there may be cases where precursor effects in a target organ consistently reported in short-term studies may be used to support a (Q)SAR prediction of a related longer-term effect (*e.g.*, carcinogenicity) in the same target organ and species by the same route of exposure.

Just as a defined mechanism of action can be an important consideration when evaluating the validity of a (Q)SAR tool (section 5.1.1.5), empirical data on mode or mechanism of action can be an important consideration when evaluating the reliability of an individual (Q)SAR prediction. Even if a (Q)SAR tool is not based on a defined mode or mechanism of action, high quality, empirical toxicological mode or mechanism of action data for a pesticide can represent a very powerful line of supporting or contradictory evidence for a prediction from that tool, including supporting or contradicting the relevance to humans and/or environmental organisms of the predicted toxicity endpoint. For instance, because of the relationship between the protein binding capability of chemicals and their skin sensitization potential, it could be assumed that a category composed of chemicals that are not only structurally similar to a test pesticide, but are also known to bind to proteins via the same mechanism as the test pesticide will likely result in a read-across skin sensitization prediction that is more reliable compared to predictions where information on protein binding mechanism is not available (Dimitrov *et al.*, 2005b). One example of the impact of mode of action data on the human relevance of a (Q)SAR prediction could involve the interpretation of a positive prediction for renal tumors in male rats that has been statistically validated and has a domain of applicability that encompasses the pesticide in question. The interpretation of such a prediction may have to be tempered if data are available from short-term or specialized mode of action studies that indicate the accumulation of alpha-2u-globulin in the kidneys of male rats administered the pesticide, a potential mode of action of questionable relevance to humans.

As outlined in section 4.4, empirical data on the pesticidal mode of action of pesticides could provide support and enhance the reliability of (Q)SAR predictions for related endpoints in humans and other non-target species. Also, a common pesticidal mode of action could be one of several lines of evidence supporting the formation of a category of related chemicals. A category that is based not only on structural and physical-chemical similarities, but also similarity of pesticidal mode of action amongst category members is likely to result in more reliable read-across and trend analysis predictions.

5.5 Documentation of (Q)SAR Tool and Prediction

In order for an evaluator to critically review the adequacy of a (Q)SAR prediction for a pesticide, the (Q)SAR tool and the prediction must be documented with a sufficient level of transparency. This is similar to the concept of sufficient documentation for empirical studies as delineated in empirical study guidelines and in guidance for producing robust study summaries for regulatory purposes. What constitutes a sufficient level of transparency will depend on the assessment context, specific data reporting requirements or policies of the regulatory agency and the type of (Q)SAR tool. When predictions of toxicity, ecotoxicity, environmental fate, etc. are used in a prioritization or screening context, it may not be necessary to provide full details on the adequacy of the (Q)SAR predictions. However, for a (Q)SAR prediction to be accepted as a critical data point in a pesticide assessment would likely require less uncertainty, and thus more extensive documentation analogous to pesticide data evaluation records (DERs) used to capture critical information from conventional toxicity, exposure, and other study types.

5.5.1 General Types of Information

At the present time, a standardized template for reporting information on (Q)SAR predictions included in pesticide assessments has not yet been developed. However, in lieu of such a template, some recommendations can be made regarding the general types of information to report (see Table 5–1). In general, for any (Q)SAR prediction, sufficient information must be provided to clearly identify the chemical for which the prediction is being made and the model or other (Q)SAR tool used to generate the prediction. A description of the results of the prediction and how they were interpreted should also be presented as well as a discussion of the validity of the tool in the context in which it is being used. These recommendations should only be considered as a starting point for what to include when documenting (Q)SAR predictions for pesticides. As indicated above, the level of detail to be included when reporting on predictions will depend on the assessment context, specific data reporting requirements or policies of the regulatory agency and the type of (Q)SAR tool.

While the recommendations in Table 5–1 are fairly general, they include a number of information elements that may be more suitable for (Q)SAR models than other tools such as analog and category approaches. Specific guidance on reporting formats for analog and category approaches has been developed by the OECD (OECD, 2007a).

Table 5–1: Recommended General Types of Information to Include when Documenting (Q)SAR Predictions

Information on the chemical
- Chemical (systematic) and common names
- CAS number
- Structural formula
- Form of the chemical (including relevant stereochemistry)
- Structural entry format

Information on the (Q)SAR model or tool
- Type of model or tool
- Name of the software platform
- Name of model/submodel or tool
- Version number/date of model/submodel or tool
- Characterization of the training set

Validity of the (Q)SAR model or tool
- Information on the predicted endpoint
- Information on the algorithm
- Domain of applicability
- Internal/external validation statistics
- Mechanistic information (if available)

Results and interpretation of the prediction
- Qualitative prediction (*i.e.*, yes/no, +/-)
- Quantitative prediction (*e.g.*, LOAEL = 50 mg/kg bw)
- Predictive probability
- Criteria for interpreting predictions

5.5.2 Documentation of (Q)SAR Predictions — EC QMRF and QPRF

The European Commission (EC) has developed the QSAR Model Reporting Format (QMRF) and the QSAR Prediction Reporting Format (QPRF) as detailed documentation templates for providing sufficient information to facilitate regulatory consideration of (Q)SAR models and predictions. The QMRF has been designed to provide information related to the OECD principles for the validation, for regulatory purposes, of (Q)SAR models and the QPRF has been developed to provide information to assist in the consideration of the adequacy of a (Q)SAR prediction for a defined regulatory purpose (EC, 2008a; 2008b). In terms of parallels with the assessment of empirical studies, the QMRF is somewhat analogous to an empirical study test guideline and the QPRF is analogous to a study DER. However, the QMRF and QPRF are very detailed and were not specifically designed for use in a pesticide context. They are discussed here only as useful examples to consider when specific (Q)SAR reporting templates for pesticides are being developed. One of the potential projects under consideration as a follow-up to this document is the development of pesticide specific (Q)SAR DERs. Further background information on the QMRF and QPRF, tables

summarizing the information fields in each, and the EC website link for the templates and additional guidance on their use are included in Appendix II.

5.6 Summary

Evaluating the adequacy of a (Q)SAR prediction is an essential step for determining whether the prediction is useful source of data for a pesticide assessment. Adequate or fit for purpose predictions can be incorporated into weight of evidence assessments, whereas inadequate predictions necessitate a reliance on other sources of data alone. Whether a prediction is adequate or fit for purpose should always be determined within a specific assessment context such that a prediction that is adequate to support a request that specific empirical studies be conducted may not be adequate enough to replace those studies in an assessment. In this section, a framework has been presented for evaluating the adequacy of predictions based on a consideration of the validity of the (Q)SAR tool, the applicability of the (Q)SAR tool to the pesticide of interest, the relevance of the (Q)SAR tool to the pesticide assessment context, and the reliability of the prediction (Figure 5–1). This framework relies on information obtained during problem formulation for (Q)SAR (see section 4) and it is flexible enough to be useful for a variety of different assessment contexts. The next section of this document (section 6) deals with combining information from multiple predictions. While there are specific issues associated with combining multiple predictions, in general, the framework for evaluating the adequacy of (Q)SAR predictions outlined in section 5 can also be applied to multiple predictions. For either single or multiple predictions that have been determined to be adequate, the next step is the incorporation into a weight of evidence assessment which is the subject of section 7 of this document.

6. Combining Information from Multiple Predictions

COMBINING INFORMATION FROM MULTIPLE PREDICTIONS

Topics Discussed in this Section:

- **Approaches to combining multiple predictions**
- **Advantages and disadvantages of combining predictions**
- **Selecting (Q)SAR tools for multiple predictions**
- **Evaluation of multiple predictions**

6.0 Introduction

Combining information from multiple (Q)SAR predictions can be thought of as analogous to combining the results of multiple *in vivo* and *in vitro* studies to strengthen a weight of evidence argument for a toxicity or ecotoxicity endpoint. Because different (Q)SAR tools may have different prediction paradigms and different strengths and limitations, combining predictions has the potential to increase the confidence in the overall prediction. However, it should be noted that combining predictions from multiple (Q)SAR tools does not eliminate the need to ensure that each prediction is adequate or fit for purpose (see section 5). In particular, it is important that each (Q)SAR tool used be scientifically valid, applicable to the pesticide of interest, and relevant to the assessment context.

Also, while there are advantages to combining predictions, it is not always necessary to do so. In some instances, relying on a single prediction from a (Q)SAR tool that is valid for the stated purpose, applicable to the pesticide in question, and relevant to the assessment context may be much more acceptable than trying to combine predictions from tools with significant limitations.

This section briefly discusses approaches to combining information from multiple predictions, some advantages and disadvantages of combining predictions, selecting (Q)SAR tools for multiple predictions, and the evaluation of multiple predictions. Although much of the discussion focuses on combining predictions from (Q)SAR models, many of the concepts mentioned can also be applied when predictions from different types of (Q)SAR tools are combined.

6.1 Approaches to Combining Multiple Predictions

Combining the output of multiple (Q)SAR models into an overall prediction has been referred to as consensus modeling, battery approaches, or weight of evidence approaches to (Q)SAR modeling (Abshear *et al*., 2006; OECD, 2007; Hewitt *et al*., 2007; Matthews *et al*., 2008, 2009a; Ellison *et al*., 2010; Hewitt *et al*., 2010). An example of a fairly simple approach to consensus predictions is the set of interpretation criteria in the KnowITAll computational system described by Abshear *et al*. (2006) which are summarized in Table 6–1 below. These criteria range from single hit/unanimity requirements for true/false predictions based on the worst case scenario and *vice versa* for the best case scenario, to the assessment of counts of true and false predictions for the majority and percentage agreement scenarios. For toxicity predictions, a value of true can be considered as equivalent to a positive prediction and value of false as equivalent to a negative prediction.

Table 6-1: Consensus Modeling Interpretation Criteria (Abshear *et al*., 2006)

Scenario	Definition
Worst Case	1. If any model returns a value of true, return a value of true 2. Only if all models return a value of false, return a value of false
Best Case	1. If any model returns a value of false, return a value of false 2. Only if all models return a value of true, return a value of true
Majority Rules	1. If the majority of the models return true, the consensus will be true 2. If the majority of the models return false, the consensus will be false
Percentage Agreement	1. If a specified percentage of the models returns a true value, the consensus will be true 2. Otherwise, the consensus will be false

In contrast, a slightly more complex method of combining predictions is the weight of evidence approach of Ellison *et al*. (2010) in which predictions from the OECD QSAR Toolbox, Derek for Windows, CAESAR and SMARTS rules were combined for skin sensitization with positive predictions given a weighting of +0.5, negative predictions given a weighting of -0.5, and the absence of structural alerts in a compound given a weighting of -0.25. The sum of the weightings was then interpreted in a weight of evidence argument as positive: ≥ +0.5, negative: ≤ -0.5, and inconclusive: > -0.5 and < +0.5.

Additional factors that could be considered when weighting individual predictions in a consensus approach include characteristics of the models or tools that are being combined such as the presence/absence of a mechanistic basis for the predictions, the characteristics of the model training set (*e.g.,* data sources and

domain of applicability), and the known predictive performance of the models or tools. Finally, when predictions are made for a quantitative (*i.e.,* continuous) endpoint (*e.g.,* LC_{50}, LOAEL, TD_{50}, EC_{50}, etc.), it may, in some cases, be possible to average the predicted numerical values or combine them using other statistical methods. Consensus predictions can vary in complexity depending on the tools considered and the methods of counting, scoring or weighting the individual predictions.

6.2 Advantages and Disadvantages of Combining Predictions

6.2.1 Advantages of Combining Predictions

A number of researchers have demonstrated improvement in measures of statistical fit and predictive performance when multiple (Q)SAR predictions for human health related and environmental toxicity endpoints are combined. One example is the study of Matthews *et al*. (2009a) involving the generation of consensus predictions for drug-induced adverse liver and urinary tract effects using training sets configured for four (Q)SAR programs (*i.e.,* MC4PC, MDL-QSAR, BioEpisteme and Predictive Data Miner). Consensus predictions where a positive prediction from at least one program was considered as an overall positive result resulted in an increase in sensitivity from 39% for one program to 56% for two, and to 68% for four programs. Increased sensitivity was at the cost of specificity which decreased from 86% for one program, to 78% for two, and to 67% for three. Consensus predictions requiring agreement between two, three and four programs increased the specificity by 4, 9, and 12%, respectively, compared to single program predictions. In this case, sensitivity was increased for consensus predictions from two or more programs, but reduced by 11 and 26% for predictions requiring agreement between three and four or more programs, respectively. Similar improvements in predictive performance were obtained for carcinogenicity predictions using models built from the four (Q)SAR programs above in an earlier study (Matthews *et al*., 2008). Matthews *et al*. (2009a) concluded that no one (Q)SAR model can provide both high specificity and high sensitivity. Combining models that have good specificity individually in a consensus approach can enhance the overall sensitivity, which can be an important consideration for models used in the assessment of pesticides.

Another example of improved predictive performance from consensus modeling, is the work of Lewis *et al*. (2002) in which a predictive concordance of 100% was obtained from combined COMPACT and Hazard Expert predictions for a small group (14) of carcinogens. The concordances were 71% for COMPACT alone

and 57% for Hazard Expert alone.[4] However, not all studies have demonstrated improved model statistics for consensus versus individual models. Hewitt *et al*. (2007) used genetic algorithms to construct a range of models for four different data sets (silastic membrane flux, toxicity of phenols to the ciliated protozoan *Tetrahymena pyriformis*, acute toxicity in fathead minnow and flash point). There was no consistent improvement in statistical fit or predictivity (*i.e.,* R^2, Q^2 root mean square error) for average predictions from a consensus of the 10 best models (*i.e.,* models with the highest R^2 and Q^2 values); or a consensus of a diverse set of models that best covered the available model space compared to the single regression model with the best R^2 and Q^2 values.

The potential for multiple (Q)SAR models to provide complementary or confirmatory information compared to individual predictions is another advantage of consensus approaches. Matthews *et al*. (2009a) defined complementary models as two or more models that predict different sets of active and inactive chemicals when used on the same test set and noted that combining complementary models can enhance predictive performance (*e.g.,* sensitivity). Combining complementary models could also enhance or expand the overall coverage or domain of applicability. For instance, if within a set of 100 pesticides, 40 contain key structures that are in the domain of applicability of (Q)SAR model A, whereas the domain of applicability of (Q)SAR model B covers half of those same 40 pesticides plus key structures found in the 60 pesticides not covered by model A, then combining predictions could provide complementary coverage of all 100 pesticides. Combining predictions from multiple models for similar or related endpoints could also provide complementary information that increases the reliability of the overall assessment. A single positive prediction for mutagenicity in a (Q)SAR model for one specific strain of *Salmonella typhimurium* provides limited information on the potential microbial (prokaryotic) mutagenicity of a compound. Combining multiple positive predictions from models for several *Salmonella* strains, models for *in vitro* mutagenicity in mammalian cell systems, and models for *in vitro* and *in vivo* chromosomal aberration assays could provide complementary information that when combined with evidence of carcinogenicity from *in vivo* bioassays may provide an initial indication of a potential genotoxic mode of action for a carcinogenic compound.

Models that are based on different predictive paradigms (*e.g.,* molecular fragment versus molecular descriptor paradigms) that predict the same chemicals to be active and inactive from a single test set provide confirmatory

[4] Although the COMPACT model is not currently available, the example was included to illustrate some of the advantages of combining predictions from different (Q)SAR models.

information (Matthews *et al*., 2009a). This type of confirmatory information can increase the overall confidence in predictions (Contrera *et al*., 2007), especially when additional mechanistic insights are provided by one or more of the models. Individual models can emphasize a set of structural features in a molecule while placing reduced or no emphasis on other features (Gramatica *et al*., 2007). Consequently, combining multiple (Q)SAR models that are based on different methodologies can help to relate the activity of a compound to different aspects of its structure, confirming the impact of key structural features on activity or providing additional insights into the key parts of a compound's structure that influence activity (Contrera *et al*., 2007). Confirmatory predictions also increase the likelihood that the structures in the active compounds are causally related to the activity in question and that the compounds come from clusters with the same mechanism of action (Matthews *et al*., 2008; 2009a). However, combining predictions from models based on the same methodology (*e.g.,* several statistical (Q)SAR models based on similar descriptors) and developed from the same training set would not be expected to provide much additional information.

6.2.2 Disadvantages of Combining Predictions

Combining information from multiple predictions has the potential to greatly increase the complexity of the predictive process in terms of selecting models to be combined, approaches to combining models, and interpreting the combined predictions. In the study of Hewitt *et al*. (2007) described above, the authors noted the complexity of generating multiple models from a range of descriptors for each endpoint, and in assessing which models and how many should be used from the global model space to construct consensus models. Also, when a large number of descriptors is considered, consensus models may be based on descriptors that are difficult to interpret mechanistically.

Conflicting predictions from individual models for the same endpoint may represent another source of complexity in consensus approaches. While the evaluation of the adequacy of the individual (Q)SAR predictions (see section 5) may help to resolve some differences, conflicting predictions may still occur if different (Q)SAR paradigms are employed and/or the models are based on different training data sets. While simplified approaches for reconciling conflicting predictions may be adopted, such as the criteria summarized by Abshear *et al*. (2006), quantitative or semi-quantitative weighting of models may be necessary to account for differences in model domains of applicability, scoring criteria for training set data, performance measures, mechanistic bases and other characteristics. When multiple factors are influencing the weighting, multiple predictions can be very complex to interpret, especially when there is a need to

additionally weight the (Q)SAR predictions against other data available on a chemical.

Resources needed to develop optimized consensus approaches could be another disadvantage of combining multiple predictions. While it may be possible to predefine weightings of (Q)SAR tools based on their known characteristics, finalized weightings of tools may require multiple rounds of testing and analysis (EC, 2010).

6.3 Selecting (Q)SAR Tools for Multiple Predictions

Many of the considerations in selecting models for multiple predictions encompass similar issues to those discussed in the problem formulation section (section 4) of this guidance document — *i.e.,* the assessment context, characteristics of the pesticide, characteristics of the (Q)SAR tool, and available empirical data including information on mode of action.

The assessment context, including whether the prediction is used to support a data waiver or to identify data requirements, what type of endpoint or property is being predicted, and whether the predicted endpoint or property is critical to a pesticide assessment will influence the level of confidence or reliability required in the combined predictions which, in turn, can be factored into the selection of appropriate (Q)SAR tools. Similarly, the amount and quality of the available supporting empirical data including data on mode of action can also impact on the required level of confidence in the combined predictions and the choice of tools used to make those predictions.

As discussed in section 4, certain (Q)SAR models may not be compatible with the molecular structure of some pesticides (*e.g.,* ionic compounds, complex mixtures, polymers, etc.) and others, though compatible, may have limitations such as insufficient structural diversity in their domains of applicability or the inability to generate quantitative predictions. As a result, the characteristics of the pesticide and of the (Q)SAR tool can also influence the selection of appropriate (Q)SAR tools for the generation of multiple predictions for a pesticide.

While selection of models for specific scenarios will most likely be made on a case by case basis taking into account the factors described above and expert judgment, one example of an approach that has been used previously is the selection of a molecular fragment-based method (Q)SAR model and a descriptor-based model, both having high predictive performance for a toxicological endpoint of interest. Such a combination could have a good chance of improving the combined domain of applicability of the models, and positive predictions

could provide strong evidence that the fragments/descriptors associated with the toxicological activity are highly significant and well separated from the structural features of inactive molecules.

Another example of an approach to selecting (Q)SAR tools for multiple predictions would be combining complementary models with different prediction paradigms, such as molecular fragment-based (Q)SAR methods that detect small molecular fragment alerts, large molecular fragments, and very large molecular fragments. Relative confidence in the predictions could be increased if these different methods identify a common region of a molecule that is correlated with activity. Similarly, approaches based on combinations of predictions from molecular descriptor-based methods that use different pools of descriptors and/or different statistical methods to detect activity could also be considered. Confidence in the predictions is increased when these different methods identify a common physicochemical structural feature of a molecule that is correlated with activity. The work of Matthews *et al.* (2008, 2009a) provides some examples of combining predictions from (Q)SAR models with different prediction paradigms.

6.4 Evaluation of Multiple Predictions

Evaluation of multiple predictions involves similar considerations to those discussed in section 5 of this guidance document relating to evaluating the adequacy of single predictions — *i.e.,* scientific validity of the (Q)SAR tools, applicability of the (Q)SAR tools to the pesticide, reliability of the combined predictions, relevance of the (Q)SAR tools to the assessment context, and documentation of the tools and predictions. Also, as mentioned previously, combining predictions from multiple (Q)SAR tools does not eliminate the need to ensure that each prediction is adequate or fit for purpose.

In addition to the concepts discussed in section 5, another issue to consider when evaluating the results of multiple or consensus predictions is the overall objective of combining the predictions. Have predictions been combined to confirm or to increase the confidence in the predictions from one (Q)SAR model based on the results of other models that use different predictive paradigms? As mentioned above, identical predictions from multiple (Q)SAR models developed from the same training set using very similar predictive methodologies may not provide much additional information, whereas models based on different paradigms that identify the same chemicals to be active and inactive may increase the confidence in the predictions (Contrera *et al*., 2007). Alternatively, is there an interest in improving the overall predictive performance for a given endpoint or parameter by combining complementary models that individually predict different chemicals to be active and inactive? Such an approach could

potentially enhance predictive performance compared to the individual models (Matthews *et al*., 2009a). Knowing the objective of combining predictions and the characteristics of the (Q)SAR tools used can help the pesticide evaluator determine whether it is appropriate to combine predictions or not.

The adequacy of the training sets in the models is another important factor in determining whether it is appropriate to combine predictions (Matthews, 2009b). If confirmatory predictions are required from multiple models then the models should have comparable training sets in terms of coverage of the chemicals to be predicted. Also, the scoring systems or criteria used to characterize the data on the training set chemicals should be comparable. For example, two models based on similar training sets of Ames test results, but constructed using different scoring systems for what constitutes a positive versus a negative assay result may not give reliable predictions when combined. Whether the training sets have been designed to be balanced or heavily weighted towards active or inactive chemicals should also be considered. If the training set for one model has a relatively low ratio of active to inactive chemicals (A/I ratio) and a high sensitivity prediction is desired, then it may be better to combine the predictions from this model with predictions from models that have higher A/I ratios in order to enhance the chance of correctly predicting positive chemicals.

If it is considered appropriate to combine predictions, then how the predictions are combined and interpreted (see section 6.1) is another important question. For example, a pesticide applicant may put a higher priority or heavier weight on models with a high level of specificity when addressing submission data requirements in an effort to support waiver rationales for multiple *in vivo* studies. Conversely, for a metabolite or residue of potential concern for which little empirical data are available, pesticide evaluators may assign a greater weight to models with high sensitivity in order to ensure potential endpoints of concern are flagged for additional data requirements. Interpretation criteria for multiple predictions may also be designed to compensate for the strengths and limitations of the individual models being combined such as giving precedence or higher weighting to predictions from tools with more extensive domains of applicability encompassing the test compound, superior overall predictive performance, clearer mechanistic bases, greater transparency, etc. Regardless of the approach used to combine and interpret multiple predictions, the interpretation criteria should be transparent and the rationale for their use should be included when documenting predictions. Documentation for combining predictions may consist of separate records for each prediction (*e.g.,* QPRFs) with an accompanying rationale for combining predictions, or, alternatively a single

document may be used to capture the individual predictions and the rationale for combining predicitions.

Finally, if it is determined that combining (Q)SAR predictions is appropriate for a particular assessment context, sufficient information on training sets, scoring systems, interpretation criteria, etc., should be available on each model used in the combined prediction so that their strengths and limitations are transparent and if necessary, their predictions can be weighted before they are combined.

6.5 Summary

In general, the aims of combining multiple predictions include enhancing predictive performance, expanding domain of applicability, obtaining complementary or confirmatory information, and ultimately increasing confidence in (Q)SAR predictions. Although there can be advantages to combining multiple predictions, it is not always necessary, as a single prediction from a validated, applicable and relevant (Q)SAR tool is likely to be more acceptable than combined predictions from tools with significant limitations. The selection of appropriate tools for generating multiple predictions for a pesticide will involve a trade-off of the desired advantages against the potential disadvantages and a consideration of the assessment context, characteristics of the pesticide, characteristics of the (Q)SAR tools, and the available empirical data. Combining information from multiple predictions does not represent a new data stream for consideration in pesticide assessments. Rather it is a variation in the (Q)SAR data stream that needs to undergo the same types of problem formulation, evaluation of adequacy, and weight of evidence considerations as single (Q)SAR predictions prior to being incorporated into pesticide assessments.

7. Integration of (Q)SAR Predictions into Hazard Assessments

INTEGRATION OF (Q)SAR PREDICTIONS INTO HAZARD ASSESSMENTS

Topics Discussed in this Section:

- **Incorporating (Q)SAR in hazard characterizations: Overview**
- **Problem formulation and Adequacy Determination**
- **Evaluating empirical data versus (Q)SAR predictions**
- **Mode of action considerations**
- **Overall weight of evidence**
- **Hazard characterization and risk communication**

7.0 Introduction

Traditional pesticide risk assessments in regulatory agencies have routinely been based on the results of laboratory animal testing and estimates of exposure according to the following four key steps from the National Academy of Sciences (NAS) risk assessment paradigm (NRC, 1983):

- Hazard Identification
- Dose Response Assessment
- Exposure Assessment
- Risk Characterization

Twenty four years later, the NAS presented a vision for toxicity testing and risk assessment in the document *Toxicity Testing in the 21st Century A Vision and a Strategy*; this document recommended the use of predictive tools such as (Q)SAR (NRC, 2007). In the outline of this vision, the NAS described some of the risk assessment-related applications of (Q)SAR including the prediction of toxicity, ADME properties, environmental fate, and ecologic effects for chemicals. Consistent with the NAS vision and the existing risk assessment paradigm, the emphasis in this section is on the integration of (Q)SAR tools into the hazard identification (biological endpoint) component of the risk assessment process for pesticides. This section builds upon concepts discussed in section 4 (Problem Formulation for (Q)SAR) and section 5 (Evaluating the Adequacy of (Q)SAR

Predictions), and discusses the process of integrating the overall toxicity database (including (Q)SAR predictions) to arrive at conclusions regarding hazard with consideration of confidence and level of uncertainty. As with traditional pesticide risk assessments, the characterization of the hazards and associated uncertainties are communicated to risk managers for consideration in regulatory decision making.

7.1 Incorporating (Q)SAR in Hazard Characterizations: Overview

In general, the hazard identification and characterization process should not be greatly different for situations where (Q)SAR predictions are an additional source of data compared to traditional hazard assessments in which empirical data alone are considered. As discussed in section 4, there are two situations where (Q)SAR predictions are likely to be considered by a pesticide regulatory authority: a pesticide applicant submits a (Q)SAR prediction to address a data requirement, either to fulfill or support the requirement or a waiver for the requirement or the evaluator of a pesticide risk assessment uses a (Q)SAR prediction to support the case for an additional or refined data requirement. For the first situation where a pesticide applicant has submitted a (Q)SAR prediction, the primary tasks for the evaluator are to determine the purpose of the (Q)SAR submission, determine whether the (Q)SAR prediction fulfills the intended purpose and to factor the (Q)SAR prediction into the overall weight of evidence for the pesticide chemical in the appropriate context. If the (Q)SAR prediction is not clearly deficient, the prediction is evaluated according to the principles and procedures discussed in sections 4 and 5 to determine whether the prediction accomplishes its intended purpose as defined in the problem formulation. At the same time, the empirical data are evaluated with respect to the validity and acceptability of each study. If mode of action data have been submitted, these data are also evaluated for individual validity and acceptability.

For the second situation where an evaluator uses a (Q)SAR prediction, usually there is no empirical information involved and it is still important to assess the validity and reliability of the chosen (Q)SAR method or model for the evaluator’s purpose.

The basic steps in the integration of (Q)SAR predictions and empirical data into a weight of evidence analysis are listed in Figure 7–1 and described in more detail in subsequent sections.

Figure 7–1: Weight of Evidence Analysis: Integration of (Q)SAR Predictions and Empirical Data

1. **Problem formulation**: What is the goal of the assessment? And the role of (Q)SAR in that assessment?
2. **Determination of Adequacy:** What is the adequacy of the (Q)SAR prediction? Is it fit for the purpose intended in the problem formulation step?
3. **Weight of Scientific Evidence:** Integration of the existing empirical data and (Q)SAR predictions. Weighing the scientific data including mode of action information, if available.*
4. **Data base sufficient:** Is the data base sufficient for risk assessment? What data are missing? What is the level of data base uncertainty?
5. **Hazard characterization**: Telling a clear and transparent hazard story and presenting the determination of confidence level and level of uncertainty, and risk communication.

* For most pesticides, a complete understanding of the mode of toxicological action may be absent. To the extent that a toxicological mode of action is postulated for an analog of the pesticide of interest, it would be important to consider this information to help build confidence in a predicted endpoint.

7.2 Problem Formulation and Adequacy Determination

As discussed in section 4, problem formulation in the context of (Q)SAR prediction for pesticides involves asking and answering a number of key questions during the course of the review with respect to assessment context, characteristics of the chemical subject to the (Q)SAR prediction, characteristics of the selected (Q)SAR tool(s) and the prediction, and identification of empirical data including mode of action data that are relevant to the (Q)SAR prediction.

The assessment context is essentially a determination of the objective of the (Q)SAR analysis, the specific endpoint that is predicted, the role that the (Q)SAR prediction plays in the risk assessment and the acceptable level of reliability that the (Q)SAR prediction must have if the prediction is to be accepted for its proposed role.

Pesticide characterization involves identifying whether the prediction is for a pesticide active ingredient, an impurity, a metabolite or transformation product, a pesticide inert/formulant ingredient, or an analog of the pesticide of interest. It

also involves ensuring that the correct structure was the subject of the prediction and whether it is appropriate to use (Q)SAR predictions for that structure.

The characterization of the (Q)SAR tool and the prediction includes a consideration of the general methodology behind the tool, the empirical data on which the tool is based, the endpoint predicted by the tool, other details on the tool, and details on the prediction. As discussed in section 5, this information is used in the evaluation of the adequacy of the prediction that includes the scientific validity of the (Q)SAR tool (*i.e.*, OECD (Q)SAR validation principles), applicability of the tool to the pesticide, reliability of the prediction, and the relevance of the tool to the assessment question (context). The characterization of the (Q)SAR tool and the evaluation of the adequacy of the prediction can aid in the interpretation of the prediction relative to any available empirical data.

Empirical data related to the (Q)SAR prediction and available mode of action data are important to consider if they either support the (Q)SAR prediction or contradict it. These data will be discussed further in the following sections.

7.3 Evaluating Empirical Data versus (Q)SAR Predictions

For many pesticide active ingredients the empirical database will consist of a suite of guideline toxicity studies that are typically required by regulatory agencies for the registration and reregistration of pesticide chemicals. These studies address systemic and local effects (*e.g.*, acute toxicity, skin irritation, developmental toxicity, genotoxicity, carcinogenicity), multiple routes of exposure (*e.g.*, oral, inhalation, dermal), multiple durations of exposure (*e.g.*, acute, short-term, long-term), and are conducted across multiple taxa (*e.g.*, freshwater fish, aquatic invertebrates, rats, mice, rabbits).

Toxicity studies tend to be very detailed. The description of the conduct of the study (*i.e.,* materials and methods) is typically very extensive. The results sections of the studies are also reported in even greater detail and many of the in vivo studies cover a variety of biological endpoints. The studies also have a conclusion section in which the data submitter will often propose a point of departure for the study depending on how the submitter interprets the outcome of the study. The reviewer or evaluator of each study prepares a written evaluation of the study (“Data Evaluation Record” or “DER” at EPA and PMRA), and in this DER the evaluator summarizes the key points of the study. Perhaps most importantly, the evaluator records his or her own conclusions about the adequacy of the study and the appropriate points of departure for the endpoints supported by the study. In the normal review of empirical data for pesticide risk assessment, each study is evaluated individually for scientific rigor and those studies that are

considered acceptable are integrated to “tell a story” or “paint a picture” of the hazard profile of the pesticide chemical.

The review of a (Q)SAR prediction is similar to the review of empirical data. In the typical scenario, the pesticide applicant will submit a documented (Q)SAR prediction including the purpose of the prediction, the rationale for selection of a model (or models), information on how the query structure was entered into the model, a discussion of the OECD validation principles as applied to the model, why the training set chemicals are applicable to the query structure, why the (Q)SAR prediction satisfies a data requirement (or supports a waiver from the data requirement) and a discussion of limitations and uncertainties associated with the prediction. The evaluator reviews the submitted (Q)SAR prediction and all of the supporting documentation and prepares a written record of the review much like the DER for empirical data (see section 5.5).

Once the review of the empirical data and the (Q)SAR prediction have taken place, and provided that the review indicates that the prediction is scientifically valid, the evaluator is in a position to determine whether the prediction is reliable, *i.e.,* whether it fills a data gap or supports a waiver from a data requirement.

Evaluating empirical data with (Q)SAR predictions enables a determination of whether relevant empirical data on structurally related chemicals will support, detract from, or influence the weighting of the (Q)SAR prediction in the assessment. For example, if there are only limited empirical data available from short-term animal studies on a pesticide impurity of potential concern and the objective is to determine whether the impurity has carcinogenic potential and, if so, to obtain some information on how it might elicit carcinogenicity, then a weight of evidence approach could be considered. Such an approach could combine information from the available short-term animal dataset on the impurity, genotoxicity/carcinogenicity studies for structurally similar compounds if applicable, and carcinogenicity and genotoxicity predictions for the impurity in question. In this example, the short-term animal study data may provide information on precursor effects that may support the carcinogenicity of the impurity. Similarly, if the studies on the structurally similar compounds are in agreement with the (Q)SAR predictions for the impurity, this would enhance the confidence in the predictions.

On occasion the (Q)SAR prediction may be in apparent conflict with empirical data. In such situations, it is useful to re-examine both the empirical database and the data supporting the prediction for a possible resolution of the apparent conflict. A theoretical example of using a closer examination of the empirical database to resolve a conflict with a (Q)SAR prediction would involve a pesticide

that demonstrates clear systemic toxicity in short-term dermal toxicity tests. If read-across extrapolations from analog chemicals indicate low potential for dermal absorption, then there is an apparent conflict. However, if a closer examination of the physical-chemical properties of the pesticide and skin irritation testing data reveals that the pesticide is likely to be poorly absorbed at low concentrations, but at high concentrations, it is corrosive, destroying the barrier properties of the skin, facilitating access to the systemic circulation then it may be possible to resolve the differences between the (Q)SAR predictions and the empirical data.

Depending on the basis for a conflict between a (Q)SAR prediction and the results of empirical studies, it may be necessary to fully examine the adequacy of the (Q)SAR prediction (see section 5), and also the adequacy of the empirical data. (Q)SAR tools are reductionist methods that may not fully account for the impact of physical-chemical properties and pharmacokinetics/dynamics, may over-emphasize the contribution of a particular structural alert or property or may miss a toxicologically relevant alert because of database/training set limitations. In general, (Q)SAR predictions should not be used to override the results of well-conducted, guideline type studies for the same endpoint. However, in cases where the empirical studies are of questionable reliabilty because they are non-guideline studies, conducted according to older protocols, restricted to examining specific research endpoints or have other limitations, it may be necessary to give greater weight to reliable and relevant (Q)SAR predictions and/or develop recommendations for further testing to help resolve the conflict. Weighting of predictions and empirical data is discussed further in section 7.5.

As discussed in section 6 (Q)SAR evidence may actually consist of multiple predictions from multiple (Q)SAR tools (*e.g.*, read-across from a chemical category, QSAR model prediction, SAR prediction, etc.) each based on different prediction paradigms with different strengths and limitations. There are a number of important advantages of combining predictions from multiple (Q)SAR models (*e.g.*, improved predictive performance, enhanced domain of applicability, complementary information, increased confidence in the predictions). Because of these advantages, it may be possible to assign a greater level of confidence and a greater weighting to combined or consensus predictions when comparing them to the available empirical data and the other streams of evidence.

There are three likely outcomes to the weight of evidence evaluation: 1) The (Q)SAR prediction adequately addresses the data requirement, *i.e.,* the data requirement is satisfied or the study need not be done; 2) The (Q)SAR prediction is not relevant to the data requirement, *i.e.*, the (Q)SAR prediction is scientifically

adequate but does not address the specific data requirement; and 3) The (Q)SAR prediction is scientifically acceptable and addresses the data requirement, but the data requirement is critical, requiring a high degree of certainty and confidence that the prediction as submitted is unable to meet. In this last instance, the result of reviewing the empirical data and the (Q)SAR prediction together may point the way to follow up action such as additional information from the submitter on the identity of, and relevant data for the training set chemicals; or perhaps targeted mode of action testing on the query structure and specific training set chemicals that could help fulfill the data requirement.

7.4 Mode of Action Consideration

If the weight of evidence includes a known or suspected mode of toxicological action (MOA), this understanding could substantially strengthen the overall database and provide additional support to determine whether the (Q)SAR prediction(s) is biologically plausible and consistent with what is known about the chemical of interest. The mode of action for the chemical of interest provides the overall biological basis for the phenotypically expressed adverse effects reported among the traditional *in vitro* and *in vivo* toxicology studies, and can also support a (Q)SAR prediction for the specific endpoint of interest (*e.g.*, cancer, genotoxicity). Some of the most robust (Q)SARs are ones with the greatest confidence that all the chemicals being combined in a model are producing toxicity through a single molecular initiating event (mechanism).

7.5 Overall Weight of Evidence

Evaluating empirical data versus (Q)SAR predictions is part of the overall concept of integrating multiple lines of evidence into a weight of evidence. While (Q)SAR predictions have strengths, limitations, and uncertainties that can differ from those associated with empirical studies, (Q)SAR predictions can be considered as one potential line of evidence within a weight of evidence assessment of a pesticide.

A detailed discussion of weight of evidence approaches in pesticide risk assessment is outside the scope of this document, but some general principles can be highlighted that are applicable in scenarios where (Q)SAR predictions and empirical data are being integrated. Similar to weight of evidence for empirical data alone, it is likely that qualitative, scientific expert judgement based approaches will be the most frequently used ones for integrating (Q)SAR predictions and empirical data. These approaches can involve considering information from each individual prediction, collectively examining multiple predictions within a (Q)SAR line of evidence (see section 6), considering multiple

data points within each individual empirical study, combining information from similar empirical studies within one line of evidence, and finally, integrating (Q)SAR, empirical and other lines of evidence together to arrive at an assessment conclusion (Health Canada, 2011).

Just as with empirical data based weight of evidence, approaches that integrate (Q)SAR and empirical lines of evidence usually include a qualitative weighting or ranking of the importance of the different lines of evidence for the overall assessment conclusion. Such a weighting involves a consideration of the adequacy (*i.e.*, see section 5) and uncertainties associated with the different lines of evidence. Regardless of the weighting/ranking approach adopted, transparency is critical and can be addressed via comprehensive narrative rationales outlining the approaches followed in considering each line of evidence and integrating the lines of evidence together. It is particularly important to outline the approaches taken when there are conflicts between the (Q)SAR and the empirical data lines of evidence. For ease of interpretation, tabular presentations of the (Q)SAR and empirical data lines of evidence can also be considered.

While qualitative approaches are likely to be used most often, it is also possible to consider quantitative scoring systems or mathematical algorithms that may be more systematic for weighting (Q)SAR and empirical data lines of evidence than qualitative, expert judgement-based approaches. Examples of such systems usually involve numerical weightings for each line of evidence, multiplying the scores for each line of evidence by its weighting, and summing up the weighted scores into an overall result. Just as transparency is critical for qualitative weighting/ranking approaches, it is especially critical to clearly outline the rationale behind any quantitative weight of evidence scoring systems.

7.6 Hazard Characterization and Risk Communication

The hazard characterization should include a clear, straightforward hazard narrative, piecing together all the components of the problem formulation, hazard identification, weight of evidence and addressing the level of uncertainty in the database and confidence in the overall assessment. As with all characterizations of hazard, it is imperative to consider all available hazard data whether these are *in vivo* or *in vitro* toxicity data or, in the absence of measured values, relevant data from (Q)SAR predictions or other predictive methods.

Subsequent to developing a clear, straightforward hazard narrative, communicating this narrative to risk assessors and risk managers becomes critical. The successful communication of risk is not only dependent on relaying the adverse health outcome(s), susceptible subpopulations, dose response and

exposure assessment, but also the quality of the data, level of uncertainty, and confidence in the overall assessment. Any risk mitigation decision that is based on risk assessment conclusions, must be made with a clear understanding of the level of uncertainty surrounding the risk assessment conclusions and what level of confidence should be placed on those conclusions to support a regulatory decision. If, for example, the level of uncertainty in the database is high because most of the non-cancer endpoints are predicted and the level of confidence in the overall risk assessment is weak, the risk manager should be cognizant of this low level of confidence before selecting an adequate risk mitigation option. In short, the regulatory option selected should be consistent with the level of uncertainty identified for the predicted and empirical datasets so as not to over or under-inflate the confidence in these datasets. If that level of database uncertainty can be addressed by additional research, the decision on when the data will be required to be submitted and when they can be considered in future risk assessment and management decisions may be dependent on the potential health outcome. In any scenario, it is critical that the risk manager has all the relevant information from the risk assessor in order to develop appropriate risk management options and make a good regulatory decision based on sound science.

7.7 Summary

The integration of (Q)SAR predictions into the risk assessment involves many steps which are similar to the risk assessment paradigm: hazard identification, dose response assessment, exposure assessment and risk characterization. The only difference between a risk assessment based on traditional empirical data from that which involves in silico predictions is the judgement of adequacy of the (Q)SAR predictions and determination of database completeness.

The steps involved with integrating (Q)SAR predictions rely on starting with a solid problem formulation to establish what the (Q)SAR prediction is intended to inform for the assessment and what type of assessment to be performed will indicate the amount of uncertainty that is deemed acceptable. For example, a screening level assessment would allow for more uncertainty than a risk assessment. The determination that the (Q)SAR prediction is valid and reliable for the purpose described in the problem formulation step is critical to proceed forward in the subsequent steps. Without the determination of scientific adequacy, the (Q)SAR prediction would be rendered unacceptable and therefore, could not be considered in the risk assessment.

The next step of integrating the (Q)SAR prediction with extant scientific data on that compound or a structural analog is critical to a scientific weight of evidence

analysis. Issues of reproducibility of observations, consistency of effects across species, strain, time of exposure and routes of exposure, as well as the determination of biological plausibility and incorporation of mode of action information is considered in this analysis. After developing the weight of scientific evidence, the risk assessor proceeds forward to determine the completeness of the database to support the risk assessment.

If the database is considered deficient, and missing critical studies, the weight of evidence, including any mode of action information, should be informative in determining the type of study(s) needed to fulfill the database deficiency; this may be in vitro and/or short term studies depending on the data deficiency and as the concept of the adverse outcome pathway becomes elucidated for the particular toxicity endpoint, the determination of what study will be needed will be more clearly defined. In short, the combination of the (Q)SAR prediction, empirical data, mode of action, and/or adverse outcome pathway in a weight of evidence approach will inform the risk assessment on database deficiency and identification of critical research needed to address this level of uncertainty.

8. Conclusions and Future Vision for (QSAR) and Pesticides

CONCLUSIONS AND FUTURE VISION FOR (Q)SAR AND PESTICIDES

Topics Discussed in this Section:

- **Toxicity Testing in 21st Century: Shift in the Risk Assessment Paradigm**
- **Weight of Evidence Approach: Biological Plausibility**
- **Adverse Outcome Pathway: Conceptual Framework**
- **Expert Scientific Judgment and Peer Review**

8.0 Toxicity Testing in the 21st Century: Shift in the Risk Assessment Paradigm

The NAS report *Toxicity Testing in the 21st Century: A Vision and a Strategy* emphasizes the need for moving away from prescriptive assessments based on checklists of traditional animal toxicity studies towards an integrated approach that relies on the existing knowledge-base for a class of chemicals and the results of alternative testing methods to identify toxicity pathways and to focus data requirements on more targeted toxicity testing (NRC, 2007). This is especially relevant for pesticide regulatory authorities that receive and evaluate large volumes of animal and ecological toxicity data submitted in support of new chemical registrations. To continue to meet the demands of new pesticide registrations requires new technologies that allow for faster, more efficient and effective technical reviews; this change, however, must not come at the price of public health and environmental protection.

While there is the need for more efficient review processes, there is also the impetus of research on newer technologies, a recognition of the accelerated pace of scientific innovation. New technologies will allow pesticide regulatory authorities to build upon existing knowledge of pesticide toxicity to develop integrated approaches for testing and assessment (IATA) of pesticides. These parallel regulatory and risk assessment changes and advancements in the state-of-the-science propel the agencies forward and expedite the transition towards global application of newer, swifter risk assessment and testing methodologies.

Computational tools vary widely depending on the purpose for their use in risk assessment. (Q)SAR tools represent an example of an alternative testing method that could be a useful component of integrated approaches to testing and assessment. (Q)SAR tools have had a long history of use by industry and regulatory agencies for hazard determinations and other applications; there are also many different types of commercial and non-commercial (Q)SAR tools that are either currently available or rapidly under development. However, in spite of the regulatory experience and the on-going developments in the field of (Q)SAR, there are very few examples of formal (Q)SAR guidance documents that discuss the unique issues and considerations associated with the application of (Q)SAR to pesticide regulatory risk assessments.

Because pesticide regulatory authorities are transitioning down the path of utilizing this type of predictive technology, it becomes increasingly important that a systematic and transparent approach to the use of (Q)SAR predictions in pesticide assessments be adopted and communicated to ensure that the application of sound scientific judgment. To that end, this document seeks to provide risk assessors and hazard evaluators with some general guidance on how to review (Q)SAR predictions included in pesticide submissions. It is well recognized, however, that there are a variety of different guidance documents for (Q)SAR which have been published by agencies such as the OECD, the EU, the US EPA and others. This particular NAFTA guidance document does not endeavor to reproduce or to replace any of the other guidance documents, but seeks to provide much needed ready-to-use, streamlined, plain language guidance to pesticide evaluators on the application of (Q)SAR to pesticide regulatory decision making.

8.1 Weight of Evidence Approach: Biological Plausibilty

This guidance document is consistent with the current hazard/risk paradigm in terms of the approach to the evaluation of (Q)SAR predictions. The document starts at the problem formulation stage as explained in section 4, through evaluating the adequacy of (Q)SAR predictions as described in section 5, combining information from multiple predictions in section 6 and integrating (Q)SAR predictions into hazard assessments assessment in section 7. The overall emphasis throughout this document is that (Q)SAR predictions should be considered among the many other data streams in a weight of evidence approach for determination of hazard/risk. Similar to the consideration of multiple lines of evidence when identifying a toxicological mode of action, (Q)SAR predictions should not be used in isolation when reaching human and environmental hazard assessment conclusions for a pesticide. In addition to

considerations of the reliability, validity and relevance of the individual (Q)SAR tools and predictions, the defensibility of the predictions is dependent on biological consistency and plausibility across all scientific lines of evidence in a holistic weight of evidence approach. This weight of evidence analysis should be described in the hazard characterization section of a pesticide assessment and should include the level of confidence, range of uncertainty, data gaps and any needed research for further refinement of the risk assessment.

8.2 Adverse Outcome Pathway: Conceptual Framework

As stated earlier, (Q)SAR is not a new predictive tool; it has been used for many decades within various regulatory programs to provide predictions of apical endpoints, biological and physiochemical properties. The limitations of using (Q)SAR when the domain of applicability is not inclusive of the chemical under review or when the database is not targeted for a particular endpoint of concern, are well understood. These limitations and others are discussed in this document, particularly in the context of the OECD validation principles, as described in section 5. In spite of these limitations, (Q)SAR remains a predictive tool worthy of continued use and development.

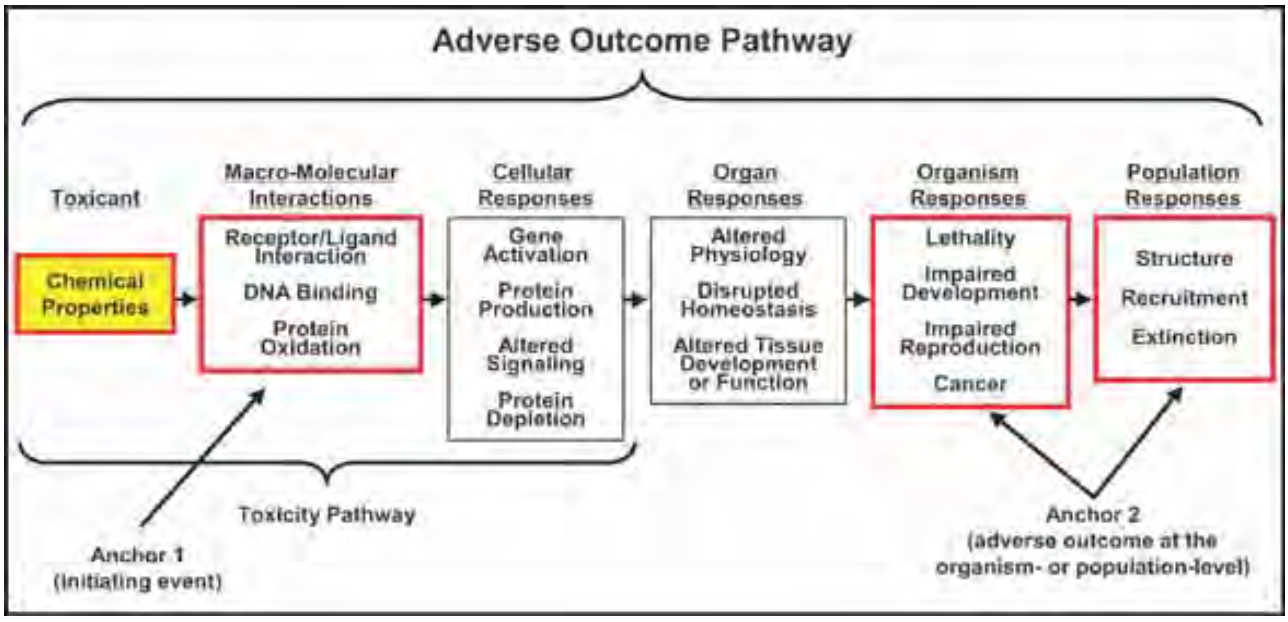
The historical use of (Q)SAR predictions were two-dimensional, stopping at the binary prediction (*e.g.*, yes/no carcinogenicity). The future use of (Q)SAR will involve anchoring (Q)SAR predictions with what is known about that chemical class/category, the biological mode of action, toxicity pathways and population effects; (Q)SAR predictions will be built into larger conceptual frameworks called adverse outcome pathways (AOPs) (Ankley *et al*., 2010). More simplistically, and as illustrated in Figure 8–1 below, AOPs delineate the documented, biologically plausible, measurable and testable processes by which chemicals induce molecular perturbations and subsequent biological responses relevant for risk assessment. The basic concept describes how molecular perturbations cause effects at different levels of biological organizations including at the subcellular, cellular, tissue, organ, and population levels (OECD 2011a). As indicated in Figure 8–1, toxicity pathways, the initial perturbations of cell-signaling motifs, genetic circuits, and cellular-response networks that might eventually result in disease, are components of AOPs, but unlike AOPs, toxicity pathways are not necessarily directly linked to apical effects (*i.e.*, disease outcomes). The mode of action (MOA) is also a component of the AOP; it is inclusive of the events from the initial molecular perturbations to an adverse effect at the individual level, but does not usually consider exposure or effects at the population level (OECD, 2012a). AOPs capture the continuum of metabolism, molecular perturbations, cellular interactions, effects on the tissue and organ leading to individual effects

and effects at the population level in a holistic approach. They allow regulatory authorities to move away from an overdependence on single chemical (Schultz and Diderich, 2011) *in vivo* animal testing and make greater use of computational, molecular and in vitro tools as described and advocated for in the 2007 NAS report *Toxicity Testing in the 21st Century: A Vision and a* Strategy.

One way in which (Q)SARs could contribute to the AOP approach would be through the identification of structural alerts associated with key events in an AOP, particularly molecular initiating events (MIEs). The OECD has noted that a close linkage between an MIE and an observed adverse outcome in vivo can be used as a basis for developing a chemical category for the relationship between chemical structure and the in vivo endpoint. Thus, rather than just relying on intrinsic chemical activity, AOPs potentially provide a comprehensive mechanistic basis for forming toxicologically meaningful categories for making predictions using read-across or (Q)SAR models (OECD, 2012b). As noted previously the European Commission Joint Research Centre is developing a reporting format for describing key events/intermediate effects in AOPs in collaboration with the OECD and ECHA (OECD, 2012a).

An example of an AOP discussed in a US EPA hosted workshop in December, 2010 involves the binding of a xenobiotic to an hepatic nuclear receptor as the molecular initiating event for a variety of toxicity pathways, including pathways leading to liver cancer (Hester *et al*., 2006). The identification of molecular initiating events of this type allows the development of methods to screen for chemical interactions with biological targets and in this case, receptors. This is where (Q)SAR models become critical, in identifying chemical categories based on chemical structures and their linkage to biological activities by understanding the toxicity pathways relevant for risk assessment (OECD, 2011b).

Figure 8–1: Adverse Outcome Pathway Diagram (Ankley *et al*., 2010)



In summary, (Q)SAR tools are important predictive technologies for today's risk assessment as well as those for tomorrow — provided that they are applied with appropriate constraints and cautionary guidance and that there is a meaningful attempt to build data bridges in a weight of evidence approach among (Q)SAR and future emerging predictive technologies to better target efforts to more efficiently and effectively maximize overall biological predictive capability (Benigni *et al*., 2007b).

8.3 Expert Scientific Judgment and Peer Review

While this guidance document is intended to cover the main issues that evaluators should consider when reviewing (Q)SAR predictions included in pesticide submissions, the document is not intended to provide stand-alone, step-by-step instructions for all potential applications of (Q)SAR tools to pesticides. This guidance document should be supplemented with expert scientific judgment and expert peer review to ensure consistency, reproducibility, and scientific defensibility in the use of (Q)SAR in pesticide hazard assessments.

To this end, the NAFTA pesticide regulatory authorities are currently assembling a (Q)SAR expert committee to provide advice to pesticide evaluators in complex assessments that seek to integrate (Q)SAR predictions with empirical data in a weight of evidence approach for hazard/risk determinations that may trigger regulatory risk management decisions. One of the mandates of this expert

committee will be to identify updates, modifications or additions to the guidance document.

Finally, as mentioned previously, there are a variety of other available guidance documents on regulatory applications of (Q)SAR (see Appendix I). Pesticide evaluators are encouraged to consult these other documents for additional information on specific topics or scenarios as required.

References

Note that all website links in the reference list were accessed at the time this document was finalized. Any any problems accessing these links after the finalization of this document should be referred to the originating organizations for the websites.

Abshear, T., G.M. Banik, M.L. D'Souza, K. Nedwed, and C. Peng. 2006. A model validation and consensus building environment. SAR and QSAR Environ. Res. 17(3):311–321.

Accelrys Inc. 2004. TOPKAT, Version 6.2, User Guide, Accelrys Inc., San Diego, CA, USA.

American Chemical Society. 2010. CAS Registry^{SM} and CAS Registry System. CAS® A Division of the American Chemical Society.

Ankley, G.T., G.P. Daston, S.J. Degitz, N.D. Denslow, R.A. Hoke, S.W. Kennedy, A.L. Miracle, E.J. Perkins, J. Snape, D.E. Tillitt, C.R. Tyler, and D. Versteeg. 2006. Toxicogenomics in regulatory ecotoxicology. Envron. Sci. Technol. 40(13):4055–4065.

Ankley, G.T., R.S. Bennett, R.J. Erickson, D.J. Hoff, M.W. Hornung, R.D. Johnson, D.R. Mount, J.W. Nichols, C.L. Russom, P.K. Schmieder, J.A. Serrrano, J.E. Tietge, and D.L. Villeneuve. 2010. Adverse outcome pathway: A conceptual framework to support ecotoxicology research and risk assessment. Environ. Toxicol. Chem. 29(3):730–741.

Arnot, J.A., F.A.P.C. Gobas. 2004. A Food Web Bioaccumulation Model for Organic Chemicals in Aquatic Ecosystems. Environ. Toxicol. Chem. 23(10):2343–2355.

Aronson, D., R. Boethling, P. Howard, W. Stiteler. 2006. Estimating biodegradation half-lives for use in chemical screening. Chemosphere 63:1953–1960.

Arvidson, K.B., R. Chanderbhan, K. Muldoon-Jacobs, J. Mayer, and A. Ogungbesan. 2010. Regulatory use of computational toxicology tools and databases at the United States Food and Drug Administration's Office of Food Additive Safety. Expert Opinion on Drug Metabolism & Toxicology 6, 793–796.

Asikainen, A.H., J. Ruuskanen, and K.A. Tuppurainen. 2004. Performance of (consensus) kNN QSAR for predicting estrogenic activity in a large diverse set of organic compounds. SAR QSAR Environ. Res. 15(1):19–32.

ATSDR. 1989. Toxicological Profile for Bis(Chloromethyl)Ether. Agency for Toxic Substances and Disease Research. U.S. Public Health Service in collaboration with: U.S. Environmental Protection Agency. December 1989. http://www.atsdr.cdc.gov/ToxProfiles/tp128.pdf

Bailey, A., N. Chanderbhan, N. Collazo-Braier, M. Cheeseman, and M. Twaroski. 2005. The use of structure-activity relationship analysis in the food contact notification program. Reg. Toxicol. and Pharm. 42:225–235.

Bassan, A., E. Fioravanzo, M. Pavan, and M. Stocchero. 2011. Applicability of physicochemical data, QSARs and read-across in Threshold of Toxicological Concern assessment. Final report of a study carried out by Soluzioni Informatiche (S-In, Vicenza, Italy) for the European Food Safety Authority (EFSA). Accessible from: http://www.efsa.europa.eu/fr/supporting/pub/159e.htm

Benfenati, E. (editor). 2007. Quantitative Structure-Activity Relationships (QSAR) for Pesticide Regulatory Purposes. Elsevier. Amsterdam.

Benfenati, E. 2010. The CAESAR project for in silico models for the REACH legislation. Chem. Cent. J. 4(Suppl 1):I1.

Benfenati, E., R. Benigni, D.M. Demarini, C. Helma, D. Kirkland, T.M. Martin, P. Mazzatorta, G. Ouedraogo-Arras, A.M. Richard, B. Schilter, W.G. Schoonen, R.D. Snyder, and C. Yang. 2009. Predictive Models for Carcinogenicity and Mutagenicity: Frameworks,State-of-the-Art, and Perspectives. J. Environ. Sci. Health. C, 27(2):57–90.

Benigni, R., and C. Bossa. 2008. Predictivity and reliability of QSAR models: The case of mutagens and carcinogens. Toxicol. Mechan. Methods 18:137–147.

Benigni, R., C. Bossa, T. Netzeva, and A. Worth. 2007a. Collection and evaluation of (Q)SAR models for mutagenicity and carcinogenicity. European Commission, Joint Research Centre, EUR 22772 EN, 126 p. Ispra, Italy. http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/doc/EUR_22772_EN.pdf

Benigni, R., T.I. Netzeva, E. Benfenati, C. Bossa, R. Franke, C. Helma, E. Hulzebos, C. Marchant, A. Richard, Y.-T. Woo, and C. Yang. 2007b. The expanding role of predictive toxicology: An update on the (Q)SAR models of mutagens and carcinogens. J. Environ. Sci. Health C 25:53–97.

Boobis, A.R., J.E. Doe, B. Heinrich-Hirsch, M.E. Meek, S. Munn, M. Ruchirawat, J. Schlatter, J. Seed, and C. Vickers. 2008. IPCS framework for analyzing the relevance of a noncancer mode of action for humans. Crit. Rev. Toxicol. 38(2):87–96.

Bradbury, S.P. 1994. Predicting modes of toxic action from chemical structure: an overview. SAR QSAR Environ. Res. 2:89–104.

Bradbury, S.P., C.L. Russom, G.T. Ankley, T.W. Schultz, and J.D. Walker. 2003. Overview of data and conceptual approaches for derivation of quantitative structure-activity relationships for ecotoxicological effects of organic chemicals. Environ. Toxicol. Chem. 22(8):1789–1798.

Bradbury, S.P., T.C.J. Feijtel, and C.J. van Leeuwen. 2004. Meeting the scientific needs of ecological risk assessment in a regulatory context. Environ. Sci. Technol. 38(23):463A–470A.

Brown, T.N. and F. Wania. 2008. Screening chemicals for the potential to be persistent organic pollutants. A case study of arctic contaminants. Environ. Sci. Technol. 42:5202–5209.

Clark, M. 2005. Generalized fragment-substructure based property prediction method. J. Chem. Inf. Model. 45(1):30–38.

Contrera, J.F., N.L. Kruhlak, E.J. Matthews, and R.D. Benz. 2007. Comparison of MC4PC and MDL-QSAR rodent carcinogenicity predictions and the enhancement of predictive performance by combining QSAR models. Reg. Toxicol. Pharm. 49:172–182.

CCA. 2012. Integrating Emerging Technologies into Chemical Safety Assessment. The Expert Panel on the Integrated Testing of Pesticides. Council of Canadian Academies. http://www.scienceadvice.ca/en/assessments/completed/pesticides.aspx

Cronin M. 2010. Quantitative structure-activity relationships (QSARs) — applications and methodology. Chapter 10 in Recent Advances in QSAR Studies: Methods and Applications. T. Puzyn, J. Leszczynski, and M.T.D. Cronin (eds). Springer, Heidelberg, Germany, pp. 3–11.

Cronin, M.T.D., J.D. Walker, J.S. Jaworska, M.H.I. Comber, C.D. Watt, and A.P. Worth. 2003. Use of QSARs in International Decision-Making Frameworks to Predict Ecologic Effects and Environmental Fate of Chemical Substances. Environ Health Perspect 111:1376–1390.

Daginnus, K., S. Gottardo, A. Mostrag-Szlichtyng, H. Wilkinson, P. Whitehouse, A. Paya-Perez, and J.M. Zaldivar. 2009. A modeling approach for the prioritization of chemicals under the water framework directive. European Commission, Joint Research Centre, EUR 24292 EN, 54 p. Ispra, Italy

Dalby, A., J.G. Nourse, W.D. Hounshell, A.K.I. Gushurst, D.L. Grier, B.A. Leland, and J. Laufer. 1992. Description of several chemical structure file formats used by computer programs developed at molecular design limited. J. Chem. Inf. Comput. Sci. 32:244–255.

Daylight Chemical Information Systems. 2008. Simplified Molecular Input Line Entry System (SMILES™) Daylight Chemical Information Systems, Inc. http://www.daylight.com/smiles/index.html

Dearden, J.C., M.T.D. Cronin, and K.L.E. Kaiser. 2009. How not to develop a quantitative structure-activity or structure-property relationship (QSAR/QSPR). SAR and QSAR in Environmental Research 20:241–266.

Dearden, J. and A. Worth. 2007. In silico prediction of physicochemical properties. European Commission, Joint Research Centre, EUR 23051 EN, 73 p. Ispra, Italy

Demchuk, E., P. Ruiz, J.D. Wilson, F. Scinicariello, H.R. Pohl, M. Fay, M.M. Mumtaz, H. Hansen, and C.T. De Rosa. 2008. Computational toxicology methods in public health practice. Toxicol. Mech. Methods 18:119–135.

Dimitrov, S.D., G.D. Dimitrova, T.S. Pavlov, N. Dimitrova, G.Y. Patlewicz, J. Niemela, and O.G. Mekenyan. 2005a. A stepwise approach for defining the applicability domain of SAR and QSAR models. J. Chem. Inf. Model. 45:839–849.

Dimitrov, S.D., L.K. Low, G.Y. Patlewicz, P.S. Kern, G.D. Dimitrova, M.H.I. Comber, R.D. Philips, J. Niemela, P.T. Bailey, and O.G. Mekenyan. 2005b. Skin sensitization: Modeling based on skin metabolism simulation and formation of protein conjugates. Int. J. Toxicol. 24:189–204.

Doull, J., J.F. Borzelleca, R. Becker, G. Daston, J. DeSesso, A. Fan, P. Fenner-Crisp, M. Holsapple, J. Holson, G.C. Llewellyn, J. MacGregor, J. Seed, I. Walls, Y.-T. Woo, and S. Olin. 2007. Framework for use of toxicity screening tools in context-based decision making. Food and Chem. Toxicol. 45:759–796.

EC. 1995a. Overview of structure-activity relationships for environmental endpoints. Part 1: General outline and procedures. European Commission. Report of the EU-DG-XII Project QSAR for predicting fate and effects of chemical in the environment. Contract No. EV5V-CT92-0211. 149 p. Ispra, Italy. http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/information-sources/qsar-document-area/QSAR_environ_background.pdf

EC. 1995b. Overview of structure-activity relationships for environmental endpoints. Part 2: Description of selected models. Report of the EU-DG-XII Project QSAR for predicting fate and effects of chemical in the environment. European Commission. Contract No. EV5V-CT92-0211. 216 p. Ispra, Italy. http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/information-sources/qsar-document-area/QSAR_environ_models.pdf

EC. 2008a. QSAR Model Reporting Format (Version 1.2). European Commission Directorate General Joint Research Centre. Institute for Health and Consumer Protection. Toxicology and Chemical Substances Unit. Ispra 26/05/2008. http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/qsar_tools/qrf/QMRF_version_1.2.pdf

EC. 2008b. QSAR Prediction Reporting Format (QPRF) (version 1.1, May 2008). European Commission Directorate General Joint Research Centre. Institute for Health and Consumer Protection. Toxicology and Chemical Substances Unit. Ispra 26/05/2008. http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/qsar_tools/qrf/QPRF_version_1.1.pdf

EC. 2010. Applicability of QSAR analysis to the evaluation of the toxicological relevance of metabolites and degradates of pesticide active substances for dietary risk assessment. Scientific report submitted to EFSA. Prepared by Computational Toxicology Group, Institute for Health & Consumer Protection, European Commission — Joint Research Centre, Ispra, Italy. http://www.efsa.europa.eu/en/supporting/pub/50e.htm

ECHA, 2010. Practical Guide 5. How to report (Q)SAR. European Chemicals Agency. http://echa.europa.eu/documents/10162/13655/pg_report_qsars_en.pdf.

ECHA. 2008. Guidance on information requirements and chemical safety assessment. Chapter R.6: QSARs and grouping of chemicals. May 2008. Guidance for the implementation of REACH. European Chemicals Agency. http://echa.europa.eu/documents/10162/13632/information_requirements_r6_en.pdf

Ellison, C.M., J. C. Madden, P.N. Judson, and M.T.D. Cronin. 2010. Application of in silico tools in a weight of evidence approach to aid toxicological assessment. Poster 4.4. 14th International Workshop on Quantitative Structure-Activity

Relationships in Environmental and Health Sciences. May 24–28, 2010. Montreal, Quebec, Canada.

Enoch, S.J. 2010. Chapter 7. Chemical category formation and read-across for the prediction of toxicity. In: Puzyn, T., J. Leszczynski, and M.T. Cronin (eds.) Recent Advances in QSAR Studies, Springer, pp. 209–219.

Enoch, S.J., M.T.D. Cronin, J.C. Madden, and M. Hewitt. 2009. Formation of structural categories to allow for read-across for teratogenicity. QSAR Comb. Sci. 28(6-7):696–708.

Eriksson, L., J. Jaworska, A.P. Worth, M.T.D. Cronin, R.M. McDowell, and P. Gramatica. 2003. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. Environ. Health Perspect. 111(10):1361–1375.

Ferguson, J. 1939. The use of chemical potentials as indices of toxicity. Proc R Soc Lond Series B Biol Sci 127:387–404.

Gallegos-Saliner, A., A. Poater, N. Jeliazkova, G. Patlewicz, and A.P. Worth. 2008. Toxmatch — A chemical classification and activity prediction tool based on similarity measures. Reg. Toxicol. Pharm. 52:77-84.

Geiss, K.T. and J.M. Frazier. 2001. QSAR modeling of oxidative stress in vitro following hepatocyte exposures to halogenated methanes. Toxicology in Vitro 15(4-5), 557–563.

Gramatica, P. 2007. Principles of QSAR models validation: Internal and external. QSAR & Comb. Sci. 26(5):694–701.

Gramatica, P. and E. Papa. 2007. Screening and ranking of POPs for global half-life: QSAR approaches for prioritization based on molecular structure. Environ. Sci. Technol. 41:2833–2839.

Gramatica, P., E. Giani, and E. Papa. 2007. Statistical external validation and consensus modeling: A QSPR case study for K_{OC} prediction. J. Mol. Graph. Model. 25:755–766.

Greene, N. 2002. Computer systems for the prediction of toxicity: an update. Adv. Drug Deliv. Rev. 54:417–431.

Hansch, C. and T. Fujita. 1964. A method for the correlation of biological activity and chemical structure. J. Am. Chem. Soc. 86:1616–1626.

Hansch, C. and W.J. Dunn. 1972. Linear relationships between lipophilic character and biological activity of drugs. J. Pharm. Sci. 61:1–19.

Health Canada. 2011. Weight of Evidence: General Principles and Current Applications at Health Canada. November 2011. Prepared by the Task Force on Scientific Risk Assessment's Weight of Evidence Working Group. Tao, T., Y. Bhuller, Y. Bonvalot, M. Hill, A. Klein, G. Kozak, I. Plante, and N. Robert.

Hester, S.D., D.C. Wolf, S. Nesnow, and S.F. Thai. 2006 Transcriptional profiles in liver from rats treated with tumorigenic and non-tumorigenic triazole conazole fungicides: Propiconazole, triadimefon, and myclobutanil. Toxicol. Pathol. 34:879–94.

Hewitt, M., C.M. Ellison, E.J. Enoch, J.C. Madden, and M.T.D. Cronin. 2010. Integrating (Q)SAR models, expert systems and read-across approaches for the prediction of developmental toxicity. Repro. Tox. 30:147–160.

Hewitt, M., M.T.D. Cronin, J.C. Madden, P.H. Rowe, C. Johnson, A. Obi, and S.J. Enoch. 2007. Consensus QSAR Models: Do the Benefits Outweigh the Complexity? J. Chem. Inf. Model 47:1460–1468.

Hulzebos, E.M., P.A.H. Janssen, L. Maślankiewicz, M.C.M. Mijerink, J.J.A. Muller, S.M.G. Pelgrom, L. Verdam, and T.G. Vermeire. 2001. The application of structure-activity relationships in human hazard assessment: a first approach. National Institute of Public Health and the Environment. RIVM report 601516 008. http://rivm.openrepository.com/rivm/bitstream/10029/9562/1/601516008.pdf

Hulzebos, E.M., P.C.J.I. Schielen, and L. Wijkhuizen-Maslankiewicz. 1999. (Q)SARs for human toxicological endpoints: a literature search. RIVM, National Institute of Public health and the Environment Report No. 601516.001, 65 p. Bilhoven, The Netherlands. http://www.rivm.nl/bibliotheek/rapporten/601516001.pdf

IUPAC. 2010a. Nomenclature and Symbols. International Union of Pure and Applied Chemistry. http://old.iupac.org/general/FAQs/ns.html

IUPAC. 2010b. The IUPAC International Chemical Identifier (InChI™). International Union of Pure and Applied Chemistry. http://www.iupac.org/home/publications/e-resources/inchi.html

Jaworska, J., N. Nikolova-Jelizkova, and T. Aldenberg. 2005. Review of methods for QSAR applicability domain estimation by the training set. A.T.L.A. 33:445–459.

Jaworska, J.S., M. Comber, C. Auer, and C.J. van Leeuwen. 2003. Summary of a workshop on regulatory acceptance of (Q)SARs for human health and environmental endpoints. Environ. Health Perspect. 111(10):1358–1360.

Jensen, G.E., J.R. Niemela, E.B. Wedebye, and N.G. Nikolov. 2008. QSAR models for reproductive toxicity and endocrine disruption in regulatory use — a preliminary investigation. SAR QSAR Environ. Res. 19(7–8):631–641.

Judson,R., A. Richard, D.J. Dix, K.Houck, M. Martin, R. Kavlock, V. Dellarco, T. Henry, T. Holderman, P. Sayre, S. Tan, T. Carpenter, and E. Smith. 2009. The Toxicity Data Landscape for Environmental Chemicals. Environ. Health Perspect. 117(5):685–695.

Judson, R.S., K.A. Houck, R.J. Kavlock, T.B. Knudsen, M.T. Martin, H.M. Mortensen, D.M. Reif, D.M. Rotroff, I. Shah, A.M. Richard, and D.J. Dix. 2010. In vitro screening of environmental chemicals for targeted testing prioritization: The ToxCast Project. Environ. Health Perspect. 118(4):485–492.

Karelson, M., V.S. Lobanov, and A.R. Katritzky. 1996. Quantum-chemical descriptors in QSAR/QSPR studies. Chem. Rev. 96(3):1027–1043.

Klopman, G. and S.K. Chakravarti. 2003a. Structure-activity relationship study of a diverse set of estrogen receptor ligands (I) using MultiCASE expert system. Chemosphere 51:445–459.

Klopman, G. and S.K. Chakravarti. 2003b. Screening of high production volume chemicals for estrogen receptor binding activity (II) by the MultiCASE expert system. Chemosphere 51:461–468.

Klopman, G., S.K. Chakravarti, N. Harris, J. Ivanov, and R.D. Saiakhov. 2003. In-silico screening of high production volume chemicals for mutagenicity using the MCASE QSAR expert system. SAR QSAR Environ. Res. 14:165–180.

Lai, D.Y., Y.-T. Woo, M.F. Argus, and J.C. Arcos. 1996. Cancer risk reduction through mechanism-based molecular design of chemicals. In: Designing Safer Chemicals: Green Chemistry for Pollution Prevention (DeVito, S.C., and R.L. Garrett, eds.). ACS Symposium Series 640. Washington, DC:American Chemical Society, 1996; 62–73.

Leonard, J.T. and K. Roy. 2006. On selection of training and test sets for development of predictive QSAR models. QSAR Comb. Sci. 25(3):235–251.

Lewis, D.F.V., M.G. Bird, and M.N. Jacobs. 2002. Human carcinogens: an evaluation study via the COMPACT and Hazard Expert procedures. Hum. Exp. Toxicol. 21:115–122.

Lipnick, R.L. 1995a. Structure-activity relationships. In: Rand, G.M., ed., Fundamentals of Aquatic Toxicology, 2nd ed. Taylor&Francis, Washington, DC, pp 609–655.

Lipnick, R.L., 1995b. Hans Horst Meyer and the lipoid theory of narcosis. Trends Pharmacol. Sci. 10(7):265–269.

Lo Piparo, E., A. Worth, M. Manibusan, C. Yang, B. Schilter, P. Mazzatorta, M.N. Jacobs, H. Steinkellner, and L. Mohimont. 2011. Use of computational tools in the field of food safety. Reg. Toxicol. Pharm. 60:354–362.

MacKay, D., S. Paterson, and M. Joy. 1983. Application of fugacity models to the estimation of chemical distribution and persistence in the environment. In: Swann, R.L. and A. Eschenroeder, eds., Fate of Chemicals in the Environment. American Chemical Society Symposium Series, Vol. 225, Chapter 9, pp 175–196.

Matthews, E.J. 2009b. Meeting on the Development of a General Guidance Document on the Application of QSAR Tools to Pesticides. July 10, 2009. 9:00am – 5:00pm Agenda/Meeting Notes.

Matthews, E.J. and J.F. Contrera. 1998. A new highly specific method for predicting the carcinogenic potential of pharmaceuticals in rodents using enhanced MCASE QSAR-ES software. Reg. Toxicol. Pharm. 28:242–264.

Matthews, E.J., C.J. Ursem, N.L. Kruhlak, R.D.Benz, D.A. Sabaté, C. Yang, G. Klopman, and J.F. Contrera. 2009a. Identification of structure-activity relationships for adverse effects of pharmaceuticals in humans: Part B. Use of (Q)SAR systems for early detection of drug-induced hepatobiliary and urinary tract toxicities. Reg. Toxicol. Pharm. 54:23–42.

Matthews, E.J., N.L. Kruhlak, R.D. Benz, and J.F. Contrera. 2008. Combined use of MC4PC, MDL-QSAR, BioEpisteme, Leadscope PDM, and Derek for Windows software to achieve high-performance, high-confidence, mode of action-based predictions of chemical carcinogenesis in rodents. Toxicol. Mech. Method. 18:189–206.

Matthews, E.J., N.L. Kruhlak, R.D. Benz, and J.F. Contrera. 2007a. A comprehensive model for reproductive and developmental toxicity hazard identification: I. Development of a weight of evidence QSAR database. Reg. Toxicol. Pharm. 47:115–135.

Matthews, E.J., N.L. Kruhlak, R.D.Benz, J. Ivanov, G. Klopman, and J.F. Contrera. 2007b. A comprehensive model for reproductive and developmental toxicity hazard identification: II. Construction of QSAR models to predict activities of untested chemicals. Reg. Toxicol. Pharm. 47:136–155.

Mekenyan, O., S. Dimitrov, N. Dimitrova, G. Dimitrova, T. Pavlov, G. Chankov, S. Kotov, K. Vasilev, R. Vasilev. 2006. Metabolic activation of chemicals: in-silico simulation. SAR QSAR Environ. Res. 17(1):107–120.

Monev, V. 2004. Introduction to similarity searching in chemistry. Match-Comm. Math. Comp. Chem. 51:7–38.

Mostrag-Szlichtyng, A. and A. Worth. 2010b. Review of QSAR Models and Software Tools for predicting Biokinetic Properties. JRC Technical Report EUR 24377 EN. Publications Office of the European Union, Luxembourg. http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/publications/

Mostrag-Szlichtyng, A. and A.P. Worth. 2010a. In silico modelling of microbial and human metabolism: a case study with the fungicide carbendazim. JRC Scientific and Technical Report EUR 24523 EN. Publications Office of the European Union, Luxembourg. http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/publications/

Mostrag-Szlichtyng, A., J.-M. Zaldívar Comenges, and A.P. Worth. 2010. Computational toxicology at the European Commission's Joint Research Centre. Expert Opinion on Drug Metabolism and Toxicology 6, 785–792.

National Research Council (NRC). 1983. Risk assessment in the federal government. Managing the process. National Academy Press, Washington, D.C.

National Research Council (NRC). 2007. Toxicity testing in the 21st century: A vision and a strategy. National Academy of Sciences, Washington, DC. http://dels.nas.edu/resources/static-assets/materials-based-on-reports/reports-in-brief/Toxicity_Testing_final.pdf

Netzeva, T., M. Pavan, and A. Worth. 2007. Review of data sources, QSARs, and integrated testing strategies for aquatic toxicity. European Commission, Joint Research Centre, EUR 22943 EN, 103 p. Ispra, Italy. http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/doc/EUR_22943_EN.pdf

Netzeva, T.I., A. Worth, T. Aldenberg, R. Benigni, M.T. Cronin, P. Gramatica, J.S. Jaworska, S. Kahn, G. Klopman, C.A. Marchant, G. Myatt, N. Nikolova-Jeliazkova, G.Y. Patlewicz, R. Perkins, D. Roberts, T. Schultz, D.W. Stanton, J.J. van de Sandt, W. Tong, G. Veith, and C. Yang. 2005. Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure-Activity Relationships. The report and recommendations of ECVAM Workshop 52. Altern. Lab. Anim. 33(2):155–173.

Nikolova, N. and J. Jaworska. 2003. Approaches to measure chemical similarity — a review. QSAR & Comb. Sci. 22:1006–1026.

OECD. 2005. The OECD Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment (OECD

series on testing and assessment No. 34). OECD Environment Directorate Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology. ENV/JM/MONO(2005)14. 18-Aug-2005. http://www.oecd.org/officialdocuments/displaydocumentpdf?cote=env/jm/mono(2005)14&doclanguage=en

OECD. 2007a. Guidance on grouping of chemicals. OECD Environmental Health and Safety Publications. Series on Testing and Assessment No. 80. ENV/JM/MONO (2007)28: 99 p. http://www.oecd.org/officialdocuments/displaydocumentpdf/?cote=env/jm/mono(2007)28&doclanguage=en

OECD. 2007b. Report on the regulatory uses and applications in OED member countries of (quantitative) structure-activity relationship [(Q)SAR] models in the assessment of new and existing chemicals. OECD Environmental Health and Safety Publications. Series on Testing and Assessment No. 58. ENV/JM/MONO (2006)25: 79 p. http://www.oecd.org/dataoecd/55/22/38131728.pdf

OECD. 2007c. The OECD Guidance Document on the Validation of (Quantitative) Structure-Activity Relationships [(Q)SAR] Models. OECD Environmental Health and Safety Publications. Series on Testing and Assessment No. 69. OECD Environment Directorate Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology. ENV/JM/MONO(2007)2. 30-May-2007. http://www.oecd.org/officialdocuments/displaydocumentpdf?cote=ENV/JM/MONO(2007)2&doclanguage=en

OECD. 2009. Guidance document for using the OECD QSAR application toolbox to develop chemical categories according to the OECD guidance on grouping of chemicals. OECD Series on Testing and Assessment No. 102. ENV/JM/MONO (2009)5: 118 p. http://www.oecd.org/dataoecd/50/60/42294034.pdf

OECD. 2011a. OECD Quantitative Structure-Activity Relationships Project [(Q)SARs]. http://www.oecd.org/document/23/0,2340,en_2649_34365_33957015_1_1_1_1,00.html

OECD. 2011b. Report of the Workshop on Using Mechanistic Information in Forming Chemical Categories. No. 138. Paris 2011.

OECD.2004. OECD Principles for the validation, for regulatory purposes, of (Quantitative) Structure-Activity Relationship Models. Organization for Economic Cooperation and Development. http://www.oecd.org/dataoecd/33/37/37849783.pdf

Overton, C.E. 1901. Studien über die Narkose. Gustav Fischer, Jena, Germany.

Patlewicz, G., A. Gallegos Saliner, M. Pavan, A. Worth, R. Benigni, A. Aptula, A. Bassan, C. Bossa, A. Galk-Filipsson, V. Gillet, N. Jeliazkova, A. McDougal, J. Mestres, I. Munro, T. Netzeva, B. Safford, B. Simon-Hettich, I. Tsakovska, M. Wallen, and C. Yang. 2005. Chemical Similarity and Threshold of Toxicological Concern (TTC) Approaches. Report of an ECB Workshop held in Ispra, November 2005 European Commission, Joint Research Centre, EUR 22657 EN, 43 p. Ispra, Italy.
http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/doc/EUR_22657_EN.pdf

Patlewicz, G., A.O. Aptulu, D.W. Roberts, and E. Uriarte. 2008. A minireview of available skin sensitization (Q)SARs/Expert Systems. QSAR Combin. Sci. 27(1):60–76.

Pavan, M., A.P. Worth, and T.I. Netzeva. 2005a. Preliminary analysis of an aquatic toxicity dataset and assessment of QSAR models for narcosis. European Commission, Joint Research Centre, EUR 21749 EN, 42 p. Ispra, Italy.
http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/doc/Report_QSAR_model_for_narcosis.pdf

Pavan, M., A.P. Worth, and T.I. Netzeva. 2005b. Comparative assessment of QSAR models for aquatic toxicity. European Commission, Joint Research Centre, EUR 21750 EN, 122 p. Ispra, Italy.
http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/doc/Report_Comparative_assessment_QSAR_models.pdf

Pavan, M., and A. Worth. 2008. A set of case studies to illustrate the applicability of DART (Decision Analysis by Ranking Techniques) in the ranking of chemicals. European Commission, Joint Research Centre, EUR 23481 EN, 80 p. Ispra, Italy.
http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/doc/EUR_23481_EN.pdf

Perkins, R., H. Fang, W. Tong, and W.J. Welsh. 2003. Quantitative structure-activity relationship methods: Perspectives on drug discovery and toxicology. Environ. Toxicol. Chem. 22(8):1666–1679.

Richard, A.M., C. Yang, and R. Judson. 2008. Toxicity data informatics: Supporting a new paradigm for toxicity prediction. Toxicol. Mechan.Methods 18(2 & 3):103–118.

Richard, A.M., I. Swirsky Gold, and M.C. Nicklaus. 2006. Chemical structure indexing of toxicity data on the internet: moving towards a flat world. Curr. Opinion Drug Disc.Dev. 9(3):314–325.

Ringeissen, S., L. Marrot, R. Note, A. Labarussiat, S. Imbert, M. Todorov, O. Mekenyan, and J. Meunier. 2010. Development of a mechanistic SAR model for the detection of phototoxic chemicals and use in an integrated testing strategy. Toxicol. in Vitro (2010), doi:10.1016/ j.tiv.2010.09.017.

Rorije, E. and E. Hulzebos. 2005. Evaluation of (Q)SARs for the prediction of skin irritation/corrosion potential. RIVM, National Institute of Public health and the Environment. September 2005. 65 p. Bilhoven, The Netherlands. 45 p. http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/information-sources/qsar-document-area/Evaluation_of_Skin_Irritation_QSARs.pdf

Ruggeri, B. 2009. Chemical exposure: Scoring procedure and uncertainty propagation in scenario selection for risk analysis. Chemosphere 77:330–338.

Russom, C.L., R.L. Breton, J.D. Walker, and S.P. Bradbury. 2003. An overview of the use of quantitative structure-activity relationships for ranking and prioritizing large chemical inventories for environmental risk assessments. Environ. Toxicol. Chem. 22(8):1810–1821.

Russom, C.L., S.P. Bradbury, S.J. Broderius, D.E. Hammermeister, and R.A. Drummond. 1997. Predicting modes of toxic action from chemical structure: Acute toxicity in the fathead minnow (Pimephales promelas). Environ. Toxicol. Chem. 16(5):948–967.

Saliner, G., G.Patlewicz, and A.P. Worth. 2005. A similarity based approach for chemical category classification. European Commission, Joint Research Centre Report number: EUR 21867 EN. Institute for Health and Consumer Protection, Toxicology and Chemical Substances Unit, European Chemicals Bureau, I-21020 Ispra (VA) Italy. http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/doc/Report_Chemical_Similarity_for_Category_Classification.pdf

Schmieder, P., O. Mekenyan, S. Bradbury, and G. Veith. 2003c. QSAR prioritization of chemical inventories for endocrine disruptor testing. Pure Appl. Chem. 75:2389–2396.

Schmieder, P.K., G.T. Ankley, O. Mekenyan, J.D. Walker, and S. Bradbury. 2003b. Quantitative structure-activity relationship models for prediction of estrogen receptor binding affinity of structurally diverse chemicals. Environ. Toxicol. Chem. 22:1844–1854.

Schmieder, P.K., M.A. Tapper, J.S. Denny, R.C. Kolanczyk, B.R. Sheedy, T.R. Henry, and G.D. Veith. 2004. Use of trout liver slices to enhance mechanistic interpretation of estrogen receptor binding for cost-effective prioritization of chemicals within large inventories. Environ. Sci. Technol. 38:6333–6342.

Schmieder, P.K., M.A. Tapper, R.C. Kolanczyk, D.E. Hammermeister, B.R. Sheedy, and J.S. Denny. 2003a. Discriminating redox cycling and arylation pathways of reactive chemical toxicity in trout hepatocytes. Toxicol. Sci. 72:66–76.

Schultz, T. and B. Diderich. 2011. Adverse outcome pathways and their role in helping to formulate mechanistically relevant chemical categories. AltTox.org Non-animal Methods for Toxicity Testing. http://alttox.org/spotlight/050.html

Tong, W., W.J. Welsh, L. Shi, H. Fang, and R. Perkins. 2003. Structure-activity relationship approaches and applications. Environ. Toxicol. Chem. 22(8):1680–1695.

Tsakovska, I., I. Lessigaiarska, T. Netzeva, and A. Worth. 2008. A mini review of mammalian toxicity (Q)SAR models. QSAR & Combin. Sci. 27(1):41–48.

Tsakovska, I., T. Netzeva, and A. Worth. 2005. Evaluation of (Q)SARs for the prediction of eye irritation/corrosion potential. Physicochemical exclusion rules. European Commission, Joint Research Centre, EUR 21897 EN, 42 p. Ispra, Italy. http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/doc/Evaluation_of_Eye_Irritation_QSARs.pdf

Urbano-Cuadrado, M., I.L. Ruiz, and M.A. Gomez-Nieto. 2008. Description and application of similarity-based methods for fast and simple QSAR model development. QSAR Comb. Sci. 27(4):457–468.

Ursem, C.J., N.L. Kruhlak, J.F. Contrera, P.M. MacLaughlin, R.D. Benz, and E.J. Matthews. 2009. Identification of structure-activity relationships for adverse effects of pharmaceuticals in humans. Part A: Use of FDA post-market reports to create a database of hepatobiliary an urinary tract toxicities. Reg. Toxicol. Pharm. 54:1–22.

US EPA. 1998. Guidelines for Ecological Risk Assessment (Published on May 14, 1998, Federal Register 63(93):26846-26924). Risk Assessment Forum. EPA/630/R-95/002F April 1998.

US EPA. 1999. Development of chemical categories in the HPV Challenge Program. Office of Pollution Prevention and Toxics. 21 p.

US EPA, 2007. Dermal Exposure Assessment: A Summary of EPA Approaches. National Center for Environmental Assessment, Office of Research and Development, Washington, DC 20460. http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=183584

US EPA. 2010. TSCA New Chemicals Program (NCP) Chemical Categories. Office of Pollution Prevention and Toxics, Washington, DC 20460. http://www.epa.gov/oppt/newchems/pubs/npcchemicalcategories.pdf

US EPA. 2011. Aster (Assessment Tools for the Evaluation of Risk). National Health and Environmental Effects Research Laboratory (NHEERL). http://www.epa.gov/med/Prods_Pubs/aster.htm

Valkó K. 2004. Application of high-performance liquid chromatography based measurements of lipophilicity to model biological distribution. J. Chromatog. A 1037(1-2):299–310.

van der Jagt, K., S. Munn, J. Torslov, and J. de Bruijn. 2004. Alternative approaches can reduce the use of test animals under REACH. European Commission, Joint Research Centre, Institute for Health Consumer Protection. EUR 21405 EN. 31 p. http://publications.jrc.ec.europa.eu/repository/bitstream/111111111/8790/1/EUR%2021405%20EN.pdf

van Leeuwen, C.J., G.Y. Patlewicz, and A.P. Worth. (2007). Intelligent testing strategies. In: Risk Assessment of Chemicals: An Introduction. 2nd Edition. Van Leeuwen, C.J. and T.G. Vermeire, (eds.).

van Leeuwen, K. T.W. Schultz, T. Henry, B. Diderich, and G.D. Veith. 2009. Using chemical categories to fill data gaps in hazard assessment. SAR QSAR Environ. Res. 20(3-4):207–220.

Veith, G.D., D.L. De Foe, and B.V. Bergstedt. 1979. Measuring and estimating the bioconcentration factor of chemicals in fish. J. Fish Res. Board Can. 36:1040–1047.

Veith, G.D., D.J. Call, and L.T. Brooke. 1983. Structure-toxicity relationships for the fathead minnow, Pimephales promelas: Narcotic industrial chemicals. Can. J. Fish. Aquat. Sci. 40:743–748.

Vonk, A.J., R. Benigni, M. Hewitt, M. Nendza, H. Segner, D. van de Meent, D. and M.T.D. Cronin. 2009. The Use of Mechanisms and Modes of Toxic Action in Integrated Testing Strategies: The Report and Recommendations of a Workshop held as part of the European Union OSIRIS Integrated Project. ATLA 37, 557–571.

Walker, J.D., D. Knaebel, K. Mayo, J. Tunkel, and D.A. Gray. 2004. Use of QSARs to promote more cost-effective use of chemical monitoring resources. 1. Screening industrial chemicals and pesticides, direct food additives, indirect food additives and pharmaceuticals for biodegradation, bioconcentration and aquatic toxicity potential. Water Qual. Res. J. Canada 29(1):35–39.

Walker, J.D., J. Jaworska, M.H.I. Comber, T.W. Schultz, and J.C. Dearden. 2003. Guidelines for developing and using quantitative structure-activity relationships. Environ. Toxicol. Chem. 22(8):1653–1665.

Walker, JD. 2003. Quantitative structure-activity relationship for pollution prevention, toxicity screening, risk assessment, and web applications. Pensacola, FL, USA: Society of Environmental Toxicology and Chemistry (SETAC). 282 p.

Weininger, D. 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J. Chem. Inform. Comput. Sci. 28:31–36.

WHO/FAO. 1996. Fenvalerate. WHO/FAO Data Sheets on Pesticides No. 90. World Health Organization. Food and Agriculture Organization of the United Nations. WHO/PCS/DS/96.90.
http://www.inchem.org/documents/pds/pds/pest90_e.htm

Williams, C., M. Wolf, and A.M. Richard. 2009. DSSTox chemical-index files for exposure-related experiments in ArrayExpress and Gene Expression Omnibus: enabling toxico-chemogenomics data linkages. Bioinformatics 25(5):692–694.

Woo, Y.T., and D.Y. Lai. 2010. (Q)SAR analysis of genotoxic and nongenotoxic carcinogens: a state-of-the-art overview. In: Cancer Risk Assessment: Chemical Carcinogenesis, Hazard Evaluation, and Risk Quantification, edited by Hsu, C.H. and T. Stedeford, Wiley. Chapter 20, 517–556.

Woo, Y.-T., D.Y. Lai, J.C. Arcos, and M.F. Argus. 1985. Chemical Induction of Cancer, Structural Bases and Biological Mechanisms. Vol IIIB: Aliphatic and Polyhalogenated Carcinogens. Orlando, FL: Academic Press.

Woo, Y.-T., D.Y. Lai, J.L. McLain, M.K. Manibusan, and V. Dellarco. 2002. Use of Mechanism-Based Structure-Activity Relationships Analysis in Carcinogenic Potential Ranking for Drinking Water Disinfection By-Products. Environ. Health Perspect. 110(Suppl 1):75–87.

Woo, Y.-T., D.Y. Lai, M.F. Argus, and J.C. Arcos. 1998. Integrative approach of combining mechanistically complementary short-term predictive tests as a basis for assessing the carcinogenic potential of chemicals. Environ. Carcinog. Ecotoxicol. Rev. (Part C of J. Environ. Sci. Health) C16(2):101–122.

Worth, A., and G. Patlewicz. 2007. A compendium of case studies that helped to shape the REACH guidance on chemical categories and read across. European Commission, Joint Research Centre, EUR 22481 EN. 188 p.
http://publications.jrc.ec.europa.eu/repository/bitstream/111111111/1234/1/7212%20-%20Compendium%20of%20case%20studies_final_060707.pdf

Worth, A., M. Fuart-Gatnik, S. Lapenna, and R. Serafimova. 2011. Applicability of QSAR analysis in the evaluation of developmental and neurotoxicity effects for the assessment of the toxicological relevance of metabolites and degradates of pesticide active substances for dietary risk assessment. Report produced for EFSA.
http://www.efsa.europa.eu/en/supporting/pub/169e.htm

Worth, A., S. Lapenna, E. Lo Piparo, A. Mostrag-Szlichtyng, and R. Serafimova. 2010. The Applicability of Software Tools for Genotoxicity and Carcinogenicity Prediction: Case Studies relevant to the Assessment of Pesticides. JRC Technical Report EUR 24640 EN. Publications Office of the European Union, Luxembourg.
http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/publications/

Worth, A., S. Lapenna, E. Lo Piparo, A. Mostrag-Szlichtyng, and R. Serafimova. 2011. A Framework for assessing in silico Toxicity Predictions: Case Studies with selected Pesticides. JRC report EUR 24705 EN. Publications Office of the European Union, Luxembourg.
http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/publications/

Worth, A.P. 2010. The role of QSAR methodology in the regulatory assessment of chemicals, Chapter 13 in Recent Advances in QSAR Studies: Methods and Applications. Puzyn, T., J. Leszczynski, and M.T.D. Cronin (eds). Springer, Heidelberg, Germany, pp. 367–382.

Worth, A.P. and A. Mostrag-Szlichtyng. 2011. Towards a Common Regulatory Framework for Computational Toxicology: Current Status and Future Perspectives, in New Horizons in Predictive Toxicology: Methods and Applications. Wilson, A.G.E. (ed). The Royal Society of Chemistry, Cambridge, UK, pp 38–69.

Yuan, H., Y. Wang, and Y. Cheng. 2007. Local and Global Quantitative Structure−Activity Relationship Modeling and Prediction for the Baseline Toxicity. J. Chem. Inf. Model., 2007, 47 (1), pp 159–169.

Zeeman, N., C.M. Auer, R.G. Clements, J.V. Nabholz, and R.S. Boethling. 1995. U.S. EPA regulatory perspectives on the use of QSAR for new and existing chemical evaluations. SAR/QSAR Environ. Res. 3:179–201.

Zvinavashe, E., A.J. Murk, and I.M.C.M. Rietjens. 2008. Promises and pitfalls of quantitative structure-activity relationship approaches for predicting metabolism and toxicity. Chem. Res. Toxicol. 21:2229–2236.

Appendix I

Where To Go to Learn More About (Q)SAR

Listed below are world wide webpages (WWW) for a number of national and international organizations involved in projects and the development of tools related to (Q)SAR and the assessment of risks from chemicals. In most cases, the descriptions of the activities and projects come directly from the webpages. These links may be useful to pesticide evaluators who are seeking additional information on general (Q)SAR concepts, and the development, validation, and evaluation of (Q)SAR tools and predictions. Some of the links and information below have been cited in various sections of this guidance document.

The list below is by no means exhaustive, and since the field of (Q)SAR is constantly expanding, pesticide evaluators are advised to regularly monitor various national and international agency websites and the open literature for developments in the area of (Q)SAR of interest to them.

Danish Ministry of the Environment Environmental Protection Agency (Danish EPA) (Q)SAR — Assessment of Chemical Properties of Substances

http://www.mst.dk/English/Chemicals/assessment_of_chemicals/qsar_assessment_chemical_properties_of_substances/

Descriptions of the key (Q)SAR activities from the Danish EPA webpage (*i.e.*, the (Q)SAR database and the Advisory list for self-classification of dangerous substances) are provided below.

“The Danish EPA has for a number of years worked with the development and use of (Q)SAR’s, also called ‘computer models’ for prediction of properties of chemical substances. (Quantitative) Structure Activity Relationships — (Q)SAR — are relations between structure properties of chemical substances and some other property. The other property can be a physical-chemical property or a biological activity, including the ability to cause toxic effects.”

“The Danish EPA has made a database, which comprise predictions from more than 70 (Q)SAR models on endpoints for physico-chemical properties, fate, eco-toxicity, absorption, metabolism and toxicity. The database is constantly growing as new models are obtained and developed. More than half of all the estimates

are for mammalian (human) toxicity endpoints and include commercial data sets from TOPKAT and MULTICASE as well as many models developed in-house."

"Lack of data on hazardous properties of chemicals makes it difficult for companies to meet their obligations to self classify the chemicals they import or produce. To address this issue, The Danish Environmental Protection Agency (DK-EPA) publishes the advisory list for self classification of chemical substances — with advisory classifications of more than 30,000 substances. Since the new regulation for classification and labelling (the CLP-regulation) came into force, the regulation is in a transitional phase until 2015 where both regulations still are relevant in certain situations. Therefore both regulations are covered."

European Commission Joint Research Centre (JRC) Institute for Health and Consumer Protection (IHCP)

http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/

A description of the EU computational toxicology and modeling activities from the IHCP website is listed below.

"One of the activities of the Institute is to support the implementation of EU chemicals policy (including the safety assessment of industrial chemicals, chemicals in consumer products, pesticides and biocides) through the development, assessment and application of computational (*in silico)* methods. These methods, sometimes referred to as 'non-testing methods', can be used to reduce our reliance on experimental testing, and in particular animal testing. In practice, these methods are used in Integrated Testing Strategies, along with experimental data generated by alternative (non-animal) tests, such as in vitro tests and high throughput screening (HTS) assays."

"Quantitative Structure-Activity Relationship (QSAR) models can be used to obtain information on the properties and activities of substances from chemical structure alone, and can thus be used to fill data gaps in the safety assessment of chemicals."

"Another method, Physiologically Based Biokinetic (PBBK) Modelling, can be used to extrapolate between in vitro and in vivo exposure conditions, thereby helping to establish the relevance of data generated by in vitro toxicity tests."

The IHCP website provides information and links to a number of documents and several downloadable (Q)SAR tools:

- Background information on non-testing methods
- Information on QSAR reporting formats
- A range of other information sources, including chemical inventories and documents on computational toxicology
- A list of IHCP publications on computational toxicology
- Webpages for free download or access to the following computational tools:
    - JRC QSAR Model Database
    - Toxtree
    - Dart
    - Toxmatch
    - Stat-4-tox

Organization for Economic Cooperation and Development (OECD) OECD (Q)SAR Project

http://www.oecd.org/env/hazard/qsar

The following is a description of the OECD (Q)SAR project from the OECD website:

"To facilitate practical application of (Q)SAR approaches in regulatory contexts by governments and industry and to improve their regulatory acceptance, the OECD (Q)SAR project has developed various outcomes, such as the principles for the validation of (Q)SAR models**,** guidance documents as well as the QSAR Application Toolbox**.** The OECD (Q)SAR Project is carried out with the financial assistance of the European Union."

The (Q)SAR project website includes links to a number of documents and software relating to OECD (Q)SAR activities including the following:

- History (of the project)
- Introduction to (Q)SARs
- Grouping of chemicals
- Validation of (Q)SAR models
- OECD QSAR Toolbox

Rijksinstituut voor Volksgezondheid en Milieu (RIVM) [National Institute for Public Health and the Environment — Netherlands]

http://www.rivm.nl/rvs/risbeoor/Modellen/QSAR.jsp

The RIVM QSAR webpage provides a definition of QSAR and links to a number of QSAR related publications produced by the Institute including:

- Report: *(Q)SARs: gatekeepers against risk on chemicals?*
- A literature review of (Q)SARs for human toxicological endpoints
- A literature review on (Q) SARs for ecotoxicological endpoints
- Report: *The application of structure-activity relationships in human hazard assessment: a first approach*
- Report: *Estimating the PBT-profile*

While much of the information on the webpage is in Dutch, the majority of the publications are available in English.

United States Environmental Protection Agency (US EPA) Office of Pollution Prevention and Toxics (OPPT) Sustainable Futures Initiative (SF)

http://www.epa.gov/oppt/sf/tools/methods.htm

A description of the SF initiative from the OPPT webpage is listed below.

“The goal of the Sustainable Futures Initiative (SF) is to make new chemicals safer, available faster, and at lower cost. It works by giving chemical developers the same risk-screening models that EPA uses to evaluate new chemicals before they enter the market.”

“Using these computer-based models, companies can identify potentially risky chemicals early in the development process and reduce risk by finding safer substitutes and/or processes before submitting them to the EPA. Also, the companies that take training and graduate from Sustainable Futures become eligible for an expedited EPA review of their pre-screened chemicals.”

The computer-based models and tools freely available for download from the OPPT webpage include the following:

- EPI Suite
- ECOSAR
- PBT Profiler
- Oncologic
- Analog Identification Methodology (AIM)
- NonCancer Screening Protocol
- E-FAST
- ChemSteer

United States Food and Drug Administration (US FDA) Center for Drug Evaluation and Research (CDER) Informatics and Computational Safety Analysis Staff (ICSAS)

http://www.fda.gov/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDER/ucm092125.htm

The following description of ICSAS is listed on the US FDA CDER webpage:

"The Informatics and Computational Safety Analysis Staff (ICSAS) is part of CDER's Office of Pharmaceutical Science. ICSAS is an applied regulatory research unit that:

- Develops databases of toxicological and clinical endpoints
- Transforms data, developing rules for quantifying toxicological and clinical effects
- Evaluates structure activity relationship (SAR) and data mining software using ICSAS databases
- Works with software developers to develop toxicology and clinical effects prediction programs through research leveraging partnerships
- Reduces the use of animals in testing by eliminating non-critical laboratory studies
- Facilitates the review process by making better use of accumulated scientific knowledge
- Supplies tools to the pharmaceutical industry to develop better means to identify and eliminate compounds with potentially significant adverse properties early in the drug discovery and development process"

The webpage also includes links to databases and further information on ICSAS activities including:

- Database Projects
- Maximum Recommended Therapeutic Dose (MRTD) Database
- Human Liver Adverse Effects Database
- Genetic Toxicity, Reproductive and Developmental Toxicity, and Carcinogenicity Database
- Salmonella Mutagenicity E-state Descriptors
- Chemical Structure Similarity Searching
- The Computational Toxicology Program and ComTox Consulting Service
- ComTox Regulatory Application of ICSAS MCASE/MC4PC-ES by the Center for Food Safety and Applied Nutrition
- Application of Computational Toxicology to Assess Clinical Adverse Drug Reactions
- Publications

Additional Useful References

The following are some additional useful references on the development, evaluation, and application of (Q)SAR tools. These references are also included in the reference section of this document.

Bassan, A., E. Fioravanzo, M. Pavan, and M. Stocchero. 2011. Applicability of physicochemical data, QSARs and read-across in Threshold of Toxicological Concern assessment. Final report of a study carried out by Soluzioni Informatiche (S-In, Vicenza, Italy) for the European Food Safety Authority (EFSA). http://www.efsa.europa.eu/en/supporting/pub/159e.htm

Cronin, M. 2010. Quantitative structure-activity relationships (QSARs) — applications and methodology. Chapter 10 in *Recent Advances in QSAR Studies: Methods and Applications.* Puzyn, T., J. Leszczynski, and M.T.D. Cronin (eds.). Springer, Heidelberg, Germany, pp. 3–11.

ECHA. 2010. Practical guide 5. How to report (Q)SAR. European Chemicals Agency, Helsinki, Finland. http://echa.europa.eu/documents/10162/13655/pg_report_qsars_en.pdf

Lo Piparo, E., A. Worth, M. Manibusan, C. Yang, B. Schilter, P. Mazzatorta, M.N. Jacobs, H. Steinkellner, and L. Mohimont. 2011. Use of computational tools in the field of food safety. Regulatory Toxicology and Pharmacology 60:354–362.

Mostrag-Szlichtyng, A. and A.P. Worth. 2010. In silico modelling of microbial and human metabolism: a case study with the fungicide carbendazim. JRC Scientific and Technical Report EUR 24523 EN. Publications Office of the European Union, Luxembourg. Available at: http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/publications/

Mostrag-Szlichtyng, A. and A. Worth. 2010. Review of QSAR Models and Software Tools for predicting Biokinetic Properties. JRC Technical Report EUR 24377 EN. Publications Office of the European Union, Luxembourg. Available at: http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/publications/

van Leeuwen, C.J., G.Y. Patlewicz, and A.P. Worth. 2007. Intelligent testing strategies. In: Risk Assessment of Chemicals: An Introduction. 2nd Edition. Van Leeuwen, C.J. and T.G. Vermeire, (eds.).

Vonk, A.J., R. Benigni, M. Hewitt, M. Nendza, H. Segner, D. van de Meent, and M.T.D. Cronin. 2009. The Use of Mechanisms and Modes of Toxic Action in Integrated Testing Strategies: The Report and Recommendations of a Workshop held as part of the European Union OSIRIS Integrated Project. ATLA **37**:557–571.

Worth, A., M. Fuart-Gatnik, S. Lapenna, and R. Serafimova. 2011. Applicability of QSAR analysis in the evaluation of developmental and neurotoxicity effects for the assessment of the toxicological relevance of metabolites and degradates of pesticide active substances for dietary risk assessment. Report produced for EFSA. Available at: http://www.efsa.europa.eu/en/supporting/pub/169e.htm

Worth, A., S. Lapenna, E. Lo Piparo, A. Mostrag-Szlichtyng, and R. Serafimova. 2010. The Applicability of Software Tools for Genotoxicity and Carcinogenicity Prediction: Case Studies relevant to the Assessment of Pesticides. JRC Technical Report EUR 24640 EN. Publications Office of the European Union, Luxembourg. Available at: http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/publications/

Worth, A.P. 2010. The role of QSAR methodology in the regulatory assessment of chemicals, Chapter 13 in *Recent Advances in QSAR Studies: Methods and Applications.* Puzyn, T., J. Leszczynski, and M.T.D. Cronin, (eds.). Springer, Heidelberg, Germany, pp. 367–382.

Worth, A.P. and A.Mostrag-Szlichtyng. 2011. Towards a Common Regulatory Framework for Computational Toxicology: Current Status and Future Perspectives, in New Horizons in Predictive Toxicology: Methods and Applications. Wilson, A.G.E. (ed.). The Royal Society of Chemistry, Cambridge, UK, pp. 38–69.

Appendix II

European Commission Joint Research Centre (JRC) QSAR Model Reporting Format (QMRF) and QSAR Prediction Reporting Format (QPRF)

The European Commission Joint Research Centre (JRC) has developed the QSAR Model Reporting Format (QMRF), a template for summarizing and reporting critical information on (Q)SAR models. The JRC also maintains a freely accessible database of QMRFs. The QSAR Prediction Reporting Format (QPRF), also developed by the JRC, is a template for summarizing and reporting individual substance-specific predictions generated by (Q)SAR models. Both templates have been designed to solicit information about (Q)SAR models and predictions that corresponds to the OECD “Principles for the Validation, for Regulatory Purposes, of (Q)SAR Models”. Also, the JRC has noted that the QPRF and QMRF are complementary and that a QPRF should always be associated with a QMRF.

While not specifically designed for documenting (Q)SAR models and predictions for pesticides, the QMRF and QPRF can be viewed as examples of detailed information templates of the type that may need to be considered when (Q)SAR predictions are to be used as critical sources of data in pesticide assessments.

Tables 12.1 and 12.2 below list the information fields included in the QMRF and the QPRF. The most current versions of the QMRF and the QPRF, guidelines for reviewing the QMRF, a QMRF editor for filling in the QMRF, guidance on creating SDF files associated with QMRFs, and examples of completed QMRFs can be downloaded from the following website:

http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/qsar_tools/QRF

The website above also provides access to the JRC database of QMRFs.

In addition, the European Chemicals Agency has developed guidance for the use of the QMRF and QPRF to report on (Q)SARs and to input information from these templates into IUCLID 5 (ECHA, 2009).

Table 1. JRC QMRF (version 1.2) Information

1. QSAR identifier
    - 1.1 QSAR identifier (title)
    - 1.2 Other related models
    - 1.3 Software coding the model

2. General Information
    - 2.1 Date of QMRF
    - 2.2 QMRF author(s) and contact details
    - 2.3 Date of QMRF update(s)
    - 2.4 QMRF update(s)
    - 2.5 Model developer(s) and contact details
    - 2.6 Date of model development and/or publication
    - 2.7 Reference(s) to main scientific papers and/or software package
    - 2.8 Availability of information about the model
    - 2.9 Availability of another QMRF for exactly the same model

3. Defining the endpoint — OECD Principle 1
    - 3.1 Species
    - 3.2 Endpoint
    - 3.3 Comment on endpoint
    - 3.4 Endpoint units
    - 3.5 Dependent variable
    - 3.6 Experimental protocol
    - 3.7 Endpoint data quality and variability

4. Defining the algorithm — OECD Principle 2
    - 4.1 Type of model
    - 4.2 Explicit algorithm
    - 4.3 Descriptors in the model
    - 4.4 Descriptor selection
    - 4.5 Algorithm and descriptor generation
    - 4.6 Software name and version for descriptor generation
    - 4.7 Descriptors/Chemicals ratio

5. Defining the applicability domain — OECD Principle 3
    - 5.1 Description of the applicability domain of the model
    - 5.2 Method used to assess the applicability domain
    - 5.3 Software name and version for applicability domain assessment
    - 5.4 Limits of applicability

6. Defining goodness-of-fit and robustness — OECD Principle 4
- 6.1 Availability of the training set
- 6.2 Available information for the training sets
- 6.3 Data for each descriptor variable for the training set
- 6.4 Data for the dependent variable (response) for the training set
- 6.5 Other information about the training set
- 6.6 Pre-processing of data before modeling
- 6.7 Statistics for goodness-of-fit
- 6.8 Robustness- Statistics obtained by leave-one-out cross-validation
- 6.9 Robustness- Statistics obtained by leave-main-out cross-validation
- 6.10 Robustness- Statistics obtained by Y-scrambling
- 6.11 Robustness- Statistics obtained by bootstrap
- 6.12 Robustness-Statistics obtained by other methods

7. Defining predictivity — OECD Principle 4
- 7.1 Availability of the external validation set
- 7.2 Available information for the external validation set
- 7.3 Data for each descriptor variable for the external validation set
- 7.4 Data for the dependent variable for the external validation set
- 7.5 Other information about the external validation set
- 7.6 Experimental design of test set
- 7.7 Predictivity- Statistics obtained by external validation
- 7.8 Predictivity- Assessment of the external validation set
- 7.9 Comments on the external validation of the model

8. Providing a mechanistic interpretation — OECD Principle 5
- 8.1 Mechanistic basis of the model
- 8.2 A priori or a posteriori mechanistic interpretation
- 8.3 Other information about the mechanistic interpretation

9. Miscellaneous information
- 9.1 Comments
- 9.2 Bibliography
- 9.3 Supporting information

10. Summary for the JRC Inventory
- 10.1 QMRF number
- 10.2 Publication date
- 10.3 Keywords
- 10.4 Comments

Table 2. EC QPRF (version 1.1) Information Fields

1. Substance
    - 1.1 CAS number
    - 1.2 EC number
    - 1.3 Chemical name
    - 1.4 Structural formula
    - 1.5 Structure code
        - a. SMILES
        - b. InChi
        - c. Other structural representation
        - d. Stereochemical features

2. General information
    - 2.1 Date of QPRF
    - 2.2 QPRF author and contact details

3. Prediction
    - 3.1 Endpoint (OECD Principle 1)
        - a. Endpoint
        - b. Dependent variable
    - 3.2 Algorithm (OECD Principle 2)
        - a. Model or submodel name
        - b. Model version
        - c. Reference to QMRF
        - d. Predicted value (model result)
        - e. Predicted value (comments)
        - f. Input for prediction
        - g. Descriptor values
    - 3.3 Applicability domain (OECD Principle 3)
        - a. Domains
            - i. Descriptor domain
            - ii. Structural fragment domain
            - iii. Mechanistic domain
            - iv. Metabolic domain
        - b. Structural analogues
        - c. Considerations on structural analogues
    - 3.4 The uncertainty of the prediction (OECD Principle 4)
    - 3.5 The chemical and biological mechanisms according to the model underpinning the predicted result (OECD Principle 5)

4. Adequacy (Optional)
    - 4.1 Regulatory purpose
    - 4.2 Approach for regulatory interpretation of the model result
    - 4.3 Outcome
    - 4.4 Conclusion

Appendix III

Listed below is a selection of example case studies of the application of (Q)SAR tools and approaches to the prediction of the ecotoxicity and toxicity of pesticides and other chemicals. These examples were prepared by various groups in the US EPA and Health Canada for applications within their respective programs. These examples were not specifically designed for use in this document, but they were generously contributed in order to illustrate various issues discussed in sections 3, 4, 5, 6, and 7 of the document.

Example No. 1

Case Study:

Use of EcoSAR QSAR Models to Estimate the Acute Toxicity of Organophosphate and Carbamate Pesticide Classes to Fish Species.

The following is a summarized version of a case study of reliability and validation testing of a set of (Q)SAR models for predicting acute toxicity to fish species. This case study involved comparing available high quality empirical data on acute toxicity for an external test set of organophosphate and carbamate pesticides with model predictions for the same pesticides generated by the US EPA's EcoSAR models. The case study was prepared by the US EPA's National Health and Environmental Research Laboratory (NHERL) and it provides a useful example of application of the OECD (Q)SAR validation principles and considerations of the adequacy of (Q)SAR predictions as discussion in Section 5 of this document.

Issue: Currently the US EPA's Office of Pesticide Programs receives acute toxicity data for fish species via the FIFRA registration process. OPP typically does not obtain test data for degradate chemicals of active ingredients and relies on QSAR approaches to determine the potential hazard associated with these substances. In addition, the Office of Water is interested in using QSAR approaches to fill data gaps to meet minimum data requirements in the development of water quality criteria for pesticide active ingredients. These are typically for other chordate and arthropod taxa, but at times data gaps exist for fish and salmonid species. A potential issue with using QSAR models is that most tools were developed to support the TSCA legislation which deals primarily with industrial organic compounds. To this end, an analysis was conducted to determine the reliability and validity of QSAR models for use in estimating the acute toxicity to fish for a set of organophosphate and carbamate pesticides acting via an acetylcholinesterase inhibition mode of action.

Approach: For this exercise Version 1.1 of the US EPA's EcoSAR modeling application was used (http://www.epa.gov/oppt/newchems/tools/21ecosar.htm). Independent assessments of EcoSAR attest to the reliability for predicting toxicity for non-specific modes of action (MOAs), and, with limited success, predicting toxicity for more specific MOAs such as reactive mechanisms (Moore et al., 2003; **Reuschenbach *et al*., 2008). With Version 1.1 of EcoSAR, the US EPA's Office of Pollution Prevention and** Toxics (OPPT), which maintains the tool, began augmenting models with a limited number of pesticide data which have been reviewed and categorized as acceptable for fulfillment of pesticide registration and re-registration guideline requirements as explained under FIFRA Subdivision E, Parts 158.145 and 158.150. This case study was conducted to determine whether these improvements led to reliable QSAR models for use in estimating hazard associated with carbamate and organophosphate pesticides with an acute mode of action of acetylcholinesterase inhibition. To make this determination, a high quality empirical data set was compiled and used to evaluate how well the QSAR models estimates agreed with empirical toxicity test data. Chemicals that were used in the EcoSAR model training set were excluded from the final evaluations of model performance, thereby being a validation of the existing models as they relate to use in estimating toxicity of acetylcholinesterase inhibitors to fish. The OECD QSAR validation principles are used as a template for examining the applicability, reliability, robustness, and predictivity of models.

QSAR Prediction Reporting Format (QPRF):

1. **Substance: Carbamates:** Initially a list of seventeen (17) carbamates and forty-six (47) organophosphate pesticides were examined (See Attachment 1 for chemical name, CAS Registry Number, and SMILES string and Attachment 2 for the logP (log of the octanol water partition coefficient.)

2. **General information:**

    a. **QPRF author and contact details:** This analysis was compiled by Chris Russom, US EPA, ORD, NHEERL, MED, Duluth, MN (Russom.chris@epa.gov and completed using EcoSAR V 1.1 with EcoSAR outputs completed on June 2011.

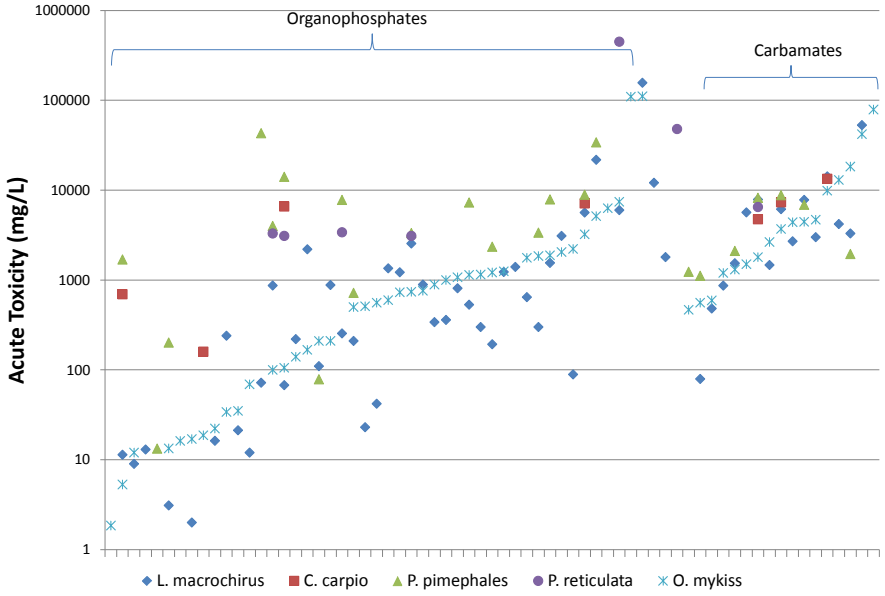
3. **Validation of QSAR model vs. High Quality Empirical Data Set:**

    a. **Empirical Data Set:** A data set was compiled for use in validating model predictions by selecting test results from the ECOTOX database (www.epa.gov/ecotox), and the OPP database of studies submitted for registration of active ingredients (Brian Montague, OPP/EFED, personal

communication). Studies were used which met the following criteria: (a) fish species identified; (b) endpoint is LC50 ; (c) test conducted in freshwater; (d) compound purity ≥90%; (e) test duration of 96 hrs; (f) effect is mortality; (g) studies conducted in a laboratory setting; (h) concentrations are not indefinite values (*i.e*., exclude >, < or ~ values); (i) documented temperature and dissolved oxygen measurements; and (j) adherence to standard test procedures (ASTM 2007; U.S. EPA 1996) *e.g.,* as they relate to organism life stage, water temperature and dissolved oxygen. Duplicate studies were removed by identifying tests where the CAS Registry number, species, age, and LC50 toxicity concentrations were the same. Once all requirements had been met an analysis of outliers was performed. In instances where more than one toxicity value was available for a chemical, species, endpoint, duration combination, the ratio of the maximum and minimum concentration values was calculated. When ratios approached or exceed 10, the original publications were examined, and if errors in data were identified these data were not included in the analysis. If an outlier could not be determined (*e.g.*, only two data points), then all data records were removed if the ratio was greater than 10.

The focus of this evaluation was the use of QSAR models to estimate acute toxicity to fulfill minimum data requirements for use by OW in deriving Agency benchmarks. Rainbow trout or bluegill sunfish tended to be the most sensitive fish species (**Figure 1**) upon examination of the empirical validation data set. Since the ECOSAR model is generic for fish (see 3b below), and a critical minimum data requirement under the Water Quality Criteria Guidelines is a salmonid (USEPA 1985), comparisons of the QSAR model estimates were made using the average test concentration of empirical test results for rainbow trout when possible, and bluegill test data only when rainbow trout data were not available (See Attachment 2).

Figure 1: Average toxicity value by for OP and carbamate acetylcholinesterase inhibitors from ICE model data set



b. **QSAR Model Endpoint (*OECD Principle 1 — A defined biological endpoint*):**

i. *Species/Endpoint*: This analysis compares QSAR estimates to a high quality empirical data set using the EcoSAR QSAR models for acute (96 hr) LC50 to freshwater fish species. The EcoSAR QSAR models provide an estimate for a generic fish, not for a particular species of fish, although frequently used species to develop the QSAR models included bluegill sunfish (*Lepomis macrochirus*), common carp (*Cyprinus carpio*), fathead minnow (*Pimephales promelas*), guppy (*Poecilia reticulata*), rainbow trout (*Oncorhynchus mykiss*), medaka (*Oryzias latipes*), or zebrafish (*Brachydanio rerio*). Model output to the user is presented in concentration units of mg/L.

ii. *Test Protocols*: Acute toxicity test data used in the training set followed either ASTM or OPP standard testing procedures (ASTM, 2007; U.S. EPA 1996). Therefore the data were from several laboratories.

iii. *Dependent variable*: The model calculates log millimole/liter LC50 and EcoSAR software converts the value to mg/L for the report page provided to the user.

3.1 Algorithm (OECD Principle 2)

a. Model, Version, and QSAR sub-model name:

i. *QSAR and model name*: **Table 1** provides details on the linear regression models from EcoSAR Version 1.1 used in this model validation exercise. The supporting files within EcoSAR provide a list of all test data included in the model training set. Attachment 3 identifies the models used for each chemical in the validation set. None of the chemicals in the validation set exceeded model limits. These are chemical class-based models and guidance in the model documentation states:

1. The Carbamate Esters, phenyl chemical class model may be used to estimate toxicity of O-phenyl substituted carbamate esters (*i.e.*, RNC(=O)OPh; where R is anything except two hydrogens; and R is not a primary amine. The phenyl (Ph) can have substitutions on the ring.)

2. The Carbamate Esters, Oxime chemical class model may be used to estimate toxicity of O-oxime substituted carbamate esters (*i.e.*, RNC(=O)ON=C; where R is anything except two hydrogens; and R is not a primary amine.)

3. The Esters, Phosphates chemical class model may be used to estimate toxicity for phosphate esters (*i.e.*, $R^1OP(=O)(R^2)R^3$ OR $R^4SP(=O)(R^2)R^3$ where R^1 can be alkyl carbon, olefinic carbon, acetylenic carbon, aromatic carbon, a carbonyl, phosphorus, sulfur, oxygen or nitrogen; R^2 and R^3 can be anything EXCEPT a hydroxy group (OH), and R^4 can be an alkyl carbon, olefinic carbon, acetylenic carbon, aromatic carbon, or phosphorus.) EcoSAR flagged halogenated tri-alkylphosphate esters as being significantly more toxic than would be estimated by this model. EcoSAR documentation also commented that this model may over or under estimate toxicity for acetylcholinesterase inhibitors, and that proper classification of these substances is ongoing.

4. The Esters, Dithiophosphates chemical class model may be used to estimate toxicity for dithiophosphate esters (*i.e.*, $R^1P(=S)(SR^2)OR^3$; where R^1 can be anything except sulfur; R^2 must be a hydrogen, carbon (alkyl, olefinic, acetylenic, aromatic, or carbonyl), or phosphorus, and R^3 must be a

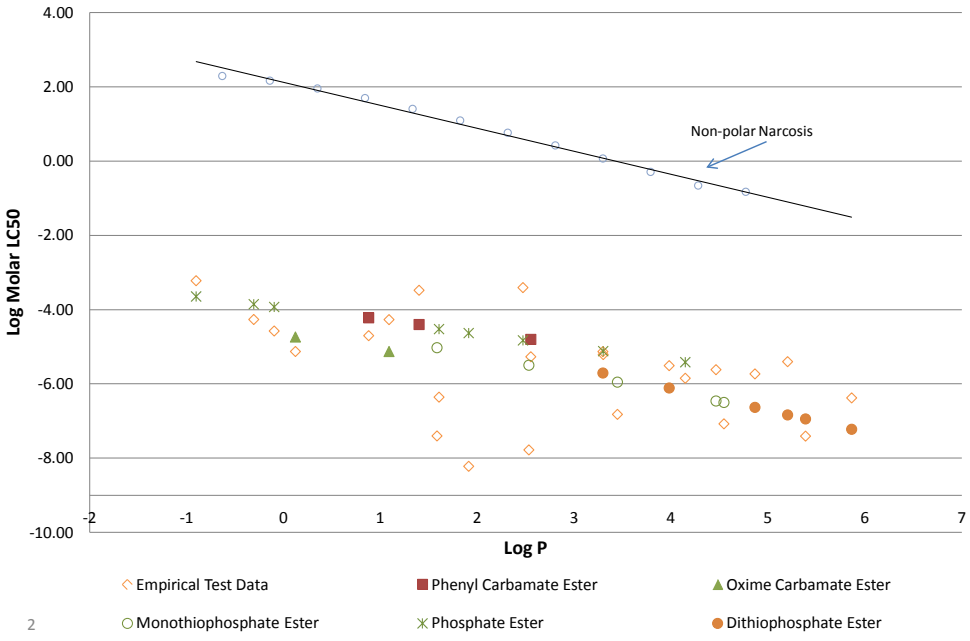
hydrogen, carbon (alkyl, olefinic, acetylenic, aromatic, or carbonyl), nitrogen, oxygen, phosphorus, or sulfur.)

5. The Esters, Monothiophosphates chemical class-based model may be used to estimate toxicity for monothiophosphate esters (*i.e.*, $R^1P(=S)(R^2)OR^3$; where R^1 and R^2 can be anything except sulfur, and R^3 must be a hydrogen, carbon (alkyl, olefinic, acetylenic, aromatic, or carbonyl), nitrogen, oxygen, phosphorus, or sulfur.)

Table 1: EcoSAR V 1.1 Fish Acute LC50 QSAR model information:

Model name	***QSAR***	***Statistics***
Carbamate Esters, phenyl	Log 96-h LC50 (mmol/L) = -0.3478 (logP) - 0.9147	R^2 = 0.1697; N=23
Carbamate Esters, oxime	Log 96-h LC50 (mmol/L) = -0.4048 (logP) – 1.6878	R^2 = 0.4475; N=18
Esters, Phosphate	Log 96-h LC50 (mmol/L) = -0.3504 (logP) - 0.9625	R^2 = 0.2551; N=27
Esters, Dithiophosphate	Log 96-h LC50 (mmol/L) = -0.4981 (logP) – 1.2363	R^2 = 0.2355; N=28
Esters, Monothiophosphate	Log 96-h LC50 (mmol/L) = -0.5902 (logP) – 0.7618	R^2 = 0.1508; N=44

Figure 2: Log molar LC50 for fish (empirical test data and EcoSAR V1.1 QSAR estimates) vs. the log of the octanol / water partition coefficient (logP)



2

ii. *Predicted value*: Attachment 3 provides model estimates as provided in EcoSAR user outputs. As stated earlier, there were no flags on any of the estimates, therefore these chemicals were all estimated within the domain of the model as parameterized. Also, the model estimates the millimolar/liter as the unit, but EcoSAR converts the value to mg/L prior to presenting it in the EcoSAR output. **Figure 2** is a plot of the log molar LC50 (QSAR estimated and empirical) vs. log P. The non-polar narcosis toxicity line is provided as a baseline.

iii. *Input for prediction*: CAS Registry numbers were used as input for EcoSAR model prediction, and SMILES strings were available within EcoSAR for all chemicals except cis-Thiocarboxime. The EcoSAR structures were verified against a second source (Alanwood or ChemID). The SMILES string was written and used as input for EcoSAR model prediction for cis-Thiocarboxime and this structure was verified in a second source.

iv. *Descriptor values*: The logP value from KowWin Version 1.68 was retrieved by EcoSAR as the descriptor variable for the resident QSAR models.

3.2 Applicability domain (OECD principle 3)

a. Domains:

i. *Descriptor variable*: The logP values for the validation data set ranged from 0.123-2.552 for the carbamates and -0.096-5.863 for the organophosphates. Examining training sets used for each QSAR model against the chemicals with estimated values not included in the training set found that all carbamate chemicals except formetanate (logP=0.89 vs. logP for Phenyl carbamate ester training set of 1.52 to 3.06) had logP values within the model training set logP range. For the organophosphates estimated using the Monothiophosphate ester QSAR models, all chemical in the validation set were within the training set logP ranges (*i.e.*, logP range of 2.4 to 4.7) except Dichlofenthion (logP=5.202), Fenchlorfos (logP=4.865), Iodophos (logP=5.387), and Trichloronate (logP=5.863). All chemicals estimated using the Dithiophosphate ester QSAR models had logP values within the training set logP range. Dicrotophos (logP=-0.096) was the only substance estimated using the Phosphate ester QSAR model that was outside the logP range for the training set (-0.74 to 4.85). EcoSAR documentation acknowledges that in general, above a

logP of 5, research has shown that the hydrophobicity of the molecule leads to "no effects at saturation" and due to this, ECOSAR documentation recommends this as an upper limit of acute toxicity (Mayo-Bean et al., 2011.) This limitation is more biological than model domain dependant and does not appear to be an issue with these substances, so for this exercise they were considered to be within the domain of the model.

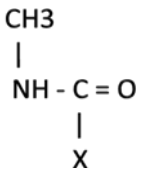
ii. *Structural fragment domain*: The EcoSAR models are chemical class-based, therefore the chemical structure domain as described under Section 3.1 defines the chemical domain of the QSAR. The model selected for each validation chemical agrees with the rules presented in the documentation and a scan of the structures, except the description provided for oxime carbamates appears to not include terminal -ON=C fragment. Both oxime carbamates in the validation set were identified by EcoSAR as oxime carbamates, so it could be the documentation is not capturing the fragments properly.

iii. *Mechanism domain*: All of the validation and training set chemicals listed in Attachments 1-4 are known to inhibit acetylcholinesterase, and documentation included in EcoSAR for Phenyl carbamate esters, oxime carbamate esters confirms this as well. The documentation for the other EcoSAR QSAR models does not mention specific modes of action associated with the chemical class.

b. Structural analogues: Training set structural analogues are presented in Attachments 1.

c. Considerations on structural analogues:

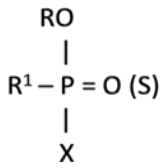
i. Structural requirements for carbamates (Fukuto 1990):



X: Leaving group; typically aryloxy or oxime (*i.e.*, -ONR); when X is a phenyl group, activity increases with 3 substitution on the ring from hydrogen, methyl, ethyl, isopropyl. Having a quarternary ammonium ion in the 3 position on the ring has a

maximum activity. Tert-butyl is less active than the isopropyl form.

ii. Structural requirements for organophosphates (Fukuto 1990):



R: Typically a methyl or ethyl group. Many times RO = $R^{1.}$ When R^1=RO, and R is either a propyl or isopropyl group acetylcholinesterase activity is very low

R^1: Methoxy, ethoxy, ethyl, phenyl, amino, substituted amino, alkylthio

X: Leaving group; typically phenoxy or aromatic group containing hetero atoms, substituted thioalkyl, or substituted alkoxy groups. When X contains a thioether group, these are susceptible to metabolic activation to sulfoxides which are metabolized to sulfones, making these more active AChE inhibitors.

O(S): Direct acting organophosphates have the oxon group; thiophosphates require metabolic activation via mixed function oxidases to the oxon prior to inhibition of AChE activity. Thiophosphates are less reactive and are more stable to hydrolytic degradation than the oxon form.

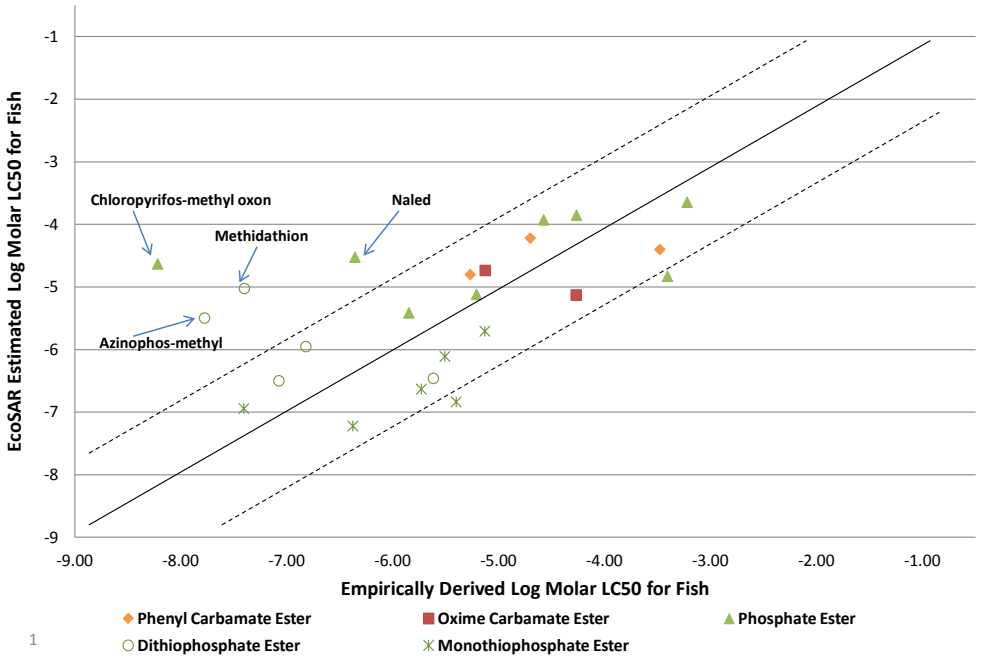
3.3 The uncertainty of the prediction (OECD principle 4)

a. Data sources: The model training set was not generated from one laboratory, therefore the data sets may have variability amongst the data related to differences in genetic stock for fish, consistency in application of test protocols such as static vs flow through exposures; use of analytical procedures to measure test concentration, or reporting of only nominal values, etc.

b. Test Species: As mentioned earlier these models are generic for fish, as can be seen in **Figure 1**, fish species can have a range of sensitivities to these substances.

c. Model Predictivity: The coefficient of correlation between predicted and empirical test data is R=0.485. As presented in **Figure 3**, most chemicals are estimated within an order of magnitude. Predictivity could be improved by the addition of new chemical categories and expanded sub-structure rules for existing chemical categories, as further described under Section 4.0.

Figure 3: Comparison of EcoSAR QSAR Estimates of the Fish Acute LC50 To Empirical Test Data for Organophosphates and Carbamates not in the Model Training Sets



1

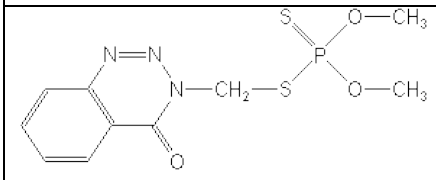
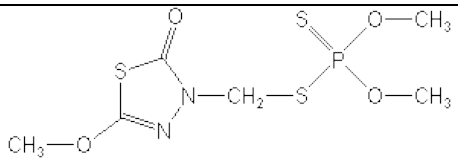
3.4 The chemical and biological mechanisms according to the model underpinning the predicted result (OECD principle 5).

Organophosphate and carbamate pesticides are indirect inhibitors of cholinesterase caused by blocking the site (serine hydroxyl group) on acetylcholinesterase where the neurotransmitter, acetylcholine, would normally attach, thereby blocking the breakdown of acetylcholine (Fukuto 1990; Mileson 1998). Acetylcholine is a major neurotransmitter in the autonomic and cholinergic (CNS) nervous system. Inhibition of acetylcholine binding at the serine site results in a build-up of acetylcholine in synapses, resulting in overstimulation of muscles, glands, and CNS. Toxic effects associated with exposure to acetylcholine inhibiting chemicals include muscle contraction and secretion, cholinergic hyperactivity, lack of muscle coordination, and respiratory depression. A rate limiting step for the thion (*i.e.*, $(RO-)_3P=S$ rather than (RO-

$)_3$P=O) forms of organophosphates is the oxidative desulfuration to create the oxon form which is required for the OP to bind to the AChE enzyme (Fukuto 1990). Documentation within EcoSAR identified the phenyl carbamate ester, the phosphate ester, and the oxime carbamate ester models as associated with an acetylcholinesterase mode of action. Although the monothiophosphate esters and dithiophosphate esters are not explicitly identified as associated with acetylcholinesterase inhibition, the structural requirements as outlined in the EcoSAR documentation align with structure requirements for acetylcholinesterase inhibitors (Fukuto 1990.)

4. Adequacy

Scientific validity of the EcoSAR model used to estimate the toxicity of chemicals listed in Attachment 2, based on the adherence to the OECD principles as outlined above, finds that for carbamates the method appears to be valid, but issues exist for some of the organophosphates. Both carbamate models provided acceptable estimates of toxicity for the five substances in the validation set, with all chemicals estimated within an order of magnitude of the empirical test value. Comparison of the EcoSAR QSAR estimates for the freshwater fish acute LC50 to empirical data sets for the organophosphate chemicals found that most values agreed within an order of magnitude (see **Figure 3**). One significant outlier was chlorpyrifos-methyl oxon, the metabolically active form of chlorpyrifos-methyl. EcoSAR categorized the degradate as a phosphate ester, and using this model, estimated the toxicity to be nearly 4000 times less toxic than found in empirical studies. Methidathion and azinphos-methyl were more than 2 orders of magnitude more toxic than estimated by the EcoSAR dithiophosphate ester model. The QSAR estimate for naled was more than an order of magnitude different than empirical test data. Once again, this substance was estimated using the phosphate ester model. The reliability of these models for use with organophosphates depends on the development of new or refined chemical class to estimate toxicity more effectively.

Azinphos-methyl	Methidathion
	

Log P is the only molecular descriptor used in the EcoSAR models evaluated in this case study. The coefficient of determination for the EcoSAR models evaluated herein (see Table 1 for R^2 values) reflects that factors other than partitioning into the organism are required to completely describe the toxic response. To this end, these QSAR models would improve by the identification and inclusion of toxicologically relevant molecular descriptors in the EcoSAR QSAR models, with linkages to key events within the acetylcholinesterase inhibition adverse outcome pathway.

Risk assessors will need to determine how relevant these model outputs are to the question being asked. For instance, are estimations within an order of magnitude acceptable? What statistical criteria must be met for the internal performance of a model to be considered acceptable (*i.e.,* $R^2 \geq 0.60$)? Similarly, what statistical benchmarks should be used to determine the predictivity of a model? The answer to these questions may depend on whether the risk assessment is for a screening and prioritization, or for deriving a final benchmark value. The purpose of this analysis was to determine whether QSAR estimates could fill data gaps used in deriving Agency benchmark values, and it is recommended that the models undergo further refinement prior to this use.

References:

American Society for Testing and Materials (ASTM). 2007. Standard Guide for Conducting Acute Toxicity Tests on Test Materials with Fishes, Macroinvertebrates, and Amphibians. ASTM Subcommittee E46.01; E729-96(2007). 22 pp. American Society for Testing and Materials, West Conshohocken, PA.

Fukuto, T.R. 1990. Mechanism of action of organophosphorus and carbamate insecticides. Environ. Health Perspect. 87:245-254.

Mayo-Bean, K., J.V. Nabholz, R. Clements, M. Zeeman,, T. Henry, D. Rodier, K. Moran, B. Meylan, and P. Ranslow. 2011. Methodology Document for the ECOlogical Structure-

Activity Relationship Model (ECOSAR) Class Program: Estimating Toxicity of Industrial Chemicals to Aquatic Organisms using ECOSAR Class Program (Version 1.1). U.S. Environmental Protection Agency, Office of Chemical Safety and Pollution Prevention, Office of Pollution Prevention and Toxics, Washington DC.

Mileson, B.E., J.E. Chambers, W.L. Chen, W. Dettbarn, M. Ehrich, A.T. Eldefrawi, D.W. Gaylor, K. Hamernick, E. Hodgson, A.G. Karczmar, S. Padilla, C.N. Pope, R.J. Richardson, D.R. Saunders, L.P. Sheets, L.G. Sultatos, and K.B. Wallace. 1998. Common mechanism of toxicity: A case study of organophosphorus pesticides. Tox. Sci. 41:8-20.

Moore, D.R.J., R.L. Breton, and D.B. MacDonald., 2003. A comparison of model performance for six quantitative structure–activity relationship packages that predict acute toxicity to fish. Environ. Toxicol. Chem. 22, 1799–1809.

Reuschenbach P, Silvani M, Dammann M, Warnecke D, Knacker T. 2008. ECOSAR model performance with a large test set of industrial chemicals. Chemosphere 71:1986–1995.

U.S. EPA. 1996. Ecological Effects Test Guidelines: OPPTS 850.1075. Fish acute toxicity test, freshwater and marine. U.S. EPA, Office of Prevention, Pesticides, and Toxic Substances, EPA 712-C-96-118. Public Draft; 13 pp.

U.S. EPA. 1985. Guidelines for deriving numerical national water quality criteria for the protection of aquatic organisms and their uses. U.S. Environmental Protection Agency, Office of Research and Development, Environmental Research Laboratories. PB85-227049.

Attachment 1: Carbamate and organophosphate pesticides structures and CAS Registry Numbers.

Greyed cells are substances that were not included in the final evaluation of model performance because the chemical was part of the EcoSAR model training set.

Pesticide	CAS Registry Number	Empirical Fish 96 hr LC50 data	EcoSAR Training Set	Chemical Class: Specific	Structure: SMILES
Carbamates (N=17 total, with N=12 included in the EcoSAR model training set, N=5 used in model validation.)					
Bendiocarb	22781-23-3	Yes	No	Carbamate	CNC(=O)Oc1cccc2OC(C)(C)Oc12
cis-Thiocarboxime	29118-87-4	Yes	No	Oxime carbamate	CNC(=O)ON=C(C)SCCC(#N)
endo-3-Chloro-exo-6-cyano-2-norbornanone, o-(Methylcarbamoyl) oxime	15271-41-7	Yes	No	Oxime carbamate	CNC(=O)ON=C1C(C2)C(C#N)CC2C1Cl
Formetanate hydrochloride	23422-53-9	Yes	No	Formamidine	CNC(=O)Oc1cccc(N=CN(C)C)c1
Pirimicarb	23103-98-2	Yes	No	Dimethylcarbamate	CN(C)C(=O)Oc1nc(nc(C)c1C)N(C)C
3,4,5-Trimethylphenyl methylcarbamate	2686-99-9	Yes	Yes	Phenyl methylcarbamate	CNC(=O)Oc1cc(C)c(C)c(C)c1
Aldicarb	116-06-3	Yes	Yes	Oxime carbamate	CNC(=O)ON=CC(C)(C)SC
Aldoxycarb	1646-88-4	Yes	Yes	Oxime carbamate	CNC(=O)ON=CC(C)(C)S(C)(=O)=O
Aminocarb	2032-59-9	Yes	Yes	Phenyl methylcarbamate	CNC(=O)Oc1ccc(N(C)C)c(C)c1
Carbaryl	63-25-2	Yes	Yes	Carbamate	CNC(=O)Oc1cccc2ccccc12
Carbofuran	1563-66-2	Yes	Yes	Benzofuranyl methylcarbamate	CNC(=O)Oc1cccc2CC(C)(C)Oc12
Methiocarb	2032-65-7	Yes	Yes	Phenyl methylcarbamate	CNC(=O)Oc1cc(C)c(SC)c(C)c1
Methomyl	16752-77-5	Yes	Yes	Oxime carbamate	CNC(=O)ON=C(C)SC

Attachment 1: Carbamate and organophosphate pesticides structures and CAS Registry Numbers.

Greyed cells are substances that were not included in the final evaluation of model performance because the chemical was part of the EcoSAR model training set.

Pesticide	CAS Registry Number	Empirical Fish 96 hr LC50 data	EcoSAR Training Set	Chemical Class: Specific	Structure: SMILES
Mexacarbate	315-18-4	Yes	Yes	Phenyl methylcarbamate	CNC(=O)Oc1cc(C)c(N(C)C)c(C)c1
Oxamyl	23135-22-0	Yes	Yes	Oxime carbamate	CNC(=O)ON=C(SC)C(=O)N(C)C
Propoxur	114-26-1	Yes	Yes	Phenyl methylcarbamate	CNC(=O)Oc1ccccc1OC(C)C
Thiodicarb	59669-26-0	Yes	Yes	Oxime carbamate	CSC(C)=NOC(=O)N(C)SN(C)C(=O)ON=C(C)SC
Organophosphates (N=47 total, with N=28 included in the EcoSAR model training set, N=19 used in model validation.)					
Acephate	30560-19-1	Yes	No	Phosphoramidothioate	COP(=O)(NC(C)=O)SC
Azinphos-methyl	86-50-0	Yes	No	Benzotriazine organothiophosphate	S=P(OC)(OC)SCN1N=Nc2ccccc2C1(=O)
Carbophenothion	786-19-6	Yes	Yes	Phenyl organothiophosphate	CCOP(=S)(OCC)SCSc1ccc(Cl)cc1
Chlorfenvinphos	470-90-6	Yes	No	Organophosphate	CCOP(=O)(OCC)OC(=CCl)c1ccc(Cl)cc1Cl
Chlorpyrifos	2921-88-2	Yes	Yes	Pyridine organothiophosphate	CCOP(=S)(OCC)Oc1nc(Cl)c(Cl)cc1Cl
Chlorpyrifos-methyl oxon	5598-52-7	Yes	No	Degradate — Oxon	O=P(OC)(OC)Oc1nc(Cl)cc(Cl)c1Cl
Crufomate	299-86-5	Yes	No	Phosphoramidate	CNP(=O)(OC)Oc1ccc(cc1Cl)C(C)(C)C
Dichlofenthion	97-17-6	Yes	No	Phenyl organothiophosphate	CCOP(=S)(OCC)Oc1ccc(Cl)cc1Cl
Dicrotophos	141-66-2	Yes	No	Organophosphate	COP(=O)(OC)OC(C)=CC(=O)N(C)C
Dioxathion	78-34-2	Yes	No	Heterocyclic organothiophosphate	CCOP(=S)(OCC)SC1OCCOC1SP(=S)(OCC)OCC
Fenchlorphos	299-84-3	Yes	No	Phenyl organothiophosphate	COP(=S)(OC)Oc1cc(Cl)c(Cl)cc1Cl
Fenitrothion	122-14-5	Yes	No	Phenyl organothiophosphate	COP(=S)(OC)Oc1ccc(N(=O)=O)c(C)c1
Fosthiazate	98886-44-3	Yes	No	Heterocyclic organothiophosphate	CCOP(=O)(SC(C)CC)N1CCSC1=O

Attachment 1: Carbamate and organophosphate pesticides structures and CAS Registry Numbers.

Greyed cells are substances that were not included in the final evaluation of model performance because the chemical was part of the EcoSAR model training set.

Pesticide	CAS Registry Number	Empirical Fish 96 hr LC50 data	EcoSAR Training Set	Chemical Class: Specific	Structure: SMILES
Iodofenphos	18181-70-9	Yes	No	Phenyl organothiophosphate	COP(=S)(OC)Oc1cc(Cl)c(I)cc1Cl
Methidathion	950-37-8	Yes	No	Thiadiazole organothiophosphate	COc1nn(CSP(=S)(OC)OC)c(=O)s1
Methyl carbophenothion	953-17-3	Yes	No	Phenyl organothiophosphate	COP(=S)(OC)SCSc1ccc(Cl)cc1
Monocrotophos	6923-22-4	Yes	No	Organophosphate	CNC(=O)C=C(C)OP(=O)(OC)OC
Naled	300-76-5	Yes	No	Organophosphate	COP(=O)(OC)OC(Br)C(Cl)(Cl)Br
SD-7438	2782-70-9	Yes	No	Organothiophosphate	COP(=S)(OC)SCc1ccc(SP(=S)(OC)OC)cc1
Sulfotep	3689-24-5	Yes	No	Aliphatic organothiophosphate	CCOP(=S)(OCC)OP(=S)(OCC)OCC
Trichloronate	327-98-0	Yes	No	Phenyl ethylphosphonothioate	CCOP(=S)(CC)Oc1cc(Cl)c(Cl)cc1Cl
Chlorpyrifos-methyl	5598-13-0	Yes	Yes	Pyridine organothiophosphate	COP(=S)(OC)Oc1nc(Cl)c(Cl)cc1Cl
Coumaphos	56-72-4	Yes	Yes	Heterocyclic organothiophosphate	S=P(OCC)(OCC)Oc1ccc2C(C)=C(Cl)C(=O)Oc2c1
Demeton	8065-48-3	Yes	Yes	Aliphatic organothiophosphate	S=P(OCC)(OCC)OCCSCC + O=P(OCC)(OCC)OCCSCC
Diazinon	333-41-5	Yes	Yes	Pyrimidine organothiophosphate	CCOP(=S)(OCC)Oc1cc(C)nc(n1)C(C)C
Dichlorvos/DDVP	62-73-7	Yes	Yes	Organophosphate	COP(=O)(OC)OC=C(Cl)Cl
Dimethoate	60-51-5	Yes	Yes	Aliphatic amide organothiophosphate	CNC(=O)CSP(=S)(OC)OC
Disulfoton	298-04-4	Yes	Yes	Aliphatic organothiophosphate	CCOP(=S)(OCC)SCCSCC
EPN	2104-64-5	Yes	Yes	Phenyl phenylphosphonothioate	CCOP(=S)(Oc1ccc(cc1)N(=O)=O)c2ccccc2
Ethion	563-12-2	Yes	Yes	Aliphatic organothiophosphate	CCOP(=S)(OCC)SCSP(=S)(OCC)OCC

Attachment 1: Carbamate and organophosphate pesticides structures and CAS Registry Numbers.

Greyed cells are substances that were not included in the final evaluation of model performance because the chemical was part of the EcoSAR model training set.

Pesticide	CAS Registry Number	Empirical Fish 96 hr LC50 data	EcoSAR Training Set	Chemical Class: Specific	Structure: SMILES
Ethoprophos	13194-48-4	Yes	Yes	Aliphatic organothiophosphate	CCCSP(=O)(OCC)SCCC
Fensulfothion	115-90-2	Yes	Yes	Phenyl organothiophosphate	CCOP(=S)(OCC)Oc1ccc(cc1)S(C)=O
Fenthion	55-38-9	Yes	Yes	Phenyl organothiophosphate	COP(=S)(OC)Oc1ccc(SC)c(C)c1
Fonofos	944-22-9	Yes	Yes	Phenyl ethylphosphonothioate	CCOP(=S)(CC)Sc1ccccc1
Isazofos	42509-80-8	Yes	Yes	Triazole organothiophosphate	CCOP(=S)(OCC)Oc1nc(Cl)n(n1)C(C)C
Isofenphos	25311-71-1	Yes	Yes	Phosphoramidothioate	CCOP(=S)(NC(C)C)Oc1ccccc1C(=O)OC(C)C
Malathion	121-75-5	Yes	Yes	Aliphatic organothiophosphate	CCOC(=O)CC(SP(=S)(OC)OC)C(=O)OCC
Oxydemeton-methyl	301-12-2	Yes	Yes	Aliphatic organothiophosphate	CCS(=O)CCSP(=O)(OC)OC
Parathion	56-38-2	Yes	Yes	Phenyl organothiophosphate	CCOP(=S)(OCC)Oc1ccc(cc1)N(=O)=O
Parathion-methyl	298-00-0	Yes	Yes	Phenyl organothiophosphate	S=P(OC)(OC)O-c(ccc1N(=O)=O)cc1
Phorate	298-02-2	Yes	Yes	Aliphatic organothiophosphate	CCOP(=S)(OCC)SCSCC
Phosmet	732-11-6	Yes	Yes	Isoindole organothiophosphate	COP(=S)(OC)SCN2C(=O)c1ccccc1C2=O
Profenofos	41198-08-7	Yes	Yes	Phenyl organothiophosphate	CCCSP(=O)(OCC)Oc1ccc(Br)cc1Cl
Propetamphos	31218-83-4	Yes	Yes	Phosphoramidothioate	CCNP(=S)(OC)OC(C)=CC(=O)OC(C)C
Tebupirimfos	96182-53-5	Yes	Yes	Pyrimidine organothiophosphate	CCOP(=S)(OC(C)C)Oc1cnc(nc1)C(C)(C)C
Temephos	3383-96-8	Yes	Yes	Phenyl organothiophosphate	COP(=S)(OC)Oc2ccc(Sc1ccc(OP(=S)(OC)OC)cc1)cc2
Trichlorfon	52-68-6	Yes	Yes	Phosphonate	COP(=O)(OC)C(O)C(Cl)(Cl)Cl

Attachment 2: Empirical data set of acceptable quality data for acetylcholinesterase inhibitors for use in comparison to QSAR model estimates. The ECOSAR Version 1.1 tool included generic freshwater fish QSAR models. For this exercise, fish model estimates were compared to either rainbow trout (*Oncorhynchus mykiss*) or bluegill (*Lepomis macrochirus*) empirical data.

CAS Registry Number	Pesticide	Species	Minimum Toxicity (ug/L)	Maximum Toxicity (ug/L)	Average Concentration (ug/L)	Number of Toxicity Values
2686999	3,4,5-Trimethylphenyl methylcarbamate	*Oncorhynchus mykiss*	4700	4700	4700	1
30560191	Acephate	*Oncorhynchus mykiss*	110000	110000	110000	1
116063	Aldicarb	*Oncorhynchus mykiss*	560	560	560	1
1646884	Aldoxycarb	*Oncorhynchus mykiss*	42000	42000	42000	1
2032599	Aminocarb	*Oncorhynchus mykiss*	12000	25000	18314.29	7
86500	Azinphos-methyl	*Oncorhynchus mykiss*	4.3	6.3	5.3	2
22781233	Bendiocarb	*Oncorhynchus mykiss*	1200	1200	1200	1
741582	bensulide	*Oncorhynchus mykiss*	720	1400	1073.33	3
63252	Carbaryl	*Oncorhynchus mykiss*	780	3500	1796.63	16
1563662	Carbofuran	*Oncorhynchus mykiss*	380	600	466.67	3
786196	Carbophenothion	*Lepomis macrochirus*	13	13	13	1
470906	Chlorfenvinphos	*Oncorhynchus mykiss*	510	510	510	1
2921882	Chlorpyrifos	*Oncorhynchus mykiss*	7.1	25	13.37	3
5598130	Chlorpyrifos-methyl	*Oncorhynchus mykiss*	120	301	210.5	2
5598527	Chlorpyrifos-methyl oxon	*Oncorhynchus mykiss*	1.7	2	1.85	2
29118874	cis-Thiocarboxime	*Oncorhynchus mykiss*	1500	1500	1500	1
56724	Coumaphos	*Oncorhynchus mykiss*	890	890	890	1
299865	Crufomate	*Lepomis macrochirus*	1800	1800	1800	1

Attachment 2: Empirical data set of acceptable quality data for acetylcholinesterase inhibitors for use in comparison to QSAR model estimates. The ECOSAR Version 1.1 tool included generic freshwater fish QSAR models. For this exercise, fish model estimates were compared to either rainbow trout (*Oncorhynchus mykiss*) or bluegill (*Lepomis macrochirus*) empirical data.

CAS Registry Number	Pesticide	Species	Minimum Toxicity (ug/L)	Maximum Toxicity (ug/L)	Average Concentration (ug/L)	Number of Toxicity Values
8065483	Demeton	*Oncorhynchus mykiss*	520	600	560	2
333415	Diazinon	*Lepomis macrochirus*	136	460	254.67	3
97176	Dichlofenthion	*Oncorhynchus mykiss*	1250	1250	1250	1
62737	Dichlorvos/DDVP	*Oncorhynchus mykiss*	100	100	100	1
141662	Dicrotophos	*Oncorhynchus mykiss*	6300	6300	6300	1
60515	Dimethoate	*Oncorhynchus mykiss*	6200	8600	7433.33	3
78342	Dioxathion	*Oncorhynchus mykiss*	69	69	69	1
298044	Disulfoton	*Oncorhynchus mykiss*	1850	1850	1850	1
15271417	endo-3-Chloro-exo-6-cyano-2-norbornanone, o-(Methylcarbamoyl) oxime	*Oncorhynchus mykiss*	13000	13000	13000	1
2104645	EPN	*Oncorhynchus mykiss*	210	210	210	1
563122	Ethion	*Oncorhynchus mykiss*	500	500	500	1
13194484	Ethoprophos	*Oncorhynchus mykiss*	1150	1150	1150	1
299843	Fenchlorphos	*Oncorhynchus mykiss*	550	645	597.5	2
122145	Fenitrothion	*Oncorhynchus mykiss*	1000	2700	2050	7
115902	Fensulfothion	*Lepomis macrochirus*	72	72	72	1
55389	Fenthion	*Oncorhynchus mykiss*	550	840	740	3
944229	Fonofos	*Oncorhynchus mykiss*	20	50	35	2
23422539	Formetanate hydrochloride	*Oncorhynchus mykiss*	4400	4400	4400	1
98886443	Fosthiazate	*Oncorhynchus mykiss*	111000	111000	111000	1

Attachment 2: Empirical data set of acceptable quality data for acetylcholinesterase inhibitors for use in comparison to QSAR model estimates. The ECOSAR Version 1.1 tool included generic freshwater fish QSAR models. For this exercise, fish model estimates were compared to either rainbow trout (*Oncorhynchus mykiss*) or bluegill (*Lepomis macrochirus*) empirical data.

CAS Registry Number	Pesticide	Species	Minimum Toxicity (ug/L)	Maximum Toxicity (ug/L)	Average Concentration (ug/L)	Number of Toxicity Values
18181709	Iodofenphos	*Oncorhynchus mykiss*	16.2	16.2	16.2	1
42509808	Isazofos	*Oncorhynchus mykiss*	18.7	18.7	18.7	1
25311711	Isofenphos	*Lepomis macrochirus*	1400	1400	1400	1
121755	Malathion	*Oncorhynchus mykiss*	30	200	105.6	5
950378	Methidathion	*Oncorhynchus mykiss*	10	14	12	2
2032657	Methiocarb	*Oncorhynchus mykiss*	436	750	593	2
16752775	Methomyl	*Oncorhynchus mykiss*	1050	1600	1308.33	6
953173	Methyl carbophenothion	*Oncorhynchus mykiss*	760	760	760	1
315184	Mexacarbate	*Oncorhynchus mykiss*	4450	15000	9887.5	4
6923224	Monocrotophos	*Lepomis macrochirus*	12100	12100	12100	1
300765	Naled	*Oncorhynchus mykiss*	132	195	167.33	3
23135220	Oxamyl	*Oncorhynchus mykiss*	4200	4700	4450	2
301122	Oxydemeton-methyl	*Oncorhynchus mykiss*	730	730	730	1
56382	Parathion	*Oncorhynchus mykiss*	864	1430	1214.67	3
298000	Parathion-methyl	*Oncorhynchus mykiss*	2750	3700	3225	2
298022	Phorate	*Oncorhynchus mykiss*	13	21	17	2
732116	Phosmet	*Oncorhynchus mykiss*	105	4700	1142.08	12
23103982	Pirimicarb	*Oncorhynchus mykiss*	79000	79000	79000	1
41198087	Profenofos	*Oncorhynchus mykiss*	21	23.5	22.25	2
31218834	Propetamphos	*Oncorhynchus mykiss*	940	2600	1770	2

Attachment 2: Empirical data set of acceptable quality data for acetylcholinesterase inhibitors for use in comparison to QSAR model estimates. The ECOSAR Version 1.1 tool included generic freshwater fish QSAR models. For this exercise, fish model estimates were compared to either rainbow trout (*Oncorhynchus mykiss*) or bluegill (*Lepomis macrochirus*) empirical data.

CAS Registry Number	Pesticide	Species	Minimum Toxicity (ug/L)	Maximum Toxicity (ug/L)	Average Concentration (ug/L)	Number of Toxicity Values
114261	Propoxur	*Oncorhynchus mykiss*	3700	3700	3700	1
2782709	SD-7438	*Oncorhynchus mykiss*	34	34	34	1
3689245	Sulfotep	*Oncorhynchus mykiss*	1000	1000	1000	1
96182535	Tebupirimfos	*Oncorhynchus mykiss*	2220	2220	2220	1
3383968	Temephos	*Oncorhynchus mykiss*	3490	6800	5145	2
59669260	Thiodicarb	*Oncorhynchus mykiss*	2650	2650	2650	1
52686	Trichlorfon	*Oncorhynchus mykiss*	370	8800	1882.57	28
327980	Trichloronate	*Oncorhynchus mykiss*	140	140	140	1

Attachment 3: Comparison of ECOSAR model (Version 1.1) estimates for fish to the average LC50 value (mg/L) from acceptable empirical data collection (See Supplemental Information Table 2 for details related to empirical test data.) ECOSAR Class Model is the QSAR chemical class-based equation used to estimate toxicity for this exercise.

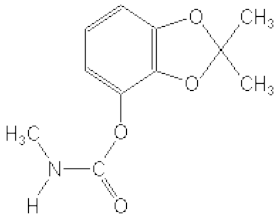
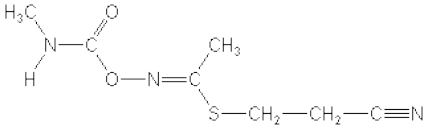
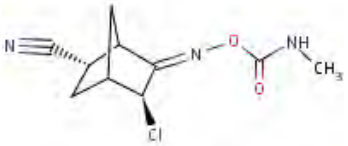
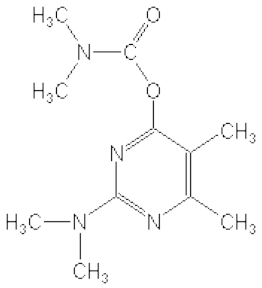
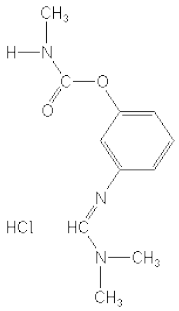
ECOSAR Version 1.1 Model Estimates and Average Empirical Test Data — Fish

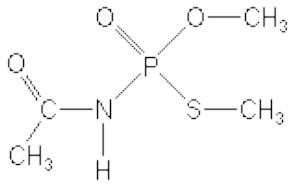
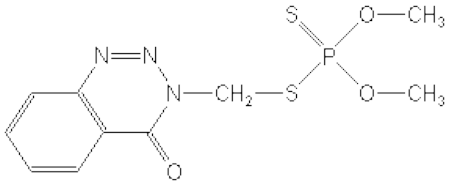
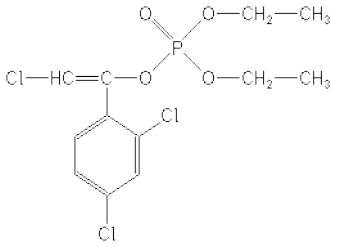
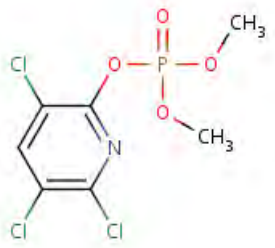
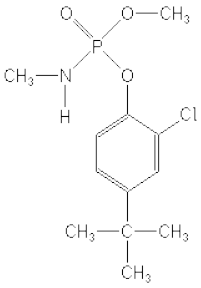
CAS Registry Number	Pesticide	General Chemical Class	Version 1.1 — ECOSAR Class Model	Version 1.1 — KowWin LogP	Version 1.1 ECOSAR Fish-96 hr LC50 Estimate (mg/L)	Average of Acceptable Empirical Fish 96 hr LC50 values (mg/L)	Ratio of ECOSAR Estimated value by Empirical Test data
Carbamate Validation Data Set: Chemical was NOT part of EcoSAR Model Training Set							
22781233	Bendiocarb	Carbamate	Carbamate esters, phenyl	2.552	3.521	1.20	2.93
29118874	cis-Thiocarboxime	Carbamate	Oxime Carbamate Ester	0.123	3.684	1.50	2.46
15271417	endo-3-Chloro-exo-6-cyano-2-norbornanone, o-(Methylcarbamoyl) oxime	Carbamate	Oxime Carbamate Ester	1.089	1.797	13.00	0.14
23422539	Formetanate hydrochloride	Carbamate	Carbamate esters, phenyl	0.879	13.318	4.40	3.03
23103982	Pirimicarb	Carbamate	Carbamate esters, phenyl	1.399	9.456	79.00	0.12
Organophosphate Validation Data Set: Chemical was NOT part of EcoSAR Model Training Set							
30560191	Acephate	Organophosphate	Esters (phosphate)	-0.902	41.359	110.00	0.38
86500	Azinphos-methyl	Organophosphate	Esters, Dithiophosphates	2.532	1.01	0.01	190.57
470906	Chlorfenvinphos	Organophosphate	Esters (phosphate)	4.146	1.382	0.51	2.71
5598527	Chlorpyrifos-methyl oxon	Organophosphate	Esters (phosphate)	1.911	7.151	0.00	3865.41
299865	Crufomate	Organophosphate	Esters (phosphate)	3.299	2.22	1.80	1.23
97176	Dichlofenthion	Organophosphate	Esters, Monothiophosphates	5.202	0.046	1.25	0.04
141662	Dicrotophos	Organophosphate	Esters (phosphate)	-0.096	27.934	6.30	4.43
78342	Dioxathion	Organophosphate	Esters, Dithiophosphates	3.446	0.509	0.07	7.38
299843	Fenchlorphos	Organophosphate	Esters, Monothiophosphates	4.865	0.075	0.60	0.13
122145	Fenitrothion	Organophosphate	Esters, Monothiophosphates	3.296	0.544	2.05	0.27
98886443	Fosthiazate	Organophosphate	Esters (phosphate)	2.471	4.207	111.00	0.04

Attachment 3: Comparison of ECOSAR model (Version 1.1) estimates for fish to the average LC50 value (mg/L) from acceptable empirical data collection (See Supplemental Information Table 2 for details related to empirical test data.) ECOSAR Class Model is the QSAR chemical class-based equation used to estimate toxicity for this exercise.

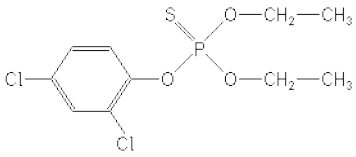
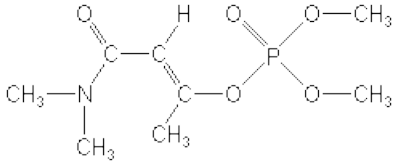
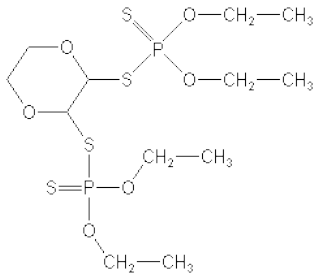
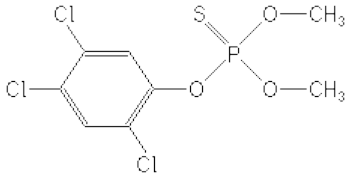
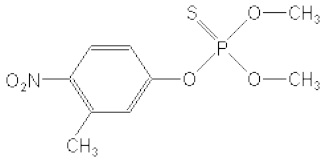
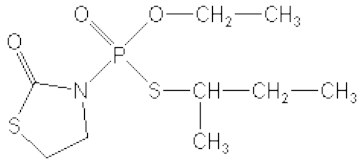
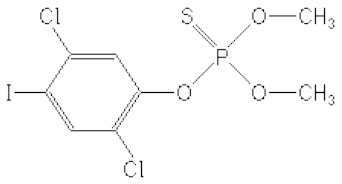
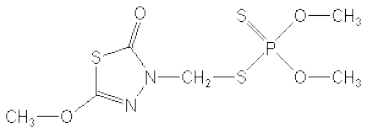
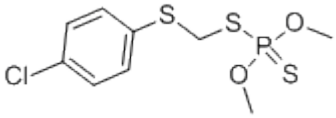
ECOSAR Version 1.1 Model Estimates and Average Empirical Test Data — Fish

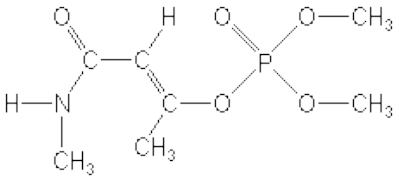
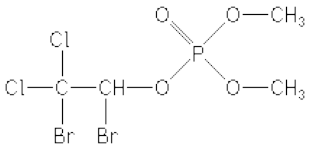
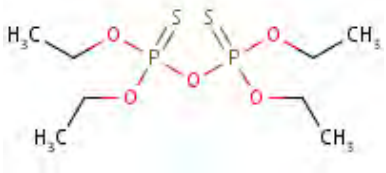
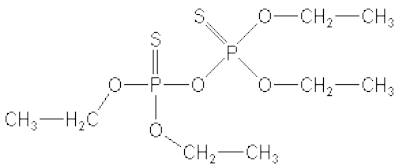
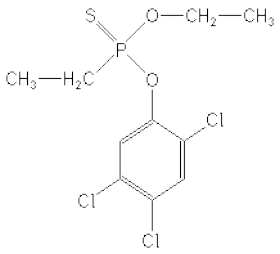
CAS Registry Number	Pesticide	General Chemical Class	Version 1.1 — ECOSAR Class Model	Version 1.1 — KowWin LogP	Version 1.1 ECOSAR Fish-96 hr LC50 Estimate (mg/L)	Average of Acceptable Empirical Fish 96 hr LC50 values (mg/L)	Ratio of ECOSAR Estimated value by Empirical Test data
18181709	Iodofenphos	Organophosphate	Esters, Monothiophosphates	5.387	0.047	0.02	2.90
950378	Methidathion	Organophosphate	Esters, Dithiophosphates	1.584	2.851	0.01	237.58
953173	Methyl carbophenothion	Organophosphate	Esters, Dithiophosphates	4.463	0.109	0.76	0.14
6923224	Monocrotophos	Organophosphate	Esters (phosphate)	-0.307	31.165	12.10	2.58
300765	Naled	Organophosphate	Esters (phosphate)	1.605	11.374 #	0.17	67.97
2782709	SD-7438	Organophosphate	Esters, Dithiophosphates	4.545	0.128	0.03	3.76
3689245	Sulfotep	Organophosphate	Esters, Monothiophosphates	3.98	0.25	1.00	0.25
327980	Trichloronate	Organophosphate	Esters, Monothiophosphates	5.863	0.02	0.14	0.14

Attachment 4: Structures for chemicals in the Validation set (not in the EcoSAR training set for the models and high quality empirical test data were available to use in model validation). Structures are from Alan Wood website (www.alanwood.net) with the exception of *endo-3-Chloro-exo-6-cyano-2-norbornanone, o-(Methylcarbamoyl) oxime , Chlorpyrifos-methyl oxon, and* SD-7438 which are from ChemIDPlus (http://chem.sis.nlm.nih.gov/chemidplus/).		
Carbamates that were not part of the EcoSAR Training Set	*Bendiocarb (CAS: 22781-23-3)*	*cis-Thiocarboxime (CAS: 29118-87-4)*
		
endo-3-Chloro-exo-6-cyano-2-norbornanone, o-(Methylcarbamoyl) oxime (CAS: 15271-41-7)	*Pirimicarb (CAS: 23103-98-2)*	*Formetanate hydrochloride (CAS: 23422-53-9)*
		

Attachment 4: Structures for chemicals in the Validation set (not in the EcoSAR training set for the models and high quality empirical test data were available to use in model validation). Structures are from Alan Wood website (www.alanwood.net) with the exception of *endo-3-Chloro-exo-6-cyano-2-norbornanone, o-(Methylcarbamoyl) oxime , Chlorpyrifos-methyl oxon, and* SD-7438 which are from ChemIDPlus (http://chem.sis.nlm.nih.gov/chemidplus/).		
Organophosphates that were not part of the EcoSAR Training Set	*Acephate (CAS: 30560-19-1)*	*Azinphos-methyl (CAS: 86-50-0)*
		
Chlorfenvinphos *(CAS: 470-90-6)*	*Chlorpyrifos-methyl oxon* *(CAS: 5598-52-7)*	*Crufomate (CAS: 299-86-5)*
		

Attachment 4: Structures for chemicals in the Validation set (not in the EcoSAR training set for the models and high quality empirical test data were available to use in model validation). Structures are from Alan Wood website (www.alanwood.net) with the exception of *endo-3-Chloro-exo-6-cyano-2-norbornanone, o-(Methylcarbamoyl) oxime , Chlorpyrifos-methyl oxon, and* SD-7438 which are from ChemIDPlus (http://chem.sis.nlm.nih.gov/chemidplus/).

Dichlofenthion (CAS: 97-17-6	*Dicrotophos (CAS: 141-66-2)*	*Dioxathion (CAS: 78-34-2)*
		
Fenchlorphos (CAS: 299-84-3)	*Fenitrothion (CAS: 122-14-5)*	*Fosthiazate (CAS: 98886-44-3)*
		
Iodofenphos (CAS: 18181-70-9)	*Methidathion (CAS: 950-37-8)*	*Methyl carbophenothion* *(CAS: 953-17-3)*
		

Attachment 4: Structures for chemicals in the Validation set (not in the EcoSAR training set for the models and high quality empirical test data were available to use in model validation). Structures are from Alan Wood website (www.alanwood.net) with the exception of *endo-3-Chloro-exo-6-cyano-2-norbornanone, o-(Methylcarbamoyl) oxime , Chlorpyrifos-methyl oxon, and* SD-7438 which are from ChemIDPlus (http://chem.sis.nlm.nih.gov/chemidplus/).		
Monocrotophos (CAS: 6923-22-4)	*Naled (CAS: 300-76-5)*	*SD-7438 (CAS: 2782-70-9)*
		
Sulfotep (CAS: 3689-24-5)	*Trichloronate (CAS: 327-98-0)*	
		

Example No. 2

The following case study was prepared by the Office of Pesticide Programs (OPP), US EPA.

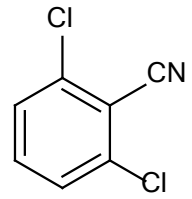
Case Study:

Use of Analog Data to Determine Whether Additional Data Should be Required for a Pesticide Degradate

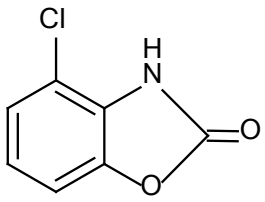
In pesticide risk assessment there is often an abundance of toxicity data on the parent active ingredient and very little, if any, data on pesticide metabolites or environmental degradation products. This can be a problem in trying to assess the risks of metabolites or environmental degradates. In the case of environmental degradates, a screening level risk assessment may be performed to determine if additional toxicity data on the degradate should be called in. The hazard component of the screening level assessment is often based on structural analogy of a degradate to the parent active ingredient. If parent and degradate are closely related structurally then toxicity data on the parent ai can be used to estimate the toxicity of the degradate. If the margin of exposure between estimated toxicity and estimated exposure is not considered large enough, additional toxicity data may be called in to enable a more comprehensive risk assessment of a degradate. On occasion, the metabolite or degradate bears little resemblance to the parent and an alternative analog with associated data must be found to support the screening level risk assessment.

The herbicide dichlobenil is relatively stable in the environment except for aqueous photolysis. A major photodegradate of dichlobenil in water (up to 19% of applied dichlobenil) has been identified as 4-chloro-2(3H)benzoxazolone (BZZ). This photodegradate bears little resemblance to the parent dichlobenil and therefore toxicity data on the parent are not considered useful for assessing the toxicity of BZZ. In the absence of appropriate toxicity data the degradate, termed BZZ, was determined to be of potential concern.

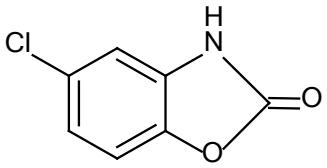
A number of online sources of information were consulted in search of appropriate analogs to BZZ including EPA/OPPT's Analog Identification Methodology (AIM) (http://aim.epa.gov/), ChemSpider (http://www.chemspider.com/) and Chemicalize (http://www.chemicalize.org/). A close analog (and isomer) of BZZ, 5-chloro-2(3H)benzoxazolone, was identified and this particular analog has a wealth of health effects data since it is an often prescribed muscle relaxant (common name chlorzoxazone).



Dichlobenil



BZZ



Chlorzoxazone

Chlorzoxazone has estimated physical properties very close to BZZ (see table below) and so bioavailability is likely to be similar. Chlorzoxazone is pharmacologically active as a muscle relaxant at doses of 10-30 mg/kg/day (http://www.drugs.com/monograph/chlorzoxazone.html), so the potential exists for BZZ to be biologically active at the same dosage, although no assumption is made about the kind of effects that might be observed at this dose of BZZ.

Estimation of selected physical-chemical properties using Episuite v4.1

Property	**Dichlobenil**	**BZZ**	**Chlorzoxazone**	**Benzoxazolone***	**Model**
log P	1.68	1.59	1.59	0.95	KOWWIN v1.68
K_{OC} (L/kg)	257	15.06	14.76	10	KOCWIN/MCI v2.00
water sol. (mg/L)	36	361	361	1414	Water NT v1.01
pK_a	n/a	8.38	8.43	8.91	SPARC v4.6
Henry's Law (atm-m3/mole)	2.86E-05	2.72E-08	2.72E-08	3.66E-08	Henry v3.20, bond estimation

*Benzoxazolone is the unsubstituted fused ring structure common to BZZ and chlorzoxazone

Screening Risk Assessment. The theoretical upper limit for BZZ based on estimated surface water concentration of parent dichlobenil is 0.005 mg/L and the corresponding dosage in a young child is approximately 0.0005 mg/kg/day for a 10 kg child ingesting 1 liter of BZZ contaminated water per day. The MOE between the lowest effective pharmacological dose of chlorzoxazone (10 mg/kg/day and the theoretical intake is 10 mg/kg/day divided by 0.0005 mg/kg/day or 20,000. Although there are many uncertainties in a screening level risk assessment such as this, the MOE is sufficiently large to conclude that additional toxicity data are unlikely to result in risks of concern from ingestion of BZZ formed in drinking water as a result of registered uses of dichlobenil.

Example No. 3

The following example has been provided by the Office of Pollution Prevention and Toxics (OPPT) and the Office of Pesticide Programs (OPP) at the US EPA. It serves to illustrate the potential use of SAR analysis as one line of evidence to support a mode of action analysis for a pesticide active ingredient.

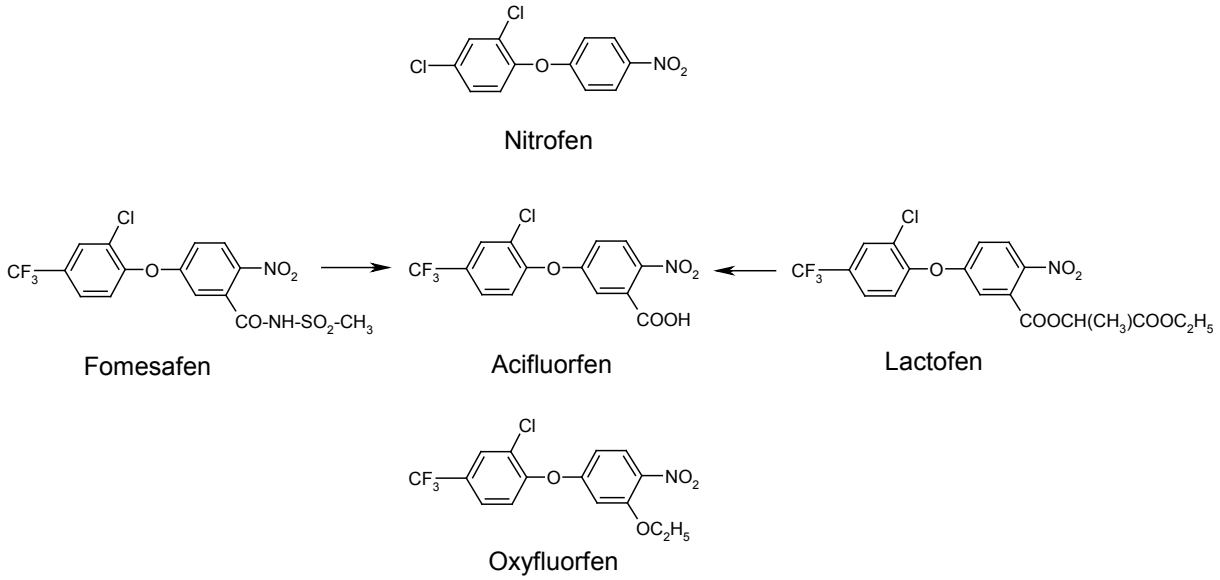
Case Study: Fomesafen cancer assessment and mode of action: use of mechanism-based SAR

Description of the case:

Fomesafen, a diphenyl ether herbicide, was submitted to OPP's Cancer Assessment Review Committee (CARC) for re-evaluation of its carcinogenic potential to humans. The herbicide was previously shown to be a mouse hepatocarcinogen by the submitter and classified as a Category C possible human carcinogen by OPP. The new data provided by the submitter included: (a) consistent negative genotoxicity data, (b) some evidence of involvement of peroxisome proliferator-activated receptor alpha receptor (PPARα) as a possible nongenotoxic mode of action for carcinogenicity, and (c) metabolism data. No SAR study was attempted. The Committee concurred that the pesticide should be nongenotoxic but considered the PPARα evidence inadequate.

SAR approaches conducted:

Several structurally related diphenyl ether pesticides with carcinogenicity data were identified. Among these, Nitrofen, Lactofen, Acifluorfen and Oxyfluorfen were considered the closest. Like Fomesafen, all four were hepatocarcinogenic in mice with Oxyfluorfen being weakly/marginally active. The chemical structures are shown in the figure below.



The presence of a nitro group in aromatic ring is generally considered a genotoxic structural alert. Indeed, there was some evidence that Nitrofen was positive in the Ames test but the evidence was complicated by the presence of impurities. In addition to mouse liver tumors, there was some evidence that Nitrofen may induce pancreatic tumors in the rat. The mode of action of Nitrofen has not been thoroughly studied.

The mode of action of rodent hepatocarcinogenesis for both Lactofen and Acifluorfen (HED MTARC; TXR #s 0051907 and 0052006, respectively) has been extensively studied and shown to involve PPARα-medicated peroxisome proliferation. Lactofen can be readily hydrolyzed by esterases to yield Acifluorfen as its primary metabolite. Structure-activity relationships studies have shown that one of the major structural requirements/alerts of most peroxisome proliferators is the presence of an acidic functional group (e.g., carboxylic, sulfonic) either in the parent compound or a metabolite (Woo and Lai 2003). The key question is whether Fomesafen can be hydrolyzed to a carboxylic acid metabolite. In general, the amide (-CO-NH-) bond is quite resistant to enzymatic hydrolysis. However, in Fomesafen, the presence of a sulfonyl group adjacent to the amide linkage can significantly facilitate hydrolysis. Indeed, a metabolism study by the submitter showed that up to 10% of Fomesafen may be hydrolyzed to yield a carboxylic acid metabolite as the most significant metabolite. Thus, Fomesafen, Acifluorfen, and Lactofen may actually have common carboxylic acid metabolite(s). It is interesting to note that, despite structural similarity, Oxyfluorfen, which cannot be metabolized to a carboxylic acid metabolite, is only weakly/marginally active as a hepatocarcinogen . Attempts to demonstrate possible PPARα-mediated activity were unsuccessful for Oxyfluoren. Overall, these findings strengthen the biological

plausibility of PPARα mode of action for Fomesafen-induced liver tumor formation in mice**.**

Outcome of the SAR study:

The SAR study provided significant support to the weight of evidence of a PPARα mode of action of Fomesafen-induced mouse liver tumors. Based on the current scientific understanding of peroxisome proliferation (e.g., Klaunig et al., 2003) and previous EPA decisions on structurally related herbicides (e.g., Lactofen and Acifluorfen), the level of confidence in this assessment is high. While the proposed mode of action for liver tumors in mice is theoretically plausible in humans, it is quantitatively implausible and unlikely to take place in humans based on quantitative species differences in PPAR□ activation and toxicokinetics. In accordance with the EPA Final Guidelines for Carcinogen Risk Assessment (March 29, 2005), the CARC classified Fomesafen as "Not Likely to be Carcinogenic to Humans".

References:

Klaunig JE, Babich MA, Baetcke KP, Cook JC, Corton JC, David RM, DeLuca JG, Lai DY, McKee RH, Peters JM, Roberts RA, Fenner-Crisp PA. (2003). PPARalpha agonist-induced rodent tumors: modes of action and human relevance. Crit Rev Toxicol 33(6):655–780.

Woo, Y.T., and Lai, D.Y. (2003). Mechanism of action of chemical carcinogens and their role in structure-activity relationships (SAR) analysis and risk assessment. In: *Quantitative Structure-Activity Relationship (QSAR) Models of Mutagens and Carcinogens*, R. Benigni, ed., CRC Press, p. 41.

Example No. 4

The following example has been abstracted from the Screening Assessment for the Challenge for Methylium, {4-(dimethylamino)phenyl]bis{4-(ethylamino)-3-methyphenyl]-, acetate prepared by Environment Canada and Health Canada in July, 2010 pursuant to section 74 of the *Canadian Environmental Protection Act 1999* (CEPA, 1999).

This example does not include the entire text or conclusions of the screening assessment document. The example is only an abstract of the following sections: Substance Identity, Physical and Chemical Properties, Health Effects Assessment, Appendix 6, and Appendix 7.

These sections have been abstracted to illustrate the application of (Q)SAR and information on analog substances to assess the toxicity of a substance in a weight-of-evidence type approach.

For a copy of the complete screening assessment document, please consult the following website:

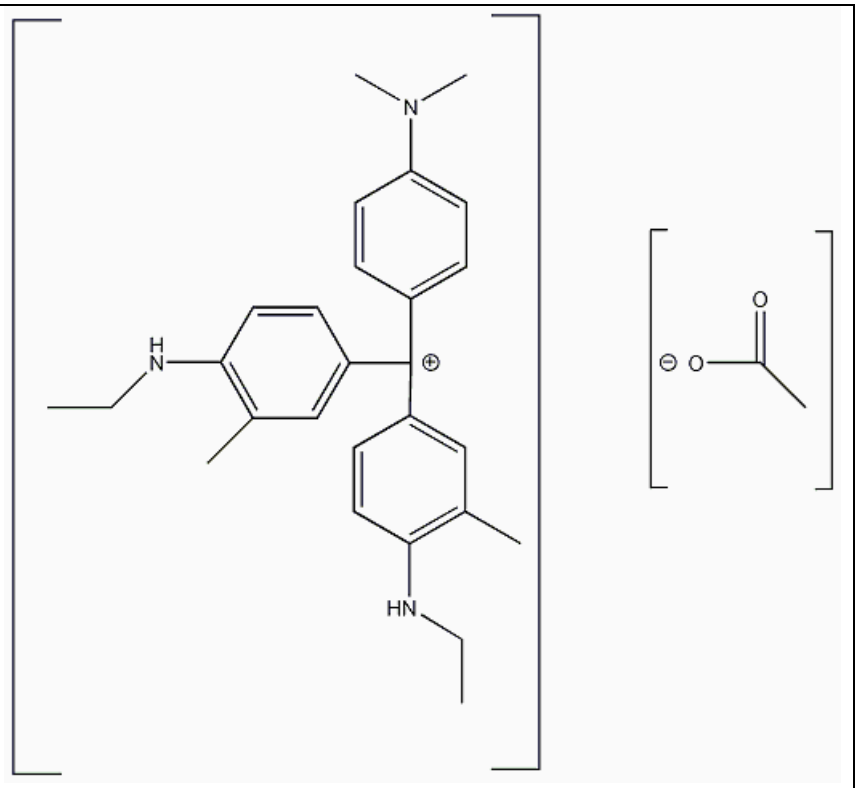
http://www.ec.gc.ca/ese-ees/default.asp?lang=En&n=403207BF-1

Substance Identity

For the purposes of this document, this substance will be referred to as MAPBAP acetate, derived from the DSL name. MAPBAP acetate belongs to a class of dyes known as cationic triarylmethanes. The class can be further sub-divided into those where the charge on the cation (triarylmethane moiety) is localized or delocalized. MAPBAP acetate belongs to the latter sub-category (Hunger 2003) implying that the bond holding the cationic and anionic components of the structure together is at least partly covalent.

Table 2. Substance identity for MAPBAP acetate.

Chemical Abstracts Service Registry Number (CAS RN)	**72102-55-7**
DSL name	**Methylium, [4-(dimethylamino)phenyl]bis[4-(ethylamino)-3-methylphenyl]-, acetate**
National Chemical Inventories (NCI) names[1]	*Methylium, [4-(dimethylamino)phenyl]bis[4-(ethylamino)-3-methylphenyl]-, acetate (1:1)* (TSCA) *Methylium, [4-(dimethylamino)phenyl]bis[4-(ethylamino)-3-methylphenyl]-, acetate* (AICS, PICCS, ASIA-PAC, NZIoC)
Other names	*[4-(Dimethylamino)phenyl]bis[4-(ethylamino)-3-methylphenyl]methylium acetate*
Chemical group (DSL Stream)	Discrete organics
Major chemical class or use	Cationic triphenylmethanes; anilines;
Major chemical sub-class	Secondary Aromatic Amines, Secondary Amines, Tertiary Amines, Tertiary Aromatic Amines
Chemical formula	$C_{27}H_{34}N_3.C_2H_3O_2$

Chemical structure [2]	
SMILES [3]	CN(c2ccc(cc2)C[(OC(=O)C)](c3cc(c(cc3)NCC)C)c1cc(c(cc1)NCC)C)C
Molecular mass (g/mol)	459.64

[1] National Chemical Inventories (NCI). 2006: AICS (Australian Inventory of Chemical Substances); ASIA-PAC (Asia-Pacific Substances Lists); PICCS (Philippine Inventory of Chemicals and Chemical Substances); NZIoC (New Zealand Inventory of Chemicals) and TSCA (Toxic Substances Control Act Chemical Substance Inventory).

[2] This substance is an organic salt, comprising a carbocation and an acetate anion.

[3] SMILES stands for : Simplified Molecular Line Input Entry System. This SMILES notation was used to generate predictions. It is for the neutral form of the molecule and indicates a covalent bond between the carbocation and acetate anion. This is typically how they are shown in EPIWIN. It is not fully established what effect using this SMILES will have on the predictions. The acetate part of the SMILES is placed in square brackets here to highlight the fact that the molecule is at least partly ionic.

I.

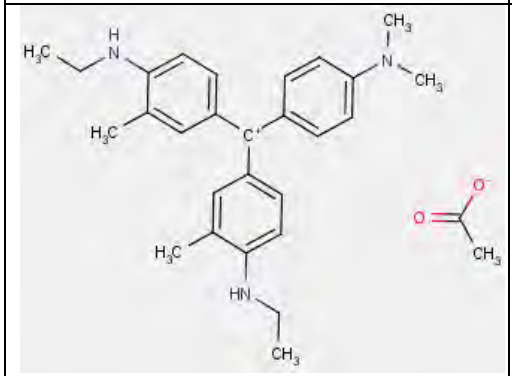
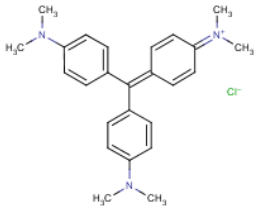
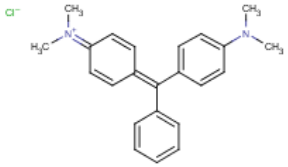
II. Physical and Chemical Properties

No experimental data are available for MAPBAP acetate. At the Environment Canada-sponsored Quantitative Structure-Activity Relationship (QSAR) Workshop in 1999 (Environment Canada 2000) modelling experts identified many structural classes of pigment and dyes as being “difficult to model” using QSARs. Some physical and chemical properties of many of the structural classes of dyes and pigments are not amenable to prediction by models. Under such circumstances, a "read-across" approach is considered which employs close analogues, to determine the approximate physical and chemical properties of MAPBAP acetate. A search of the ChemIDPlus (2009) database yielded a number of suitable analogues which are described in Table 2. Experimental data for these analogues, when available, were used as extrapolated (read-across) values for MAPBAP acetate or as supporting values for the weight of evidence.

A limited number of read-across data were found for the selected analogues and, therefore, predicted values are also used for MAPBAP acetate and the uncertainties of the predictions are noted.

Table 3 below contains predicted physical-chemical properties of the neutral form of MAPBAP acetate that are relevant to its environmental fate. Analogue data are available for water solubility and log K_{ow}. The water solubility of Ethyl Violet (CAS RN 2390-59-2) is 9000 mg/L (Green 1990). There is an indication that triphenylmethane acetates are more soluble than the chlorides (Pfenninger and Bruttel 1985) indicating the water solubility of MAPBAP acetate is high.

Table 2. MAPBAP acetate and its structural analogues

Methylium, [4-(dimethylamino)phenyl]bis[4-(ethylamino)-3-methylphenyl]-, acetate MAPBAP acetate (CAS RN 72102-55-7)	**Analogue 1** Ethanaminium, N-[4-[bis[4-(diethylamino)phenyl]methylene]-2,5-cyclohexadien-1-ylidene]-N-ethyl-, chloride Ethyl Violet (CAS RN 2390-59-2)	**Analogue 2** N-(4-(Bis(4-(dimethyl amino) phenyl) methylene) -2,5-cyclo hexadien-1-ylidene)-N-methyl methanaminium, Chloride Gentian violet (CAS RN548-62-9)	**Analogue 3** Methanaminium, N-(4-((4-(dimethylamino)phenyl)phenyl-methylene)-2,5-cyclohexadien-1-ylidene)-N-methyl-, chloride Malachite Green (CAS RN 569-64-2)
	**No Structure**		

Comparative analysis:

The differences between the chemical structures of MAPBAP acetate (i) and analogues 1,2 and 3 are:

- the number and position of the methyl, or ethyl, groups;
- the counteranions: acetate, for MAPBAP acetate and chloride (Cl-) for the analogues.

For all substances, the charge on the cation is de-localized. Resonance hybrids can occur and these affect the position of the counteranion (acetate for (i) and chloride for the analogues).

Table 3. Physical and chemical properties for the neutral form of MAPBAP acetate and analogues

Property	Substance	Type	Value[1]	Temperature (°C)	Reference
Melting point (ºC)	MAPBAP acetate	Modelled	236.73	-	MPBPWIN 2008
Boiling point (ºC)	MAPBAP acetate	Modelled	551.67	-	MPBPWIN 2008
Vapour pressure (Pa)	MAPBAP acetate	Modelled	9.13×10^{-10} (6.85×10^{-12} mmHg)	25	EPIWIN 2004
Henry's Law constant ($Pa \cdot m^3/mol$)	MAPBAP acetate	Modelled	1.92×10^{-10} (1.895×10^{-15} $atm \cdot m^3/mole$)	25	HENRYWIN 2008
Log K_{ow} (Octanol-water partition coefficient) (dimensionless)	Analogue (C.I. Basic Violet 3 CAS RN 548-62-9)	Experimental	0.51	-	Tsai et al. 1991

Property	Substance	Type	Value[1]	Temperature (°C)	Reference
K_{oc} (Organic carbon-water partition coefficient) (dimensionless)	MAPBAP acetate	Modelled	10.26 [2]	-	PCKOCWIN 2008
Water solubility [3] (mg/L)	Analogue (CAS RN 2390-59-2)	Experimental	9000	-	Green 1990
	MAPBAP acetate	Modelled	475 [2]	25	WSKOWWIN 2008

[1] Values and units in brackets represent those originally reported by the authors or estimated by the models.

[2] This value was modelled using the experimental analogue logKow of 0.51 as input,.

[3] Importer of MAPBAP acetate has indicated that it is completely soluble at environmental pHs (eg. pH 7).

Health Effects Assessment

No empirical toxicity data were identified for MAPBAP acetate. Sources of health hazard information considered included examination of available international reviews, assessments or classifications, reviewing the available empirical data where available and the use of predictive models as appropriate. The outputs of predictive models were also considered using five different QSAR models: TOPKAT (2004), CASETOX (2008), Toxtree (2009), DEREK 2008, and Model Applier (2009).

Using the representative molecular structure of MAPBAP (with the acetic acid fragment (acetate) attached to the carbon atom (attached to three aromatic rings)), the following results were obtained. Positive predictions were obtained on five different genotoxicity endpoints and only one of these (i.e. rodent micronucleus assay) is corroborated by more than one model (CASETOX and Toxtree). The Benigni-Bossa model within the Toxtree also predicts it to be a *Salmonella typhimurium* TA100 mutagen with metabolic activation. On the other hand, the female rat cancer models of both CASETOX and Model Applier gave positive predictions. The male rat cancer model of Model Applier as well as both mice models (male and female) of CASETOX gave positive predictions. The presence of a structural alert indicative of genotoxic carcinogenicity is another piece of supporting information that has been obtained from the Benigni-Bossa model within Toxtree. Applying the OncoLogic model to a nearly similar structure containing hydroxyl group in place of the acetate group results in a positive carcinogenicity prediction. This prediction is based on presence of Nitrogen substituted groups on the aromatic rings.

It is important to note that the Toxtree micronucleus model is a coarse grain filter for preliminary screening of potential *in vivo* mutagens and the OncoLogic does not use the identical structure for prediction purposes. Also, the Ames point mutation models of CASETOX and Model Applier predict negative results whereas TOPKAT and DEREK fail to provide any information. However, in the case of cancer models, there are at least three models (CASETOX, Model Applier and Toxtree) that classify this chemical as a potential carcinogen. The CASETOX, Model Applier and the Toxtree models are based on unique methodologies for making predictions and since they point towards a similar outcome, it carries more weight.

Thus the model predictions were mixed for carcinogenicity (6 positive and 4 negative), genotoxicity (6 positive and 7 negative), developmental (2 positive; 18 negative and 10 no result) and reproductive toxicity (1 positive and 12 no result).

Potential structural analogues of MAPBAP acetate for the purposes of read-across for human health toxicity information were identified using Leadscope (Leadscope 2008)

and ChemID (ChemIDPlus 2009) along with professional judgement. As the main structure would distribute the positive charge across the molecule through resonance structures, the acetate counter ion would likely interchange with other ions or substrates when the dye is used. Therefore, the moiety of interest from a human health toxicological perspective would be the parent molecule itself. Other similar triarylamine substances, that have empirical data, include gentian violet (CAS 548-62-9), malachite green (CAS 569-64-2), C.I. Basic Violet 4 (CAS 2390-59-2) and leucomalachite green (CAS 129-73-7) as shown in Appendix 7.

Gentian violet has been classified by the European Union as Carcinogenicity Category 2 (ECB 2002) based on carcinogenicity in experimental animals. One study did report negative *in vitro* genotoxicity for mutations in a reverse mutation assay in several *S. typhimurium* strains after exposure to gentian violet at concentrations ranging from 5 – 1000 µg/plate (NICNAS 1999). Malachite green has been classified by the European Union as Reproductive Toxicity Category 3(ECB 2003) based on developmental toxicity in experimental animals. Also, the U.S. NTP (2005) reported equivocal evidence of carcinogenicity in female rats and negative results for genotoxicity from an *in vivo* micronucleus assay and an *in vitro* assay in *S. typhimurium* (NTP 1997, 1994). C.I Basic Violet 4 had negative *in vitro* genotoxicity data for chromosomal aberrations in Chinese Hamster Ovary cells (NICNAS 1999) and was also found to be predominately negative *in vitro* in assays conducted in *S. typhimurium* and mouse lymphoma cells (CCRIS 2009). Leucomalachite green was found to have some evidence of carcinogenicity in female mice and had positive *in vivo* genotoxicity data (NTP 1996, 2005).

The information obtained from the QSAR models as well as potential analogues, suggest that there may be potential carcinogenic or developmental toxicity hazards associated with the substance.

The confidence in the toxicity database is considered to be low due to the lack of available data for MAPBAP acetate.

References

CASETOX [Prediction module]. 2008. Version 2.0. Beachwood (OH): MultiCASE. [cited 2009 September 30]. Available from: http://www.multicase.com/products/prod03.htm [restricted access].

[CCRIS] Chemical Carcinogen Research Information System. 2009. Mutagenicity studies for C.I basic violet 4. Available from: http://toxnet.nlm.nih.gov/cgi-bin/sis/search/r?dbs+ccris:@term+@rn+2390-59-2

ChemIDPlus. 2009. Tool for searching the NLM ChemIDplus database using Chemical Name, CAS Registry Number, Molecular Formula, Classification Code, Locator Code, and Structure or Substructure. http://chem.sis.nlm.nih.gov/chemidplus/

[DEREK] Deducting Estimation from Existing Knowledge [Prediction module on CD ROM]. 2008. Version 10.0.2. Cambridge (MA): Harvard University, LHASA Group. [cited 2010 05 31]. Available from: http://lhasa. harvard.edu/?page=toxicology.htm [restricted access]

[ECB] European Chemicals Bureau. 2002. Summary record. Commission Working Group on the Classification and Labelling of Dangerous Substances. Meeting at ECB Ispra,16–18 January 2002. European Commission, Directorate-General Joint Research Centre, Institute for Health and Consumer Protection, European Chemicals Bureau. ECBI/15/02-Rev. 3. Available from: http://apps.kemi.se/hclass/DocumentDownload.aspx?DocId=922706

[ECB] European Chemicals Bureau. 2003. Summary record. Commission Working Group on the Classification and Labelling of Dangerous Substances. Meeting at ECB Ispra,15–17 January 2003. European Commission, Directorate-General Joint Research Centre, Institute for Health and Consumer Protection, European Chemicals Bureau. ECBI/30/03-Rev. 3. Available from: http://apps.kemi.se/hclass/DocumentDownload.aspx?DocId=923632

Environment Canada. 2000. Chemicals Evaluation Division. *Environmental Categorization for Persistence, Bioaccumulation and Inherent Toxicity of Substances on the Domestic Substances List Using QSARs. Final Report.* Environment Canada.

Green, FJ. 1990. The Sigma-Aldrich handbook of stains, dyes and indicators. Aldrich Chemical Company, Inc. Milwaukee, Wisconsin.

Hunger, K. (ed.). 2003. Industrial Dyes: Chemistry, Properties, Applications. Wiley-VCH, Weinheim.

Model Applier [Prediction module] 2009. Columbus (OH): Leadscope Inc. [cited 2009 September 30]. Available from: http://www.leadscope.com/all_products.php [restricted access].

[NCI] National Chemical Inventories [database on a CD-ROM]. 2006. Issue 1. Columbus (OH). American Chemical Society, Chemical Abstracts Service Columbus, Ohio. Available from: http://www.cas.org/products/cd/nci/require.html

[NICNAS] National Industrial Chemicals Notification and Assessment Scheme 1999. Full Public Report. Component of Basazol Violet 46 L. File No: NA/538. Available at: http://www.nicnas.gov.au/publications/car/new/na/nafullr/na0500fr/na538fr.pdf

[NTP] National Toxicology Program 1994. Study report number A90714. Available from: http://ntp-apps.niehs.nih.gov/ntp_tox/index.cfm?fuseaction=salmonella.salmonellaData&endpointlist=SA&study%5Fno=A90714&cas%5Fno=569%2D64%2D2&activetab=summary

[NTP] National Toxicology Program 1996. Study report number A31156. Study start date 07/20/1996

[NTP] National Toxicology Program 1997. Study report number A82983. Study start date 01/16/1997 Available from: http://ntp-apps.niehs.nih.gov/ntp_tox/index.cfm?fuseaction=micronucleus.micronucleusData&cas_no=569%2D64%2D2&endpointlist=MN

[NTP] National Toxicology Program 2005. Toxicology and Carcinogenesis Studies of Malachite Green Chloride and Leucomalachite Green (CAS Nos. 569-64-2 and 129-73-7) in F344/N Rats and B6C3F1 Mice (Feed Studies). Study report number C93029. Study Report date February, 2005.

Pfenninger Heinz and Bruttel Beat. 1985. Process for converting sparingly soluble inorganic salts of cationic dyes and brighteners into more soluble salts of organic acids. United States Patent 4559144. Available from: http://www.freepatentsonline.com/4559144.html

[TOPKAT] Toxicity Prediction Program [Internet]. 2004. Version 6.2. San Diego (CA): Accelrys Software Inc. Available from: http://www.accelrys.com/products/topkat/index.html

Toxtree version 1.60. 2009. Tool to estimate toxic hazard by applying decision tree approach. Developed by Ideaconsult Ltd Bulgaria.

Appendix 6: Summary of (Q)SAR Results

(Q)SAR PREDICTIONS ON CARCINOGENICITY

Model/ Species	Mice		Rat		Rat	Mice	Rodent	Mammal
	Male	Female	Male	Female				
Model Applier	N	N	P	P	P	N	N	-
Multicase Casetox	P	P	ND*	P	-	-	-	-
Topkat	NR	NR	NR	NR	-	-	-	-
Derek	-	-	-	-	-	-	-	NR

** This one is weakly positive (30 case units & 81 % probability)*

(Q)SAR PREDICTIONS ON GENOTOXICITY

Model/endpoints	chrom. ab.	chrom. ab. other rodent	chrom. ab. rat	micronucleus mice	micronucleus rodent	drosophila	drosophila HT	drosophila SLRL	mam. mutation	mam. mutation DL	UDS	UDS human lymphocytes	UDS rat hepatocytes	mouse lymphoma mut	s. cerevisiae	yeast	hgprt	e. coli	e. coli w	microbial	salmonella	BB cancer alert
MA	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	P	N	ND	-	N	N	ND	N	ND	P	N	-
CT	N	-	-	P	-	ND	-	-	-	-	NR	-	-	ND	-	-	-	-	-	-	N	-
TK	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	NR	-
TT	-	-	-	-	P	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	P	P

(Q)SAR PREDICTIONS ON REPRODUCTIVE TOXICITY

Model Applier

Model/ endpoint	Female			Male		
Species	mice	rat	Rodent	mice	rat	rodent
repro	ND	ND	ND	ND	ND	ND
sperm	-	-	-	ND	ND	ND

Multicase Casetox

mice	rat	rabbit	human
NR	P	NR	NR

(Q)SAR PREDICTIONS ON DEVELOPMENTAL TOXICITY

Model Applier

Endpoint/ Species	mice	rabbit	rat	rodent
Retardation	N	ND	N	N
Weight decrease	N	ND	N	N
Fetal death	N	ND	N	N
Post impl. loss	ND	ND	N	N
Pre impl. loss	P	ND	N	N
Structural	N	ND	ND	N
Visceral	N	-	N	N

Multicase Casetox

Endpoint/Species	Hamster	Mammal	Miscellaneous
Teratogenicity	-	P	NR
Developmental	NR	-	-

MA – model applier;

CT – Multicase Casetox;

TK – Topkat;

TT – Toxtree;

BB – Benigni-Bossa rule;

ND – not in domain;

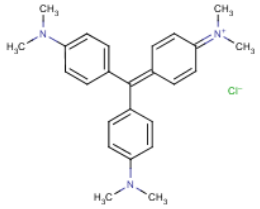
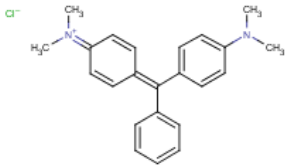
'-' no model available in QSAR suite

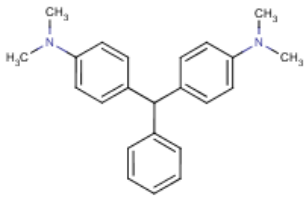
NR – no result

P – positive

N – negative

Appendix 7: Analogues of MAPBAP acetate considered in Human Health portion of assessment

Structure	Name / CAS RN	Data/Classifications
	Genitian violet 548-62-9	European Union Carcinogenicity Category 2 (ECB 2002) based on evidence in experimental animals **Genotoxicity** *In-vitro* reverse mutation: Negative in *S.typhimurium* TA98, TA100, TA1535, TA1537 with and without activation (NICNAS 1999).
	Malachite green 569-64-2	European Union Reproductive Category 3 (ECB 2003) **Carcinogenicity** Equivocal evidence of cancer in female rats (NTP 2005) **Genotoxicity** *In-vitro* gene mutation: Negative in *S.typhimurium* TA97, TA98, TA100, TA102, TA104, TA1535 with and without activation (NTP 1994). Chromosome aberration: *In vivo*: Negative in mouse micronucleus bone marrow and peripheral blood (NTP 1997).

`	C.I Basic Violet 4 2390-59-2	**Genotoxicity** *In vitro* gene mutation: Negative in *S.typhimurium* TA98, TA100, TA1537, TA1538 with and without activation; TA1535 without activation (CCRIS 2009): Positive in TA1535 with S9 activation (CCRIS 2009) Negative in Mouse Lymphoma L5178Y with and without activation (CCRIS 2009) Chromosome aberration: Negative in Chinese Hamster Ovary Cells V79 with and without S9 activation (NICNAS 1999).
	Leucomalachite green 129-73-7	**Carcinogenicity** Some evidence of carcinogenicity in female mice (NTP 2005) **Genotoxicity** *In vivo* chromosome aberration: Positive in female mouse micronucleus peripheral blood study (NTP 1996).

Example No. 5

The following example has been provided by the Office of Pesticide Programs (OPP) at the US EPA. It serves to illustrate the potential use of SAR analysis and weight of evidence approaches in risk assessment decision making.

Case Study: Use of a weight of Evidence (WOE) approach, including SAR information, to waive the chronic toxicity/carcinogenicity study requirement in a biocide reregistration decision.

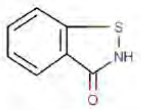
1, 2-benzisothiazolin-3-one (BIT) is a member of the isothiazolone class of biocides. Some of the registered uses of BIT involve chronic/long term exposures (e.g., use in metal working fluids). To address this type of potential exposure scenario, chronic and/or cancer studies would usually be required.

The chemical structures of the isothiazolone biocides can be divided into two sub-classes (Figure 1):

1. **General Isothiazolone Class:** Isothiazolone pesticide chemicals without a benzene ring attached at the 4-5 position of the isothiazolone ring. For example: Kathon RH287, Kathon RH886 and/or OIT, which have been registered by OPP.

2. **1,2-Benzoisothiazolone Class:** Isothiazolone pesticide chemicals with a benzene ring attached at the 4-5 position of the isothiazolone ring. In this case, BIT is the chemical being discussed.

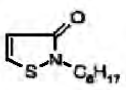
Isothiazolone Pesticidal Chemicals with benzene ring



BIT

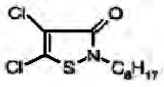
1,2-Benzisothiazolin-3-one

Isothiazolone Pesticidal Chemicals without benzene ring



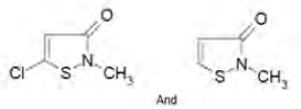
Octhilinone

2-n-Octyl-4-isothiazolin-3-one or OIT



Kathon RH287

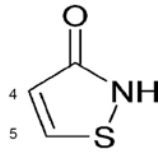
4,5-Dichloro-2-n-octyl-4-isothiazolin-3-one or DCOIT



RH 886 (Mixture of CMIT and MIT)

Figure 1. Chemical Structures of the Isothiazolone Biocides

All isothiazolone biocides contain an isothiazolone ring (Figure 1).



The issue to be discussed in the example case is whether chronic/cancer studies can be waived based on existing conditions. The issue is discussed from following different aspects.

- **Pesticidal Mode of Action:** BIT, CMIT, MIT, OIT, and DCOIT all share a common pathway for antimicrobial activity:
    - All inhibit cell respiration
    - All inhibit the same class of dehydrogenase enzymes

  These biocides react with microbial cells through cleavage of the S-N bond to form an S-S linkage with the thiol group on target enzymes. Biocidal activity is a function of the inhibition of cell respiration.

- **Structure Activity Consideration:** According to the pesticide mode of action consideration, the antimicrobial activity for all isothiazolone classes is due to the isothiazolinone ring and the sulfur nitrogen bond in the isothiazolinone ring plays a key role for efficacy as a biocide. The current issue is whether the benzene ring will be a concern for potential toxic effects of BIT. Based on SAR information, the benzene ring may prolong the biological half-life of the metabolic intermediate moieties in the body.

- **Toxicity Profile for BIT:** There are no carcinogenicity or chronic toxicity studies for any of the benzene ring-isothiazolone chemicals (such as BIT). BIT is not mutagenic as all acceptable guideline mutagenicity studies were negative. The toxicity profile of BIT shows that it is an irritant following oral and dermal exposures, and this is the effect observed following repeated dosing in subchronic toxicity studies.

    - In two oral subchronic rat studies, gastrointestinal irritation was reported at 10 mg/kg/day (lowest dose tested), and there were no other treatment-related systemic effects.

    - In a 90-day rat dermal study, skin irritation and histopathology were noted at all doses of 100, 300 and 1000 mg/kg/day, while systemic toxicity was only reported at the limit dose (1000 mg/kg/day). Gastrointestinal

irritation/histopathology was also reported at 100 mg/kg/day which may be attributed to grooming.

- The repeated-dose metabolism and disposition study indicates the metabolites associated with BIT exposure may remain in the body much longer than the parent compound. The evidence from this study indicates benzene containing metabolites may accumulate in the liver, kidneys and thyroid gland.

- **Toxicity Profile for the non-benzene ring isothiazolone pesticides:** The mutagenicity data for the non-benzene ring containing isothiazolones were largely negative except for a few positive observations *in vitro* with CMIT/MIT and DCOIT. Three chronic/carcinogenicity studies are available for non-benzene ring isothiazolone pesticides and all were negative for carcinogenicity, although two of these were found to have major deficiencies for the chronic toxicity portion of the studies.

    - One study was conducted using drinking water administration of a 14.2% CMIT/MIT mixture at doses of 2.0/3.1, 6.6/9.8, and 17.2/25.7 mg/kg/day in rats males/females. This study reported hyperplasia of the GI tract but no other systemic effects.

    - The second study used dermal administration of a single dose of 400 ppm CMIT/MIT to the skin of mice for 30 months and the only significant finding was dermal irritation.

    - A carcinogenicity study was conducted using dietary administration of OIT. There were no reported carcinogenic effects following oral exposure to up to 1000 ppm in the diet for 78 weeks. Although some tumors were reported, the incidences were within the ranges for the control animals.

  All of the isothiazolones produced toxicity at the site of contact, i.e. irritation of the gastrointestinal tract, skin and respiratory tract, when administered at high doses. These biocides produce minimal to no significant systemic toxicity; no histopathological change distant from the site of dosing was observed, which appears correlated with rapid metabolism and excretion for these chemicals. Based on read-across comparison, it is concluded that:

    - Skin irritation: Similar findings in all dermal studies (BIT, CMIT/MIT, OIT) although at different dose levels.
    - Skin histopathology: Similar for BIT and OIT, none found in rabbit study on CMIT/MIT
    - Similar clinical chemistry findings with BIT and OIT and similar to BIT oral dog study
    - Severe skin irritation in BIT dermal study

The relative potency is as follows:

Skin Irritation: CMIT, DCOIT > OIT, MIT, BIT

Skin Sensitization: CMIT ≥ DCOIT >OIT > MIT > BIT

- **Risk assessment considerations:** As a class, the isothiazolone pesticides are irritants by all routes of exposure, and are dermal sensitizers. For BIT, gastrointestinal irritation provides the basis for points of departure for short, intermediate, and chronic/long-term exposure scenarios.

Final Recommendation:

Based on a read-across comparison and weight of evidence (WOE) approach, that the chronic toxicity/carcinogenicity study for BIT is not required at this time if the risk assessment is protective of irritation. This recommendation was based on the following considerations: 1) available cancer studies for the isothiazolone pesticides are negative; 2) there is a lack of mutagenicity concern for BIT, and the other isothiazolone pesticides; 3) BIT and the other isothiazolones are irritants following oral, dermal and inhalation exposures and produce similar effects following subchronic exposures; 4) the isothiazolones as a group have a known mode of action for antimicrobial activity; 5) irritation is the predominant effect and is the basis of the PODs; 6) although the metabolism study for BIT showed an increased half life and accumulation of radioactivity in thyroid compared to other isothiazolone chemicals, these observations were determined to be not of toxicological significance, as the toxicological effects of BIT up to 90 days were not different than the effects observed with the other isothiazolone chemicals.

It is recommended that the available data be evaluated to inform the need for a UF to account for subchronic to chronic exposure durations for BIT.