(Q)SARs: Evaluation of the commercially available software for human health and environmental endpoints with respect to chemical management applications

Technical Report No. 89

ISSN-0773-8072-89
Brussels, September 2003

ECETOC Technical Report 89

© Copyright - ECETOC
European Centre for Ecotoxicology and Toxicology of Chemicals
4 Avenue E. Van Nieuwenhuyse (Bte 6), B-1160 Brussels, Belgium.

All rights reserved. No part of this publication may be reproduced, copied, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the copyright holder. Applications to reproduce, store, copy or translate should be made to the Secretary General. ECETOC welcomes such applications. Reference to the document, its title and summary may be copied or abstracted in data retrieval systems without subsequent reference.

The content of this document has been prepared and reviewed by experts on behalf of ECETOC with all possible care and from the available scientific information. It is provided for information only. ECETOC cannot accept any responsibility or liability and does not provide a warranty for any use or interpretation of the material contained in the publication.

(Q)SARS: Evaluation of the commercially available software for human health and environmental endpoints with respect to chemical management applications

CONTENTS

SUMMARY

With the marked escalation in demand for information on the potential effects of chemicals on health and the environment, there is a renewed interest in the development and use of (Q)SARs ((quantitative) structure activity relationships) to meet these demands with greater speed, and with less resources. Of particular importance being the potential for such models to minimise the need for animal testing.

Most recently the proposal for a new chemical policy in the European Union (EU (EC, 2001) represent a major challenge for industry, and indeed for all stakeholders, in the numbers of chemicals to be evaluated for their potential to affect adversely human health and the environment.

This report evaluates the current status of the commercially-available (Q)SAR approaches for human health and environmental endpoints, in the context of their applicability to product development and regulatory decision making, such as in hazard assessment (classification) or risk assessment.

(Q)SARs are simplified (mathematical) representations of complex chemical-biological interactions. They can be divided into two major types, QSARs and SARs. QSARs are all quantitative models yielding a continuous or categorical result. The most common techniques for developing QSARs are regression analysis, neural nets and classification methods. Examples of regression analysis include ordinary least squares and partial least squares. Examples of classification methods are discriminant analysis, decision trees and distance based methods of similarity analysis. SARs are qualitative relationships in the form of structural alerts that incorporate molecular substructures or fragments related to the presence or absence of activity.

(Q)SAR predictions are potentially more uncertain than the underlying test data. This imposes limitations on the acceptable use of (Q)SAR in chemical management and decision-making. Approaches to determine the acceptability of (Q)SAR predictions have been developed in the past, but because of their breadth and generality they have not been widely applied or respected by either (Q)SAR users or developers. As a consequence, decision making on the basis of existing models must be done with care and is subject to expert opinion as there is currently no framework for QSAR use and therefore a lack of confidence in these predictions.

Industry is being encouraged to lead the initiative to build more widespread confidence in the appropriate use of (Q)SARs in chemicals management and, towards this goal, to develop jointly with other stakeholders a framework to govern such approaches. This framework should address the acceptability of both negative and positive hazard predictions.

This report reviews the currently-available, most preferred (Q)SAR-based predictive software against acceptability criteria that were developed during the ICCA workshop on regulatory acceptance of QSARs for both human health and environmental endpoints (Cefic, 2002). Only endpoints for which models with sufficiently large databases are available are considered. Difficulties encountered with their modelling are discussed.

Current common modelling techniques that find relationships between parent structure and the endpoint limit applicability and predictive powers of (Q)SARs. Their use is suitable only in those cases where absorption, distribution, metabolism and excretion (ADME) is not important or can be neglected. for the endpoint in question, e.g.:

- Intensive physico-chemical properties;
- acute aquatic toxicity via nonpolar narcosis;
- chemicals with common mode of action (e.g. acute aquatic toxicity via specific mechanisms, carcinogenicity of amines).

Correspondingly, (Q)SAR predictions are not suitable for activity or effects for broad chemical classes where ADME is important (e.g. bioaccumulation, carcinogenicity, chronic toxicity, reproductive toxicity).

Prediction systems for broad chemical classes will normally require complex modelling approaches invoking a (Q)SAR model interactively communicating with ADME or receptor models. In principle it is possible to build such systems if mechanisms are understood. However, since the underlying mechanisms are rarely fully understood, such modelling approaches may not be generally possible for endpoints such as acute and chronic toxicity, carcinogenicity, reproductive toxicity, sensitisation and irritation. Currently, several metabolism simulation programs (MultiCASE: META, DEREK: Meteor, HazardExpert: MetabolExpert) are available that have the potential to improve the understanding of both toxicity mechanisms and detoxifying mechanisms. However, they are not always linked to the (Q)SAR systems, and there is a need to validate these approaches. In general, ADME processes and receptor modelling, are areas of intensive research, but links need to be established to (Q)SAR models applicable to broad chemical classes. Currently, (Q)SAR systems cannot reliably replace experimental testing, though they can provide warnings/alerts about possible toxic properties of the query compounds. These alerts need to be interpreted by experts in the relevant field of toxicology.

The availability of quality data is another important factor in assessing predictivity of (Q)SARs. Physico-chemical property models are in general well developed, as they are founded on large data sets and provide reliable estimations.

In general, (Q)SARs for environmental endpoints are founded on larger databases than those for human health endpoint models. Further, the models are often poor because the endpoints are expressed through many different mechanisms, are receptor-mediated, involve multi-stage processes comprising ADME and are site specific. At the present time, this complexity imposes severe limitations on the successful development of (Q)SARs suitable for non-congeneric sets of compounds. Hence, developments in the human health area are focused on sharing of data and expansion of databases, which include quality evaluation of the data. To aid these developments, it will be necessary to overcome the barriers to the sharing of proprietary information.

In promoting the exchange of data and building of these databases, greater emphasis should be placed on those experimental data that include mechanistic information.

For the time being, (Q)SAR models should only be used by experts looking objectively at the estimates and considering the overall weight of evidence of the available information on the substance in question.

Table 1 summarises the Task Force assessments of the applicability pertaining to (Q)SARs for the IUCLID endpoints. Where an endpoint is not mentioned it is because no suitable (Q)SAR was identified. From this analysis, it is clear that in the short term (Q)SARs will find greater acceptability for predicting environmental endpoints than human health endpoints.

Table 1: Applicability of currently available (Q)SARs for IUCLID endpoints

Physico-chemical properties		**Environmental fate and excotoxicity**		**Human health**	
Melting point	Good	Photodegradation	Limited	Acute oral toxicity	Limited
Boiling point	Good	Hydrolysis	Limited	Skin irritation	Limited
Vapour pressure	Very good	Biodegradation	Good	Eye irritation	Limited
Partition coefficient	Very good	Bioaccumulation	Good	Skin sensitisation	Limited
Water solubility	Very good	Acute tox to fish	Good	Chronic mammalian toxicity	Very limited
		Acute tox to invertebrates	Limited	Mutagenicity *in vitro*	Limited to good
		Acute tox to algae	Limited	Carcinogenicity	Very limited
		Acute tox to bacteria	Limited	Teratogenicity	Very limited
		Chronic toxicity	Very limited		

There is a need to develop (Q)SAR-based approaches, that are applicable for broad chemical classes, to address the escalating demand for information on their potential effects on health and the environment. If necessary, each endpoint could have several more-restricted (Q)SAR models combined in one interface for the user, where the most appropriate (Q)SAR is chosen by an electronic expert system based on agreed acceptability criteria. Such a system could also include a rigorous framework for regulatory use of (Q)SARs, including support of the system and maintenance of the underlying data and reports by an independent body. Any QSAR decision support system (QDSS) developed should have an open architecture, which would allow the addition of new models as they are developed.

1. INTRODUCTION

The regulatory demands have continued to escalate for information on the potential health and environmental effects of chemicals. Most recently, within the European Union (EU), the European Commission (EC) published its White Paper for a new EU Chemicals Policy (EC, 2001). Proposals in this White Paper, when translated into regulation, would require the generation of extensive data on thousands of chemicals. This represents a major challenge to both industry and to regulators in resourcing the efforts needed in the stated time frame addressing the concerns of animal rights organisations, and indeed all the stakeholders, to reduce, refine and replace animal testing wherever possible. Indeed, the White Paper explicitly recommends the use of alternative methods, both *in vitro* and *in silico*, for data gap filling. These developments have re-awakened interest in (Q)SARs and their potential for use in the regulation of chemicals.

(Q)SARs organise existing knowledge on chemicals about an endpoint (effect or property) with the purpose of generalising this knowledge and allowing for extrapolations to be made to other chemicals for which data are unavailable. They are indeed computer-based predictive models, i.e. *in silico* approaches.

However, because they are simplified mathematical representations of complex chemical-biological interactions, (Q)SAR predictions are potentially more uncertain than the underlying test data. This is because often the mechanisms of action are not known, because of the variability in the data on which they are built, e.g. the measurement uncertainty, and the number and type of chemicals in the training set. Thus, whilst the use of (Q)SARS to assist the process of chemical hazard and risk assessment is of considerable interest to industry and to regulators, there remain questions over their reliability, applicability and overall scope for use in such decision making.

One reason for this is that advances in (Q)SAR science have mainly been applied in product development. This has not been matched by increased regulatory applications of (Q)SARs and the acceptance, frameworks that delineate the correct use of (Q)SARs, and prevent their mis-use, need to be developed and respected by practitioners and other stakeholders. To make progress towards this goal there is a need to:

- Understand the current status of (Q)SARs in regulatory applications;
- identify why, up to now, these applications have been limited;
- identify a way forward for industry to meet the requirements of the EC White Paper and define the role of (Q)SARs in this process.

To address these issues ECETOC established a Task Force with the following Terms of Reference:

1. Review the use of (Q)SARs within the current regulatory decision-making frameworks in EU, North America (USA, Canada) and Japan. The review should address how regulators are using methods for reliability and uncertainty assessment and applicability evaluation of (Q)SARs;
2. Review the current use of (Q)SARs within the industry and develop a position on the future use of (Q)SARs for environmental and human health effects endpoints and physical chemistry for application in prioritisation, screening, classification and risk assessments in the context of European Commission White Paper (Com(2001) 88 Final) and High Production Volume (HPV) programme;
3. Organise a multi-stakeholder workshop (e.g. EU, US EPA, OECD, industry, academia) to scope out the needs to develop harmonised acceptability criteria and a strategy for the use of (Q)SARs in classification and labelling and risk assessment;
4. Prepare a document, based on the outcomes of the workshop;
5. Develop a targeted research proposal to address identified research gaps.

To address the first Term of Reference, the Task Force commissioned a review of the use of (Q)SARs within the current regulatory decision-making frameworks in EU, North America (USA, Canada) and Japan. Two papers have been published in Environmental Health Perspectives, namely:

- Cronin et al (2003a) - Use of QSARs in international decision-making frameworks to predict ecological effects and environmental fate of chemical substances;
- Cronin *et al* (2003b) - Use of QSARs in international decision-making frameworks to predict human health effects of chemical substances.

A summary of both papers is provided in Section 2 of this report.

Section 3 addresses the second Term of Reference, by describing the current use of (Q)SAR models in industry, and reviewing them against acceptability criteria for existing data. Sections 4 and 5 contain conclusions and recommendations respectively.

To address the third term of reference, the Task Force, under auspices of the Long-range Research Initiative (LRI), organised a multi-stakeholder workshop on regulatory acceptance of (Q)SARs for human health and environmental endpoints (March 4-6th 2002 in Setubal, Portugal). As preparation for the workshop, a thorough review of uncertainty and reliability assessment methods was prepared by Eriksson *et al* (2003). The paper has also been published in Environmental Health Perspectives.

A summary of this paper, specifically addressing the use of reliability, applicability and uncertainty methods by regulators, is given in Section 2.4. The proceedings of the Setubal workshop (Cefic, 2002), in part addresses the fourth Term of Reference; the executive summary of the workshop is included as Appendix A to this report.

Section 6 describes the current availability of experimental data and the need to improve the sharing of data. Such developments will be the foundation for further progress in computer-based predictions.

Section 7 deals with the TF recommendations on scope for use of (Q)SARs, and addresses the fourth Term of Reference.

Section 8 presents research recommendations in the area of (Q)SARs.

2. CURRENT USE OF (Q)SARS BY REGULATORY AUTHORITIES

2.1 Physico-chemical properties

Cronin *et al* (2003a) address the key physico-chemical properties (including water solubility, log K_{ow}, Henry's Law constant, melting point, boiling point and vapour pressure) in their review of the use of (Q)SARs in international decision-making frameworks to predict the ecological effects and environmental fate of chemical substances. A brief description of general reviews is provided for each endpoint. Currently, the EPIWIN software is clearly the predominant model in use.

2.2 Environmental endpoints

Cronin *et al* (2003a) describe the regulatory uses of (Q)SARs, focussing on North America and Europe. Initially the paper addresses how (Q)SARs are used, i.e. data evaluation, decision making for test strategies, establishing specific parameters for model input and identifying data needs for effects of special concern. The use of (Q)SARs by the Oslo Paris Commission (OSPAR) in identifying chemicals which persist, bioaccumulate and are toxic (PBT), by the Danish EPA as an aid to self-classification and by Germany and the Netherlands for assessing chemicals, are all described. The paper also discusses the recommendations of the Scientific Committee on Toxicology, Ecotoxicology and the Environment (CSTEE) for further uses of QSARs and the probable extension of (Q)SAR use through the EU Chemicals Policy review.

The use of Structure-Activity Relationships (SARs) in the USA for identifying chemicals of concern for further testing by the Interagency Testing Committee (ITC) is described. So too is, the need for the United States Environmental Protection Agency (US EPA) to use SARs for the assessment of chemicals submitted under the Toxic Substances Control Act Pre-Manufacturing Notification (TSCA PMN) programme, for which there is no requirement for testing. Because there are approximately 2000 PMNs per annum, there is considerable experience in the use and development of (Q)SARs within the US EPA. Canada is using (Q)SARs in categorising the Domestic Substance List. Finally the uses in other regions, specifically by Japan and Australia are briefly discussed. Relevant websites that provide further useful information are listed in Appendix B.

The activities of other bodies involved, either in the promulgation of (Q)SARs (e.g. Organisation for Economic Co-operation and Development, OECD), or the validation of alternative methods (e.g. the European Centre for Validation of Alternative Methods (ECVAM) and the US Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM)) are also described. A number of useful references to reviews and recommendations are made, including the website that gives details of a comparative study done under the tripartite assessment of (Q)SARs by the US EPA, the European Chemicals Bureau (ECB) and Japanese regulatory authorities (see Appendix B).

The paper also describes the key resources available and used by regulatory agencies for each endpoint. Acute toxicity prediction is done with the computer programs ECOSAR, TOPKAT, CASE and OASIS. The EC Technical Guidance Document (TGD) (EC, 1996, 2003) contains acute aquatic toxicity (Q)SARs (based on mode of action) but they are not integrated into the EUSES software. There are few (Q)SARs for chronic toxicity and they are not used for regulation of chemicals. The softwares used to predict biodegradation are BIOWIN (predominant), MultiCASE and CATABOL. The two most-favoured approaches among regulatory agencies to predict bioaccumulation are a log K_{ow} based (Q)SAR, with expert judgement regarding metabolism, or BCFWIN, which implicitly considers metabolism through use of correction factors. For soil and sediment sorption prediction, PCKOC or log K_{ow} based (Q)SARs are being used.

Finally the paper describes the on-going programmes addressing HPV chemicals, focussing on the US Challenge programme. The paper briefly describes each of the endpoints discussed, and their importance and summarises the approaches that may be used within the Challenge programme for developing predictions for the endpoints.

2.3 Human health endpoints

Cronin *et al* (2003b) reviewed the use of (Q)SARs, by governmental regulatory agencies and authorities worldwide, to predict human health effects. Uses of (Q)SARs described are prioritisation of chemicals for testing and filling of data gaps in risk assessment data sets. Whilst there is some application of (Q)SARs by regulatory agencies and authorities worldwide, the most widespread application is by US EPA, Food and Drug Administration (FDA) and other governmental groups in North America. The Agency for Toxic Substances and Disease Registry (ATSDR) and the FDA use (Q)SARs to provide toxicity estimates for chemicals lacking experimental data. Some agencies (e.g. the FDA) are also building electronic toxicity databases. The regulatory use of (Q)SARs in Europe and elsewhere is less widespread and formalised and is generally on a local national level by individual agencies. In Europe, (Q)SAR use has been limited to physico-chemical and environmental endpoints. However, an increasing need for prediction of human health effects endpoints is expected as the consequence of the revision of EU Chemicals Policy. The Danish EPA has drawn up a list for self-classification of dangerous substances using (Q)SAR models. In Germany, SARs are used to evaluate physico-chemical, irritation and corrosive properties of newly notified chemicals. However, the EC TGD existing substance regulation and notification of new substances does not address how and when to use (Q)SARs. ECVAM and ICCVAM are also addressed as the bodies responsible for validation of animal alternatives methods. There is also a description of a (Q)SAR validation study carried out under the auspices of the OECD.

Currently, (Q)SARs are used to predict acute and chronic toxicity, mutagenicity, carcinogenicity, teratogenicity, dermal and ocular irritation and corrosion and skin sensitisation. The most frequently used software programs are MultiCASE, TOPKAT and DEREK.

The role of (Q)SARs in the HPV challenge programme is also addressed. The paper reviews the guidance as well as scope and application for (Q)SARs. The use of (Q)SARs for human health toxicity prediction is judged to be more difficult than for physico-chemical properties and environmental safety, because of the variety in and complexity of endpoints and exposure conditions.

2.4 Use of reliability, applicability and uncertainty methods

As there is no available commercial software to assess reliability, applicability and uncertainty of (Q)SARs, the assessment of how regulatory agencies deal with these critical elements could not be made. Some of the model software give predictions with confidence intervals, but it is unclear how this information is used in decision-making. Most often, regulators take (Q)SARs at face value, without taking into account uncertainty. A critical issue is the lack of assessment of the (Q)SARs applicability domains that leads to misuse of the models. Further discussion of this topic can be found in the proceedings of the Setubal workshop (Cefic, 2002).

3. CURRENT USE OF (Q)SARS IN INDUSTRY

3.1 Physico-chemical properties

Knowledge of the physico-chemical (PC) properties of a substance is critical to human and environmental assessment. Such properties are central inputs in classification and labelling, emission estimation, exposure modelling and toxicity prediction.

The specific PC properties required as inputs for data collection in IUCLID under Council Regulation (EEC) 793/93 on the evaluation and control of the risks of existing substances (EC, 1993) are:

- Melting point;
- boiling point;
- density;
- vapour pressure;
- aqueous solubility;
- octanol-water partition coefficient (log K_{ow});
- surface tension;
- flash point;
- autoflammability;
- explosive properties;
- oxidising properties.

For many of the above endpoints, substance-specific data may be available since these properties are critical considerations in manufacturer and commercial use of the product. Therefore, QSPRs may serve more as a theoretical check on experimental measurements, or as a means to identify data outliers, than as a means to fill data gaps. Large discrepancies between measurement and model (beyond the expected uncertainty bounds of QSPR) alert the user that further scrutiny is needed in reassessing the validity of experimental measurements or in revising QSPR algorithms.

Several software packages provide predictions for multiple PC endpoints of interest (Table 2). Detailed information about each of the software packages is documented in Appendix C.

QSPRs for PC endpoint prediction are usually restricted to the evaluation of discrete, simple organic molecules. Consequently, these methods are generally not applicable to metals, organometallics, polymers or commercial mixtures.

Table 2: Physico-chemical properties estimated by different software packages

Endpoint	EPIWIN	ASTER	SPARC	ACD	PREDICT
Melting point	X				
Boiling point	X	X	X	X	X
Density			X	X	X
Vapour pressure	X	X	X		X
Aqueous solubility	X	X	X	X	
Octanol-water partition coefficient	X	X	X	X	
Henry's law constant	X	X	X		
Acid dissociation constant		X	X	X	
Surface tension				X	X

3.2 Environmental endpoints

Knowledge of the fate and distribution properties of a substance is critical to environmental exposure assessment. The properties often required as inputs for exposure assessment are:

- Organic carbon - water partition coefficients;
- photolysis;
- hydrolysis;
- atmospheric oxidation;
- aerobic and anaerobic biodegradation;
- aquatic bioaccumulation;
- terrestrial bioaccumulation.

Two properties, in particular biodegradation and bioaccumulation, are also central inputs in environmental classification. Possibilities of prediction of fate and distribution endpoints are evaluated and documented in Appendix D.

Knowledge about toxicity is needed to assess effects. Aquatic toxicity is also used in classification. Both SAR and QSAR models for acute aquatic toxicity have been developed. Verhaar *et al* (1992) developed a general rule-based classification system that is based on chemical structure and acute aquatic toxicity to fish that can be applied to discrete nonionic organic chemicals. Only the quality assured freshwater fish databases can be considered as 'gold standard' for (Q)SAR development. These databases were developed for fathead minnow (US EPA, Duluth) and guppy (IRAS, Utrecht, the Netherlands). Therefore, models of acute toxicity to fish are based on these two data sets. Unfortunately, similar 'gold standard' databases for other organisms are lacking, thus hampering further development of (Q)SARs. ECOSAR provides (Q)SARs for a variety of species but in contrast to the other models, the data are not publicly available.

Table 3 summarises software for environmental endpoints prediction. A description of each software, including capabilities and performance, is provided in Appendix C.

Table 3: Software for environmental endpoints prediction

	SRC	MultiCASE	TOPKAT	ECOSAR	ASTER	CATABOL - OASIS
FATE and DISTRIBUTION						
Abiotic degradation						
• Photolysis						
• Hydrolysis	X				X	
• Atmospheric oxidation	X					
Biotic degradation						
• Aerobic	X	X	X		X	X
• Anaerobic		X				
Sorption	X				X	
Bioaccumulation						
• Bioconcentration	X				X	
TOXICITY: Species/environment						
Fish						
• Acute		X	X	X	X*	X
• Chronic				X		
Daphnia						
• Acute			X	X		
• Chronic				X		
Algae						
• Acute/chronic				X		
Terrestrial				X		
Marine				X		

* contains the link to a database

Because of their importance, bioaccumulation, biodegradation and aquatic toxicity predictions were evaluated in greater detail, and the results are summarised in Tables 4-6, respectively.

Table 4: Summary of bioaccumulation prediction software

	BCFWIN	ASTER
Allows modifications	No	No
Availability of the training set	Yes	Yes
Documentation of the prediction method	Very good via journal publications	Good via journal publications
Internal (cross) validation	Yes	No
External validation	Yes	No
Explanation of applicability domain	No explanation	No explanation
User friendliness	Very good	Good
Operating system	Windows	Internet based
Batch mode	Yes	?
Cost	Free	Free
Who supports	US EPA, SRC	US EPA ORD

Table 5: Summary of the available features of reviewed biodegradation prediction software

	BIOWIN	TOPKAT	MultiCASE	CATABOL
Endpoint predicted	Probability of ready biodegradability, range of half lives	Probability of biodegradability	Probability of ready ready biodegradability	Extent with confidence limits in ready biodegradability test[1]
Allows modifications	No	No	Yes	Yes
Availability of the training set	Yes	Yes	Yes but has to be purchased separately	Yes
Documentation of the prediction method	Very good via journal publications	Limited, equations are not available to the user	Good via journal publications	Very good via journal publication + extensive online help and manual
Internal (cross) validation	Yes	Yes	Yes	Yes
External validation	Yes	Not known	Limited	Yes
Explanation of applicability domain	No explanation	Yes, checked OPS[1,2]	Yes	Under development
User friendliness	Very good	No experience	No experience	Very good
Operating system	Windows	Windows	VAX/VMS + Windows	Windows
Batch mode	Yes	Yes	Yes	Yes
Cost	0	~10 k USD/y	~ 20 k USD/y	~20 k USD/y
Who supports	US EPA	www.accelerys.com	CASE Western University	Procter and Gamble and LMC University, Bourgas, Bulgaria

[1] also generates probable aerobic biodegradation pathway

[2] Optimum Prediction Space (US patent, 2001)

There are two programs available, BCFWIN and ASTER, to predict BCF. Both are based on log K_{ow}. These and other popular bioaccumulation models are discussed in detail in Appendix C.

Table 6: Summary of the available (commercial) systems for aquatic toxicity prediction

	ECOSAR	TOPKAT	MultiCASE	OASIS	ASTER
Endpoint predicted	L(E)C_{50} for various organisms, in some cases NOEC[1]	Fish L(E)C_{50}, the prediction is supported by information about molecular descriptors, e-states, fragments in equation, moiety effect and similarity search	Fish LC_{50} and its probability	Fish LC_{50} with confidence intervals	Fish LC_{50}
Allows modifications	No	No	Yes	No	No
Availability of the training set	No	Yes, and training set is searchable for most similar compounds with respect to whole molecule, substructural contributions and activities.	No, has to be purchased separately	Yes	Yes
Documentation of the prediction methods	On-line + publications	Limited, equations are not available to the user, developers offer courses	Via journal publications publications	On-line and publications	Via publications
Internal validation	No	Cross-validated	No	Cross-validated	No
External validation	Limited	Limited	Limited	Limited	Limited
Explanation of applicability domain	No	Yes, checked automatically[2]	Yes with limitations	Yes, checked automatically and visualised	Yes with limitations
Operating system	Windows	Windows	Available on VAX, limited versions under Windows	Windows	Internet based
User friendliness	Very good	Good but no output file provided.	Windows version very good	Very good	Good
Batch mode	Yes	Yes, at ~18 K EUR/y	Yes	Yes	Unknown
Cost	Free	~10 k EUR/y per each module	~ 20 k USD/y	~20 k EUR/y	Free
Who supports	US EPA	www.accelerys.com	CASE Western University	LMC Bulgaria	ORD EPA Duluth office

[1] ECOSAR is chemical class specific and therefore may not contain (Q)SARs for each species for a substance of interest

[2] Optimum Predictive Space (US patent, 2001)

3.3 Human health endpoints

Commercial SAR and (Q)SAR systems are available for a wide variety of toxicological endpoints. The two major approaches for commercial prediction systems are knowledge-based and statistically-based. Knowledge-based systems, such as DEREK, OncoLogic and HazardExpert, use rules about generalised relationships between structure and biological activity that are derived from expert opinion and interpretation of toxicological data to predict the potential toxicity of novel chemicals, i.e. computerised systems that mimic the thinking and reasoning of human experts; statistically-based systems, such as TOPKAT and MultiCASE, derive mathematical relationships for a training set of non-congeneric compounds in an unbiased manner.

A summary of the features of the most important commercial software programs is outlined in Table 7 and a summary of strengths and limitations of the commercial (Q)SAR software for human health endpoints is given in Table 8.

Other systems which also provide limited information on toxicological properties (e.g. PASS – Prediction of Activity Spectra for Substances (www.ibmh.msk.su/PASS) and ASTER) were not included in this review.

Table 7: Comparison of commercial software features (partial list)

Endpoint	**DEREK**	**Hazard Expert**	**MultiCASE**	**TOPKAT**	**TOXSYS**	**OncoLogic**	**DSS**
Carcinogenicity	x	x	x	x		x	
Mutagenicity *in vitro*	x	x	x	x	x		
Acute oral toxicity (rat, LD_{50})				x	x*		
Chronic toxicity (rat, oral)				x			
Developmental toxicity / teratogenicity	x	x	x	x			
Skin sensitisation	x	x	x	x			
Irritancy/ corrosion	x	(x)	x	x			x
Maximum tolerated dose			x	x			
Neurotoxicity	x	x					
Immunotoxicity		x					
Interface							
Model	SAR	SAR	QSAR	QSAR	Neural net	SAR	SAR
Output file	Yes		Yes	No			
Operating system	Windows/Unix		Windows/open vms	Windows	Windows		
Compound input format	Chemdraw, ISIS draw, mol, sd		Smiles, mol	Smiles			Special input format
Batch mode	Yes		Yes	Yes	Yes		No
Website	www.chem.leeds.ac.uk	www.compudrug.com	www.multicase.com	www.accelrys.com	www.scivision.com	www.logichem.com	

*mouse LD_{50}

Table 8: Summary of the software for human health endpoints evaluation

	DEREK	MultiCASE	TOPKAT	DSS
Allows modifications	Yes	Yes	No	Yes
Availability of the training set	Only non-proprietary data are available	No, has to be purchased separately	Yes. Training set is searchable for most similar compounds with respect to whole molecule, substructural contributions. and activities	No, only proprietary data
Documentation of the prediction methods	Compounds and number of compounds used to derive a rule are not generally given	Via journal publications	Limited, equations are not available to the user, developers developers offer courses	Poor
Internal validation	Not applicable	No	Cross-validated	Not known
External validation	Scattered	Scattered	Not known	With about 500 compounds
Explanation of applicability domain	Not applicable	Yes with limitations	Checked automatically with OPS, If the query compound is outside the OPS, the number of dimensions in which the query compound is outside and the respective distance is displayed.	No
User friendliness	Good	Good for Windows version; poor for VMS	Good, but the program does not provide an output file. Printing result reports is uncomfortable.	Limited owing to very complicated structure input.
Cost	~22,000 EUR	~ 20 k USD /y	10 k USD /y per module	Free of charge
Who supports	LHASA	CASE Western University	Accelerys	BgVV

4. CONCLUSIONS

(Q)SARs organise existing knowledge about an endpoint (effect or property) into mathematical forms with the purpose of generalising this knowledge and making inferences about chemicals for which data do not exist.

(Q)SAR predictions are uncertain because (Q)SARs are simplified (mathematical) representations of complex chemical-biological interactions. The uncertainty is greater for statistical models than for mechanistic-based models.

The size and type of information in a training set for the endpoint of interest, including its inherent variability, is an additional factor influencing uncertainty and imposing limitations on the acceptable scope of (Q)SAR use in chemical management and decision-making.

The size and diversity of chemical inventories for which safety data are needed force industry to use a variety of tiered testing approaches, each with the potential for requiring large investments in testing resources and animals. In order to reduce the number of chemicals requiring extensive animal testing and to refine the use of the higher-tiered testing to ensure that maximum information is gained for the resources invested, a targeted and tiered testing approach is warranted. The existing tiered approaches can be streamlined and focused by 'initial evaluation' tiers that use predictive (Q)SARs.

Followed by intensive computational toxicology research, the use of *in silico* approaches has become increasingly popular over recent years. One of the biggest limitations in the development of predictive systems is the lack of reliable and consistent data available to the developers. Much of the data are generated by methods that were mainly developed for classification purposes. In many cases, especially in human health, these data are not suitable for modelling purposes, i.e. for (Q)SAR development.

Several crucial assumptions are made about the data during (Q)SAR model building particularly : homogeneity with respect to type of experimental tests and measured endpoint, homogeneity of species, achievement of a steady state. Such data sets are in short supply.

An equally important limitation is the limited knowledge about mechanisms of action underlying many toxicological observations and endpoints. Thus, the performance of the 'all chemicals' oriented predictive systems is still limited. Context specific models, e.g. models for classes of chemicals acting through the same mechanisms of action, or models for single mechanistic steps, produce more accurate results. However there are no computerised systems giving guidance to the user which (Q)SAR is appropriate to use.

Table 9 summarises the TF's assessments of the applicability of (Q)SARs for the IUCLID endpoints. If the endpoint is not mentioned, it is because no suitable (Q)SAR was identified for that endpoint. From analysis of this table it is clear that in the short term (Q)SARs will find greater acceptability for predicting environmental endpoints than human health endpoints.

Table 9: Applicability of (Q)SARs for IUCLID endpoints

Physico-chemical properties		**Environmental fate and ecotoxicity**		**Human health**	
Melting point	Good	Photodegradation	Limited	Acute oral toxicity	Limited
Boiling point	Good	Hydrolysis	Limited	Skin irritation	Limited
Vapour pressure	Very good	Biodegradation	Good	Eye irritation	Limited
Partition coefficient	Very good	Bioaccumulation	Good	Skin sensitisation	Limited
Water solubility	Very good	Acute toxicity to fish	Good	Chronic mammalian toxicity	Very limited
		Acute toxicity to invertebrates	Limited	Mutagenicity *in vitro*	Limited to good
		Acute toxicity to algae	Limited	Carcinogenicity	Very limited
		Acute toxicity to bacteria	Limited	Teratogenicity	Very limited
		Chronic toxicity	Very limited		

To fulfill their potential, (Q)SARs predictions need to be scientifically valid, reliable and accepted by decision makers.

Given that a (Q)SAR meets all validation and acceptability requirements. and the prediction is made within the applicability domain of that (Q)SAR, it can potentially be used in chemicals' management, particularly prioritisation for product development, and if there is a need for further safety assessment. Validated (Q)SARs are most useful for risk-based priority setting, and as one of the tools in the risk assessment of substances. In the initial phase of the process, both risk-based prioritisation and risk assessment allow for error and relatively large uncertainty. These gradually decrease in the later phases of the process. This is a principle of a tiered risk assessment framework and (Q)SARs fit well into this approach.

The tiered refinement of risk allows for the collection of more accurate exposure and effects data and, if warranted, for further and higher tier animal testing. Decisions on further (animal) testing should consider how critical the impact of the (Q)SAR prediction is on the outcome of the risk assessment and/or the societal and business implications of a classification. The current approach to classification allows less opportunities for refinement, lacking a tiered approach and depending on available information. Consequently, the use of (Q)SARs, in the absence of data, for classification needs to be carefully managed.

However, for certain substances and hazard endpoints, uncertainty associated with the (Q)SAR prediction would not change the outcome of the classification. In these particular cases, the use of an accepted (Q)SAR prediction seems appropriate.

Physico-chemical endpoints

Knowledge regarding the physico-chemical properties of a substance is critical to both human and environmental assessment. Such properties are central inputs in classification, emission estimation, exposure modelling and toxicity prediction. Several user-friendly software programs are available for physico-chemical property estimation. Given the large data sets of experimental data that are available, the QSARs are often suitable for broad chemical classes, and considerable validation has been performed to characterise reliability of predictions. Thus, most physico-chemical properties can be predicted across broad chemical classes. The most reliable predictions are for neutral, nonpolar organic substances. For example, organic-carbon water partition coefficients can be reliably estimated to within an order of magnitude for these classes.

Accordingly, the applicability domain of these QSARs is generally restricted to neutral nonpolar organic substances; For other substance classes, more class-specific QSARs or experimental measurements are often required. However, it is not possible for the user to confirm if a given substance falls within the chemical applicability domain of the existing PC properties predictive models.

Environmental endpoints

Abiotic degradation in the aquatic, marine and terrestrial environment is a critical consideration in defining realistic overall degradation half-lives in these compartments. The availability of QSARs for predicting the abiotic half-lives of organic chemicals in the aquatic/marine/terrestrial environment is currently very limited. In contrast, reliable QSARs, applicable to a wide range of organic chemical classes, are available for estimating abiotic half-lives in the atmosphere.

For biodegradability, current models focus on predicting ready biodegradability owing to availability of data for this regulatory endpoint. BIOWIN, MultiCASE and CATABOL give satisfactory predictions for broad chemical classes that were included in the training sets to separate ready from not-ready biodegradable substances. CATABOL is the only program that explicitly calculates biodegradability extent (expressed as % of Th CO_2), the two other programs classify the chemicals under assessment into various bands of degradation.

There are two programs CATABOL and META, linked to MultiCASE, that simulate biodegradation pathways in an effort to identify potential metabolites. META's technical capabilities are very limited, compared to CATABOL, and were never validated. While promising, CATABOL predictions should be treated as a research tool until biodegradation pathways are further validated. There are no reliable approaches for estimating anaerobic degradation owing to limited data.

The reliability of current QSARs for estimating primary biodegradation half-lives in surface water, soil and sediment compartments, which are key inputs required for multimedia exposure models (e.g. EUSES), has not been critically assessed.

Bioconcentration is a complex endpoint that consists of uptake, distribution and elimination processes in a whole organism. While all the models that are discussed can be used to derive estimates for neutral, nonpolar and non-ionised chemicals that are known to be recalcitrant to metabolism, SRC BCFWIN model is the preferred model because it has been trained on metabolising chemicals as well as those that do not metabolise. Therefore, it is applicable to the broadest range of chemical classes, including substances that are 1) ionisable and 2) undergo biotransformation. Toxicokinetic approaches provide a promising development in modelling approaches, but critically depend on availability of metabolism rates which, due to the extensive efforts involved, will become available for only a limited number of industrial chemicals.

Most of the work on acute toxicity prediction has been focused on freshwater organisms, particularly fish. Limited work has been done on modelling effects on invertebrates (daphnia) and very little on algae. Nonionic chemicals with a narcotic mode of action can be predicted reliably with relatively simple QSARs (based on log K_{ow}) for fish, invertebrates and algae. QSARs for other modes of action are best developed for fish but still in the developmental stage for other types of species owing to limited availability of data. In general, limited validation has been performed.

Developed SARs used for categorisation of substances by mode of action, provide a guidance to mode-of-action specific QSARs that is a toxicologically preferred modelling approach. Both ASTER and OASIS provide QSAR estimations based on mode of action. The EC TGD (EC, 1996, 2003) contains aquatic toxicity QSARs (based on mode of action) but they are not integrated into the EUSES software. In contrast ECOSAR is a chemical-class specific system. In addition, the other disadvantage of ECOSAR is lack of access to the training sets used for model development, and consequently an inability for users to specify applicability domains. Nevertheless the use of ECOSAR is widespread.

In addition, SARs used for categorisation of substances by mode of action, provide a promising approach for estimating acute to chronic ratios and species-sensitivity distributions, thereby generally improving environmental effect assessments.

Human health endpoints

One of the major limitations in the development of predictive systems for human health endpoints is the limited availability of reliable and consistent data to the developers, often aggravated by limited knowledge about mechanisms of action underlying the toxicological effect. (Q)SARs for chemicals with high structural similarity, common mechanisms of action or single mechanistic steps, show relatively good performance, but the current commercial (Q)SAR systems are not generally applicable for non-congeneric groups of chemicals. The performance of the 'all chemicals' oriented predictive systems is limited. Thus, for the time being, (Q)SAR systems cannot reliably replace experimental testing, though they can provide warnings/alerts about possible toxic properties of the query compounds that need to be further interpreted by experts in the relevant field of human toxicology.

The *in vitro* mutagenicity endpoint, with a relatively large database and for which mechanisms of action are fairly well understood, is considered the most promising endpoint for (Q)SAR modelling. For this endpoint, models are being developed that will probably be suitable as screening tools.

In spite of several, well-established computerised systems such as DEREK, MULTICASE and TOPKAT for the prediction of 'local' human health effects (e.g. irritation, sensitisation) their predictions should still be treated with care. This is because historically, local human health effects were evaluated by e.g. skin grading, which did not provide data for modelling. Hence, improvements in this area should be focused on expansion of databases with quality data suitable for modelling. Current databases for these endpoints include data of highly variable quality.

Concerning systemic effects, application of common modelling techniques that find relationships between parent structure and the endpoint, and the expansion of databases with quality data suitable for modelling, will lead to improvements in predictions for chemical classes where ADME or receptor mediation is not important. However, the remaining chemicals might need to be assessed using more-complex modelling approaches rather than the current statistical approaches. Where ADME and receptor-mediated effects or species differences have to be taken into account, it will be more difficult to develop models, even if large databases become available. The prediction systems for these complex endpoints will require complex modelling approaches, invoking a (Q)SAR model interactively communicating with ADME or receptor models. Although in principle it is possible to build such systems, presently, mechanistically-based modelling approaches may not be generally possible for endpoints such as acute and chronic toxicity, carcinogenicity, reproductive toxicity, sensitisation and irritation as the underlying mechanisms are not (fully) understood. To add to the complexity, there may be multiple mechanisms that can lead to the same toxicological endpoint.

Richard and Benigni (2001) and Greene (2002) recommend that predictive models must be able to accurately delineate classes of active chemicals, and within those classes of chemicals, determine the structural features and properties responsible for modulating activity.

Currently several metabolism simulation programs (MultiCASE: META, DEREK: Meteor, HazardExpert: MetabolExpert) are available that potentially allow better understanding of toxic mechanisms or detoxifying mechanisms. However, they are not always linked to the (Q)SAR systems and there is a need to validate these approaches. In general, ADME processes and receptor modelling are areas of intensive research, but links to (Q)SAR models applicable to broad chemical classes have yet to be established.

The results of studies on the performance of rule-based systems (DEREK, BgVV DSS), show that the prediction capability of these systems is limited due to the modelling approach adopted. Expert systems based on structural alerts generate many false positives, because they can identify alerts but do not take into account any relevant detoxifying/deactivating mechanisms/structures. Alert systems that identify positive compounds are useful for product development. However, they have limited capability to reduce resources for regulatory applications because they cannot identify non-active compounds. This is because the applicability domain of such alert systems is comprised almost entirely of active compounds.

Databases

Data quality is an important aspect to be considered if databases are created or evaluated for (Q)SAR model development. A high quality database could help to develop a better mechanistic understanding and eventually a good predictive model. Quality parameters as suggested in IUCLID are acceptable, but should be generally implemented into toxicity databases to allow selection of suitable data from different sources.

Inclusion of proprietary human and environmental health data in these databases is considered important for future (Q)SAR developments. However, this has to be balanced against the often confidential nature of the chemical structures involved. Currently, there are no acceptable methodologies available that achieve the required balance.

Software

Commercially available software was evaluated against applicability criteria that were developed at the recent workshop in Setubal, Portugal (Cefic, 2002), with special attention to validation and transparency. None of the available commercial software systems, either for environmental or human health endpoints, currently meets all these criteria.

In general, while internal validation is usually done and documented, external validation is limited. Quite often the methods are not transparent, despite publications in peer-reviewed journals. A critical fault in the majority of the software is the lack of applicability domain assessment. When available, this would minimise the potential for the mis-use of the model. Furthermore documentation of the uncertainty of predictions is lacking.

As explained above, expert systems based on structural alerts such as DEREK, DSS BgVV generate many false positives. Therefore while they are useful for product development they have limited impact on the need to reduce resources for regulatory applications because with these tools non-active compounds cannot be identified and eliminated from the need for further testing. In contrast modelling approaches in which the applicability domain consist of both active and non-active compounds will be more useful in chemical management because of their greater potential to reduce the need for testing.

5. RECOMMENDATIONS

There is a need to develop confidence in the use of (Q)SARs. If this is to be achieved, agreement will be needed on the specific criteria for acceptance of (Q)SARs, and a process on how they will be applied. The expectation is that initially, only a few (Q)SARs will meet acceptability criteria and therefore be accepted for use. Developing confidence would therefore appear to be easier to achieve for environmental endpoints than mammalian endpoints.

There is an urgent need to develop a transparent, scientifically-sound framework on the use of (Q)SARs that will be recognised and respected by all users and developers of (Q)SARs. The framework should be based on harmonised and widely-recognised acceptability criteria including:

- (Q)SAR applicability domain assessment;
- (Q)SAR validation and reliability assessment;
- (Q)SAR prediction uncertainty characterisation.

A framework for promoting acceptance and regulatory use of (Q)SARs has recently been proposed by Worth *et al* (2003).

The Task Force fully supports the conclusions of the Setubal Workshop (Cefic, 2002) on regulatory acceptance of (Q)SARs and specifically proposed acceptability criteria for QSARs to be used in chemical management. The proposed criteria for validating predictive models are currently more focused on continuous and categorisation QSAR models. Thus, there is a need to further develop SAR and analogue identification acceptability criteria.

There is a need to compile and test existing (Q)SARs against the proposed criteria for use with the various regulatory applications, i.e. priority setting, risk assessment and hazard classification and labelling. For the environmental fate and effects endpoints it is recommended to build on efforts from an 1992 EU project (EU Contract EV5V-CT92-0211).

(Q)SAR predictions may be in the form of statements that a chemical is 'active' or 'not active', or the chemicals are categorised using regulatory cut-offs. These predictions are based on acceptance of certain sensitivity and specificity levels (false positive and false negative rates). It is therefore recommended that if a (Q)SAR prediction is within the accepted error tolerance region then the prediction can be used for determining activity or lack of it. However, if the prediction is outside the accepted error tolerance region, then generation of experimental data should be considered.

There is a need to develop (Q)SAR based approaches applicable for broad chemical classes to improve the safe handling and use of chemicals and to address the growing human and environmental health information needs. If necessary each endpoint may have several, more-restricted (Q)SAR models combined in one interface for the user, where the most appropriate (Q)SAR is chosen by an electronic expert system. Such a system could also include a rigorous framework for regulatory use of (Q)SARs, including support of the system and maintenance of the underlying data and reports by an independent body. Any QSAR Decision Support System (QDSS) developed should have an open architecture which allows addition of modules over time.

The databases for the (Q)SAR training and validation sets have to be considerably enlarged to allow development of generally applicable prediction systems. (Q)SARs considered for chemical management applications need to be carefully validated and their chemical applicability domain carefully defined. A QDSS needs to include automated checks of the chemical applicability domain against the chemical structure inputs for which a prediction is required. In addition, uncertainty of the predictions needs to be characterised as a function of the underlying variability in the data, variability in the endpoint, or the uncertainty in the model parameters.

Specific recommendations for physico-chemical endpoints

Physico-chemical properties models are, in general, well developed and in many cases provide reliable estimations. An important improvement in these models would be to include an assessment of the chemical applicability domain.

Specific recommendations for environmental fate and effects endpoints

Sorption: For sorption to organic carbon it is recommended that there be further development of class specific (Q)SARs that go beyond neutral, nonpolar organic substances, initially by further developing the experimental database.

Biodegradation: In order to extend the use of (Q)SARs beyond categorisation as readily/not readily biodegradable and in risk assessments, without using very conservative extrapolation factors, there is a need to develop and validate approaches for estimation of half-lives in all environmental compartments. The highest priority should be given to the aquatic compartment. The need involves development of methods to combine primary abiotic and biotic degradation rates. It is recognised that lack of data will delay progress in this area. There is also a need to develop models for anaerobic degradation. Again, it is recognised that the absence of appropriate test methods and test data will delay (Q)SAR progress in this area.

Bioconcentration: It is critical to take metabolism into account for a proper evaluation of bioconcentration potential. Therefore, there is a need to ensure that training sets of BCF models contain chemicals that are known to undergo metabolism. Further, there is a need for consistent use of parent BCF values that do not include measurements based on total radioactivity. Sources of uncertainty in bioconcentration factors for highly hydrophobic compounds need to be evaluated. Special attention should be given to exposure issues that could be potentially overcome by applying, for example, a dietary test protocol. Further refinements can be achieved by considering species variation in lipid content and development of species-specific QSARs including species-specific metabolic rates.

Aquatic toxicity: High quality experimental databases for species other than fish, as well as for chronic toxicity endpoints, will have to be collated to allow new (Q)SAR developments. Software is required, further to promote mode-of-action based categorisation of substances. This should also facilitate investigation of the applicability of this concept for human health acute toxicity endpoints once good quality data sets become available.

Human health endpoints

It is critical that currently available models are used only by expert toxicologists who must examine objectively the estimates, also considering the overall weight of evidence for the available toxicologic information of the query chemical. Expert judgement is imperative for the evaluation of predictions.

Further refinement of expert rules of current commercial knowledge-based systems is necessary to increase their performance and the reliability of predictions.

The underlying databases, for both the knowledge-based and statistically-based systems, have to be considerably enlarged to facilitate development of generally applicable prediction systems.

Currently there is insufficient information regarding mechanisms of action for many effects. The rapid development of toxicogenomics and other similar technologies, which have the potential to generate huge databases of chemical exposures and triggered molecular activity, should be investigated. The link between (Q)SAR technology and 'omics' development should be encouraged within a short time frame as broad ranging exploratory studies are being planned for toxicogenomic screening. However, the evaluation of the results of these studies should be considered with great care (ECETOC, 2001).

The need for better predictions for human health endpoints should focus on the expansion of databases suitable for modelling. In particular, attention should be directed towards the use of appropriate assays, including data quality assessments. Areas of priority are sensitisation (skin and respiratory), dermal adsorption and repeated dose toxicity. The current testing methods should be evaluated against the goal of improving general knowledge and the possibility of making inferences about the properties, fate and effects of chemicals for which experimental data do not exist. To make progress in this area it is considered important to make sharing of at least some proprietary human and environmental health data a possibility. Thus, there is a need to find an approach that balances the sharing of health data but which keeps chemical structure information confidential whilst providing possibilities for (Q)SAR development.

Development of broadly applicable (Q)SARs for complex endpoints e.g. teratogenicity, chronic toxicity, carcinogenicity, sensitisation is considered difficult to achieve with current modelling techniques. Mechanistic modelling approaches such as ADME or receptor mediation should be included in development of predictive systems for these endpoints.

Data

Predictive capabilities of (Q)SARs will benefit from compilation and centralisation of the existing quality data and knowledge. In future it will be helpful to model development for optimal experimental design techniques for testing to be employed. Such techniques will help minimise testing, while potentially expanding the applicability domain of models.

The current testing methods should be evaluated to ascertain how data, which would improve general knowledge about the properties of chemicals, could be generated. Such data would aid the development of models and is usually of the type required for risk assessment. An added bonus would be the potential also to use these data for classification purposes.

Software

Increased use of (Q)SAR based tools is contingent upon the ease of use of existing approaches and model approaches. Special care should thus be given to ease of input/output. The software should be user friendly, including facilitated structure input. For example, many systems use SMILES notations for data input and therefore, integration of software that allows conversion of CAS to SMILES will foster ease of data input. Desired output files should allow for formatting harmonised with Excel spreadsheets. Batch processing is another desirable feature to increase efficiency of (Q)SAR use.

Systems to predict environmental endpoints are usually internally validated and, where the data allows, externally validated. In contrast, and in part due to availability of data, none of the systems to predict human health endpoints is externally validated. They therefore fail to fulfil the recommendations of the Setubal workshop (Cefic, 2002) with respect to validation and transparency. The systems currently do not account for metabolism and take into account only 2-dimensional structures and fragments that may be insufficient for modelling some mechanisms.

Existing commercial systems bear high costs that limit their use to large companies and regulatory agencies. It is recommended that industry funding, combined with other funding sources (e.g. EU 6th Framework Programme), be considered as a way of meeting the resourcing needs necessary for the development of freely available (Q)SAR based decision support systems.

6. USE OF CURRENT DATABASES FOR (Q)SAR DEVELOPMENT

Development of 'global' (Q)SAR models for general prediction of a toxicity endpoint, spanning diverse chemicals and mechanisms, but also the development of class-specific or mechanism-based SARs within smaller groups of congeneric molecules, requires a broad-based effort to gather toxicity data from wide ranging information sources. The ability to assess potential toxicity of chemicals based on structure information is a shared goal of academic, commercial and government regulatory groups. They all have a need for unrestricted access to toxicity data linked with structure information. Currently, there exists no central repository of toxicology information, commercial or public, that meets adequately the data requirements for flexible analogue searching, increased reliability and speed for SAR/QSAR model development. A number of activities to develop toxicity databases are outlined in the following paragraphs.

It is generally recognised that (Q)SAR models should be developed based on sound, valid experimental data, because the accuracy and precision of the (Q)SAR will depend on the quality of the data, and the range of structures used to derive the (Q)SAR (Greene, 2002; Hulzebos *et al*, 1999). In March 2002, syndicate sessions at the Setubal workshop on regulatory acceptance of (Q)SARs for human health and environmental endpoints, highlighted the fact that (Q)SAR models, developed for certain endpoints, are currently limited by a lack of good quality databases (Cefic, 2002).

6.1 Data quality

The main pre-requisite for the development of reliable (Q)SARs is the availability of a suitable data set: large enough and of sufficient high quality. Almost all initiatives address data quantity, but only limited projects address fully the quality aspect i.e. the Fraunhofer Repeat Dose Toxicity Database. The IUCLID database offers the possibility for quality identification (including e.g. reliability criteria, good laboratory practice, test methods identification), but use of these parameters is limited and needs to be extended for this database to be useful for (Q)SAR model development. The International Toxicity Information Centre (ITIC), which is based on IUCLID, was an attempt to address gaps in the IUCLID quality identification, but it has not yet been implemented in the pilot database. With the Distributed Structure-Searchable Toxicity (DSSTox) network, it is left to the individual user to sort data for quality aspects. At present, the commercial databases are based primarily on the extensive Registry of Toxic Effects of Chemical Substances (RTECS) repository of published data. The major disadvantage of RTECS is that it only includes reports of positive toxicity data and it is lacking any measure of quality review. To address this issue, LeadScope started working with a focus group to 1) establish common protocols for toxicity databases; 2) develop a grading scheme for toxic endpoints using LeadScope's method; 3) build a small database for data mining; 4) apply the grading scheme to the database to derive toxic alerts and 5) test and validate the alerts.

In summary, data quality is an important aspect, in particular for human health endpoints, which should be considered if databases are created or evaluated for (Q)SAR model development. The generation of (Q)SARs for human health endpoints is often limited by a lack of mechanistic understanding. A high quality database could potentially help to develop a better mechanistic understanding for certain human health endpoints. Quality parameters as suggested in IUCLID are acceptable, but they should be generally implemented into toxicity databases to allow selection of suitable data.

6.2 Data sharing

Pharmaceutical and chemical industries have been leading the use of information technologies and data mining tools for large internal databases that contain public and proprietary data. While there is an overlap of public data in all databases, there is usually little overlap of proprietary data. The ability to share and consolidate toxicity information, not only public data, but also some consolidation of proprietary chemicals with public toxicity information, will allow to maximise effectiveness of data mining and (Q)SAR model construction efforts.

A limited release of proprietary information is expected following the implementation of the EC White Paper on Chemicals Policy. However, the extent to which proprietary data will improve models, might depend on chemical classes or certain endpoints. Proprietary data are often concentrated on a limited range of chemicals that are relevant to a speciality of a company. Often, industrial research and development is chemical class specific and it might thus enhance science, but not necessarily improve 'global' (Q)SARs. However, the use of proprietary data can be helpful for (Q)SAR acceptance, as e.g. (Q)SARs could be built based on public data and validated with proprietary data.

In summary, release of at least some proprietary human and environmental health data is considered important for (Q)SAR development. However, the release of data has to be balanced against the often confidential nature of the chemical structures involved. There is a need to evaluate how this issue could be addressed successfully.

7. RECOMMENDATIONS PERTAINING TO SCOPE FOR (Q)SAR USE

(Q)SAR predictions have the potential to save time and money and importantly, to minimise the use of animal testing. However, to fulfil this potential, the predictions need to be scientifically valid, reliable and accepted by all stakeholders.

Any application of (Q)SARs in regulatory decision-making, should be validated and meet agreed acceptability criteria. Several such criteria have been developed in the past (for example OECD) but because of their breadth and generality they have not been widely applied and respected either by regulatory agencies or by private software companies. The ICCA (Setubal) Workshop on Regulatory Acceptance of (Q)SARs (Cefic, 2002; Jaworska *et al* 2003), initiated the process to develop more specific guidance and applicability criteria for (Q)SAR use, as well as a mechanism for ensuring that these criteria are actually met before a (Q)SAR is used in the management of chemicals.

Given that a (Q)SAR model meets all validation and acceptability requirements, and the prediction is made within the chemical applicability domain of that (Q)SAR, it can be used for all phases of chemicals' management: prioritisation, risk assessment and classification.

First and foremost, (Q)SARs that have been thus validated are most useful for risk-based priority setting. In such applications, there is sufficient scope for tolerating inherent error and uncertainty of the model, as the purpose of the exercise is not to assess everything, but to identify those scenarios with the highest priority for assessment, triggered by specific concerns on the basis of structural alerts, analogue information etc.

Another important potential application for appropriate, validated (Q)SARs, is as one of the tools in risk assessment. Applied as part of a tiered approach, and complemented with other relevant data, such models could inform decisions for generating additional hazard or, indeed, exposure data that would be appropriate for chemicals under evaluation. Effective use of (Q)SARs in this way would doubtless contribute to economies in the scope and number of experimental studies undertaken, e.g. by indicating the most appropriate test species. In some cases, validated (Q)SARs could also be useful in classification of chemicals according to their hazardous properties. This would particularly apply for those substances and endpoints where the inherent uncertainty of the (Q)SAR prediction was unlikely to influence the category of the proposed classification.

In summary, though there are many potential benefits to be gained from the appropriate use of (Q)SARs, there has been limited use made of such models within regulations. More widespread application and further development of (Q)SAR models depends on the establishment of appropriate criteria governing the validity and reliability of their application, their acceptance by all stakeholders and the adherence of all developers and users to these bounds.

The inherent uncertainties of (Q)SARs set the boundaries for their use in chemicals management. Thus, the applications most likely to be accepted amongst stakeholders are risk-based prioritisation and risk assessment, as both processes allow for error and relatively large uncertainty in their initial phases. Only where there was the need to refine the assessment would additional data be generated. Such an approach is fundamental to tiered risk assessment and (Q)SARs fit well into such methodology.

There remains a need to develop a transparent, scientifically-sound framework on the use of (Q)SARs that will be recognised and respected by developers and users. This framework should be based on harmonised and widely-supported acceptability criteria that ensure:

- The domain for which the QSAR is validated and applicable is specified;
- the uncertainty of the QSAR prediction is quantified;
- the predictability/reliability is specified.

The level of uncertainty that is acceptable in a prediction will depend on the type of decision being made. The smaller the change in prediction value that would affect a decision, the more certain the prediction should be.

There is wide consensus that validation and acceptability requirements for application of (Q)SARs are far from being resolved. Further effort must be invested in defining the situations where a (Q)SAR may be predictive and those that are inappropriate. This may be achieved by studies in which a (Q)SAR is applied to specifically selected, well-studied compounds.

8. RESEARCH RECOMMENDATIONS

8.1 Framework for (Q)SARs use and QDSS

There is no rigorous framework for use of (Q)SAR in a regulatory situation. Such a framework is important as it supports the users, both regulatory and industry, in their decision making. It should contain the following key elements:

- Validated (Q)SAR models that meet agreed acceptability criteria;
- transparent databases with flexible search engines that, at least, contain the training set data;
- approaches for identification of structural analogues and grouping of chemicals.

The framework can be translated into an electronic (Q)SAR decision support system (QDSS) to enable a general use across industry and beyond. Attributes of such a QDSS are:

- User-friendliness in its operation;
- decision support tools incorporated for appropriate (Q)SAR selection;
- able to go a long way towards guiding non-(Q)SAR experts;
- generally available;
- accessible through the internet;
- dynamic, i.e. allowing for the continuous refinement of existing (Q)SARs[a].

The QDSS must be supported and maintained by an independent body that can hold proprietary test reports, validate data and models and can provide training on the system where needed.

Based on conclusions of the multi-stakeholder representation by international experts who were present at the LRI workshop in Setubal, Portugal (Cefic, 2002), the following general research areas and needs can be identified:

Data/endpoint assessment

Work is needed to increase the availability of good quality data where possible. Several potential research activities stem from this requirement:

- Database generation, with validated toxicity and fate data used for building the (Q)SARs or for validation purposes (i.e. training set) in order to provide full transparency on the methods included. Areas of priority are (respiratory) sensitisation (appropriate assay and data), dermal adsorption (correlation between *in vitro* and *in vivo*), repeated dose toxicity and expansion of environmentally-relevant test species/systems;

[a] It is expected that, when the acceptability criteria agreed during the Setubal LRI workshop are properly applied to a (Q)SAR, there is not a major impact on chemical specific predictions with consequent model improvements. Rather, the models would improve through expansion of the applicability domain, and thus expand to include predictions for structures not previously covered by the (Q)SAR. In most cases, the predictions for "old" structures would not be significantly altered.

- development of data exchange tools that can maintain structural information that is proprietary, whilst maintaining transparency. This will further enhance the capabilities for analogue assessments;
- variability analysis of an endpoint/value is critical to acceptance of (Q)SAR output variability.

For complex *in vivo* human health endpoints, the existing regulatory endpoints and methods need to be:

- Evaluated for the (implicit) processes that can lead to such an effect and subsequently evaluated for their potential to generate additional data that are more prone to modelling. Such a separation of events should lead to more rational testing designs and increased capacity for priority setting. An example of such a complex endpoint is endocrine disruption;
- compared to outcomes from toxicogenomics and other similar technologies testing towards identifying the potential of these approaches for (Q)SAR application in a regulatory setting.

Validation criteria

- There is a need to develop further SAR and analogue identification acceptability criteria. The proposed criteria for validating predictive models are currently more focused on continuous and classification (Q)SAR models.
- Compile and test existing (Q)SARs against the proposed criteria for use with the various applications, i.e. priority setting, risk assessment and classification, building on efforts from the EU project that addressed several environmental fate and effects QSARs only (EU Contract EV5V-CT92-0211).

Acceptability and confidence

Screening of endpoint models against the acceptability criteria and incorporation into the QDSS.

Follow up on the US EPA/ECB project within the context of OECD (1994) to increase confidence. The original project showed that, even though there was poor predictability from some QSARs, overall the toxic chemicals were picked out. It was suggested that a similar exercise be carried out on the HPV data that will be generated over the next few years.

8.2 Proof of concept of the QDSS

As a proof of concept, and before embarking on initiatives to incorporate more endpoints, industry should develop, or support the development of, a QDSS that will address priority setting issues and the use of (Q)SARs in two main areas, namely aquatic toxicity and endocrine disruption.

8.2.1 Endocrine disruption

US EPA Office of Research and Development recently embarked on a demonstration effort that will develop and apply (Q)SAR technologies to the problems of prioritising and screening endocrine disrupting chemicals. Their research approach is designed to implement a program that will provide, in the near-term, predictive computer-based structure activity models (and *in vitro* assays) that will identify those compounds most likely to disrupt endocrine systems. The aim is to develop within the next one to two years, computer models that predict from chemical structure, the potential for a compound to elicit the initial chemical-biological interactions that trigger the best-understood receptor-mediated events that, lead ultimately to adverse reproductive or developmental outcomes. The computer models are developed and evaluated in an iterative process, and improved by employing data sets derived from *in vitro* assays that are responsive to, and diagnostic of, specific toxicological pathways (e.g. binding affinities to the endocrine receptor, androgen receptor, thyroid hormone). Hence, development and refinement of such assays, which also serve to establish inherent potency measures, accompany the efforts to develop pathway-specific predictive models. Application of these computer models to large chemical inventories, will help identify compounds with the greatest likelihood of disrupting endocrine systems through known, modelled pathways, establishing a rationale for prioritising and ordering appropriate *in vitro* assessments, and ultimately reducing the number of compounds requiring *in vivo* assessments.

8.2.2 Environmental endpoints

In general, aquatic toxicity (Q)SARs are founded on relatively large quantitative databases with a reasonable mechanistic understanding. The (Q)SAR chapter in the EC TGD on risk assessment of new and existing chemicals (EC, 1996, 2003) refers to several such (Q)SARs for acute and chronic aquatic toxicity for the freshwater environment. However, application of these (Q)SARs has been limited and has focused mainly on use in the assessment of the validity of test results for substances which are difficult to test.

A (Q)SAR validation research project in the environmental fate and toxicity area was supported by the European Commission. The aim of the project was the evaluation of existing (Q)SARs according to agreed validity criteria. The chapter of the report on (Q)SARs for ecotoxicity may provide a good starting point for the identification of the QSARs to be used in the QDSS. Other sections of this report can become the starting point for further evaluation and identification of other environmental fate and effects QSARs.

8.3 Research needs that can be addressed immediately

In the absence of a QDSS it is important to increase the ease of use of existing approaches and model approaches. Many of these systems use SMILES notations for data input. Hence, a simple and stand alone software tool that allows conversion of CAS No. to SMILES is required for consistent data inputs. This tool should allow batch processing. An additional feature of the tool could be the capacity to generate SMILES based upon a drawing of the structure in the program. This tool should become incorporated as a module of the QDSS at a later stage.

'Gold standard' databases with quality assured experimental data for aquatic species other than fish, (notably daphnia and algae), need development to allow new (Q)SAR developments. Also, the existing fish toxicity database should be expanded to include chemicals that have a mode of action other than polar or nonpolar narcosis, to allow more reliable QSARs to be developed in these areas (Moore, 2000).

Applicability of the mode of action categorisation concept should be further investigated in the human health endpoints area. A first step is to develop quality assured data sets for several human health acute toxicity endpoints. If the categorisation approach is applicable to human health endpoints, it might be possible also, to, determine mode-of-action specific acute to chronic ratios.

Other approaches and tools to assess the applicability domain of a (Q)SAR will also become an important element of the QDSS. Criteria for analogue definition and chemical grouping approaches can be structurally-based (e.g. a closely related member of the same chemical class), property-based (e.g. properties fall within the range of the properties of the training set), or toxicologically-based (e.g. Verhaar categorisation scheme (Verhaar *et al*, 1992)). Structurally-based analogue definition will probably require substructure searching capabilities in a database, further augmented with similarity analysis approaches. Immediate work is required to define these approaches and to codify them properly for implementation into the DSS.

APPENDIX A: EXECUTIVE SUMMARY OF THE LRI WORKSHOP ON REGULATORY ACCEPTANCE OF QSARS

Participants of the Setubal Workshop (Cefic, 2002) on the use of (Q)SARs for regulatory purposes had both a diverse background in human and environmental safety and in associations to academic institutions, government bodies or industry from Europe, North America and Japan. They agreed, that the workshop initiated great potential for the further development of predictive models and their application for chemicals management including priority setting, risk assessment and classification and labelling. The workshop participants urged careful consideration of the way forward to ensure future successful growth, development and wider application of such models as one of the components of chemicals' safety assessments.

One of the key messages during the workshop was that both industry and regulatory authorities share the same goal, i.e. to use (Q)SARs in a much broader scope than currently for safety evaluation and chemicals management. Consequently, there was a clear agreement on the need to continue dialogue and co-operation.

(Q)SARs are simplified mathematical representations of complex chemical-biological interactions. They can be divided into two major types, QSARs and SARs. QSARs are all quantitative models yielding a continuous or categorical result. The most common techniques for developing QSARs are regression analysis, neural nets and classification methods. Examples of regression analysis include ordinary least squares and partial least squares. Examples of classification methods are discriminant analysis, decision trees and distance based methods of similarity analysis. SARs are qualitative relationships in the form of structural alerts that incorporate molecular substructures or fragments related to the presence or absence of activity.

(Q)SAR predictions have the potential to save time, money and minimise the use of animal testing. However, to fulfil this potential the predictions, especially those considered for regulatory decision-making, need to be scientifically valid, appropriate for the purpose intended, reliable and accepted by decision-makers. Approaches to determine the acceptability of (Q)SARs have been developed in the past (for example OECD guidance) but because of their breadth and generality they have not been widely applied or respected by either (Q)SAR users or developers. As a consequence, decision making with the help of existing models must be done with care and considerable knowledge. The workshop in Setubal aimed at re-opening the debate to develop more specific guidance and acceptability criteria and a system that would support the use of (Q)SARs such that the guidance and acceptability criteria were actually applied when a (Q)SAR was used for chemical management purposes.

Acceptability criteria for (Q)SARs

(Q)SAR predictions are derived from simplified mathematical representations of complex chemical-biological interactions and consequently are potentially more uncertain than the underlying test data. This imposes limitations on the acceptable scope of a (Q)SAR use in chemical management and decision-making. The general acceptability criteria developed for alternative methods to animal testing by ECVAM (European Centre for the Validation of Alternative Methods) were discussed in the context of (Q)SARs and were fully accepted. These criteria indicate that an alternative model should:

- Be associated with a defined endpoint which it serves to predict;
- take the form of an unambiguous and easily applicable algorithm for predicting a pharmaco-toxicological endpoint;
- ideally have a clear mechanistic basis;
- be accompanied by a definition of the domain of its applicability, e.g. the physico-chemical classes of chemicals for which it is applicable;
- be associated with a measure of its 'goodness-of-fit' and internal 'goodness-of-prediction' estimated with cross-validation or similar method to a training set of data;
- be assessed in terms of its predictive power by using data that were not used in the development of the model (external validation).

The workshop participants agreed that (Q)SARs are one of the alternative methods to animal testing and thus these generic criteria can and should be further refined specifically for (Q)SARs. The acceptability criteria were divided into two components, statistical and non-statistical. The discussions on statistical criteria centred around proposals circulated prior to the meeting, which focused on the regression and two-way classification models. As a result of the quantitative nature of these criteria it was possible to make them very specific.

Specific criteria for continuous (Q)SAR models

One of the clear outcomes was that for regulatory purposes a QSAR's predictive power and the prediction uncertainty must be reported along with the 'goodness-of-fit'. Typical values were recommended and it was agreed these should be used in the subsequent testing of the criteria.

Specific criteria for classification (Q)SARs

Again values describing 'goodness-of-fit' including specificity and sensitivity and predictive power (including negative and positive predictivity) were proposed and accepted. Other factors that need to be evaluated are the minimisation of false positive/negatives by sequential use of models.

Specific criteria for SARs

The issues arising from the need to assess such models were felt to be model specific, but clearly included similarity analysis. It was recognised that further research was needed to address SARs and the application of expert knowledge models.

Non-statistical criteria for (Q)SARs

The non-statistical criteria discussed were associated with the endpoint, chemical descriptors, mechanism, domain and transparency. In particular it was agreed that predictive models should be transparent. Transparency in this context means that there should be access to the training and validation data sets as well as to the methods used for the development and validation of the model. Thus an informed user with the correct tools would be able to re-create the model using the same data and techniques as the original developer.

Another outcome of the discussion about non-statistical criteria was the realisation of the difference between human health and environmental predictive models owing to differences in the nature of the endpoints studied. In general, QSARs for environmental endpoints are founded on relatively large quantitative databases with sufficient mechanistic understanding to enable the model to have useful predictive capability. Furthermore, it is relatively easy to support the prediction with subsequent testing. In contrast, the ability to predict local effects in humans is currently limited by a lack of good quality data and consequently has limited regulatory use. For systemic human health endpoints, the models are poor because the traditional endpoints, e.g. LD_{50}, NOAEL, LOAEL, while suitable for current methods of chemicals' management, are not suitable for (quantitative) modelling. These complex human health endpoints are expressed through many different mechanisms, are often receptor mediated and are multi-stage processes comprising absorption, distribution, metabolism and excretion (ADME) frequently with site specific interactions. Furthermore it was concluded that often the endpoints were not defined by a clear dose response and that steady state concentrations in animals were often not achieved. In the light of these discussions, the workshop felt that further work was needed to increase the availability of good quality data where possible. It was also recognised that before attempting improvements to the predictive models for complex *in vivo* human health endpoints, the existing methods needed to be evaluated for their potential to generate additional data more useful for modelling purposes. It should be noted that the workshop did not, owing to lack of appropriate experts, discuss reproductive toxicity and repeated dose effects.

Future (Q)SAR applications

By developing the acceptability criteria the workshop agreed that considerable progress had been made in refining when and which (Q)SARs can support chemical management decisions. This was recognised as a convergence of industry and regulatory agencies positions regarding scope of (Q)SARs use and an acceptance that both positive and negative assignment of a chemical can be achieved with (Q)SAR-models. It was agreed that:

- Acceptable levels of uncertainty in a prediction will depend upon the chemical management decision being made, i.e. the model should be fit for purpose;
- the smaller a change in the prediction that affects such decision, the more certain that prediction should be. This means that there will need to be a balance between the accuracy of a (Q)SAR prediction and the (Q)SAR applicability domain depending upon the decision to be made;
- uncertainty in a prediction should always be considered in the light of the underlying variability of the experimental data. If the prediction uncertainty matches the inherent variability associated with the endpoint then animal testing should be avoided.

It was thus concluded that if the uncertainty of the prediction was such that a 'wrong' decision might be made, targeted testing could be conducted to confirm the data point. In other circumstances the (Q)SAR prediction would be judged acceptable and no testing warranted.

The workshop participants assumed a scenario with a (Q)SAR that m*et al*l the validation and acceptability requirements previously agreed. It was agreed that such a (Q)SAR could potentially be used for prioritisation, risk assessment and/or classification and labelling. It was felt that (Q)SARs appear most useful for risk-based priority setting and, risk assessment at the lower tiers of the assessment and rational prioritisation for testing and test design. In this way it is possible to evaluate whether the substance would trigger specific concerns on the basis of structural alerts, analogue information or (Q)SAR. The tiered approach in risk assessment allows for the collection of more accurate effects data and if warranted guides further and higher tier animal testing, as well as help in decisions on test species selection. For example, a confirmatory test with one of the most sensitive species could be made instead of simultaneously generating data on all species of a regulatory scheme. If a prediction and an experimental value agree, then further testing could be derogated and (Q)SAR results for the other species used.

The workshop participants recognised that until recently there had been limited regulatory uses of (Q)SARs. The major use had been in the support of chemical assessment and notification in the USA. n the last few years, this had begun to change as both within the EU (Denmark, The Netherlands) and in other areas (e.g. Canada and Japan) programs are being developed that will considerably increase the use of (Q)SARs.

Management of accepted (Q)SARs models

It was acknowledged that there is no rigorous framework for the use of (Q)SAR. The workshop agreed that such a framework is needed as this will support the users, both regulatory and industry, in their decision-making. The system discussed had the following key elements:

- Transparent databases with flexible search engines;
- validated (Q)SAR models that meet agreed acceptability criteria;
- a biotransformation and metabolic simulation model.

It was also recognised that such a system, regardless of its complexity, should be user friendly, incorporate tools to aid the selection of appropriate (Q)SARs and be generally available via the Internet. The workshop agreed that such a decision support framework should attempt to help non-QSAR experts choose the most appropriate models and aid them in a decision-making process.

The decision support system should be dynamic, i.e. allowing for the continuous refinement of existing (Q)SARs. It is expected that when the acceptability criteria are properly applied to a (Q)SAR there will not be a major impact on chemical specific predictions. With consequent model improvements, the predictions for 'old' structures would not be significantly altered. Rather, the models would improve through expansion of the applicability domain and thus predictions for structures not previously covered by the (Q)SAR. It was recognised, however, that this was an area where there needed to be further work to help build confidence by all users in the approaches being advocated.

The decision support system should be supported and maintained by an independent organisation that can hold data (experimental or models) and/or validate data and models. This organisation might also be a potential holder of proprietary data or provide a mechanism for better sharing of data between model developers.

Actions and recommendations

Three review papers were circulated among participants prior to the workshop and provided a common background and helped in the discussions:

- Use of (Q)SARs in international decision-making frameworks to predict ecological effects and environmental fate of chemical substances;
- use of (Q)SARs in international decision-making frameworks to predict human health effects;
- existing methods for reliability and uncertainty assessment and applicability evaluations of regression based and classification (Q)SARs.

These papers have been accepted for publication in *Environmental Health Perspectives* together with the executive summary of the workshop.

The workshop urged industry to take the lead to further develop the acceptability criteria and test existing (Q)SARs against the proposed criteria for use with the various applications, i.e. priority setting, risk assessment and classification. This exercise is seen necessary to build general confidence rapidly in (Q)SAR use for these purposes and especially to address the acceptability of both negative and positive classifications by all interested parties. This work would then be used to further the development of an appropriate decision support system.

This would have the additional benefit of enhancing research within industry to focus on development and selection of alternative chemicals, which was a key component of the proposed US Sustainable Futures programme and the EU White Paper on New Chemical Policy.

Whilst it was recognised that the HPV Challenge programme was unlikely to generate large volumes of new data, the data it will generate should be used to validate existing (Q)SARs add to or develop new (Q)SARs and further test the proposed criteria.

The recommendations of the workshop, results of subsequent testing of the new criteria, a final proposal on acceptability criteria and a DSS should be submitted to the OECD's Special Session on (Q)SAR to be held in November 2002. Linkage with the Special Session is expected to provide an opportunity for a broader policy level discussion of the OECD's current efforts relating (Q)SAR and the possible need for and, nature of, an expansion of these efforts, along the lines recommended by this workshop. Thus involvement of OECD Member countries and other stakeholders in revising and expanding the current OECD guidance on (Q)SARs, along the lines described, would be a very important driver in expanding the role for and, reliance on, (Q)SARs in regulatory decision making. It is envisaged that a key milestone for this effort would be an OECD Workshop in the second half of 2003 the aim to finalise and agree the acceptability criteria and the key elements of the decision support system for (Q)SARs.

APPENDIX B: USEFUL LINKS

General

European Union Technical Guidance Document
http://ecb.jrc.it/existing-chemicals
Screening Information Data Set (SIDS) manual guidance on the use of SAR in the OECD SIDS program
www.oecd.org/ehs/sidsman.htm
OECD QSAR validation study details (EU/US EPA/Japan tripartite comparison study):
http://www.epa.gov/opptintr/MPD-SAR/
HPV challenge program chemical list
http://www.epa.gov/chemrtk/volchall.htm

Environmental

Environment Canada Web-Site:
http://www.ec.gc.ca/substances/ese/eng/esehome.cfm#autop
add DSL link TP
US EPA SAR models:
http://www.epa.gov/oppt/exposure/docs/episuitedl.htm.
US EPA validation studies of OPPT SARs:
http://www.epa.gov/oppt/MPD-SAR/
http://www.epa.gov/MPD-SAR/index.html
http://www.oecd.org/ehs/ehsmono/index.htm
US EPA database of aquatic toxicity values (from literature):
http://www.epa.gov/ecotox/

Human safety

US EPA administrator
http://www.epa.gov\opptintr\itc
(Q)SARs to assess human health effects of HPVCs
http://www.epa.gov/chemrtk/sarfinl1.htm
FDA's Center for drug evaluation and research
http://www.fda.gov/cder
NIOSH Current Intelligence Bulletins
http://www.cdc.gov/niosh

APPENDIX C: PREDICTION OF PHYSICO-CHEMICAL PROPERTIES

C.1 Software evaluated by the Task Force

The **EPIWIN** (Estimations Programs Interface for Windows) suite of models is presently among the most widely used and is available as a free download from the Syracuse Research Corporation. This program provides a convenient user interface to several model algorithms for PC property prediction.

ASTER (ASsessment Tools for the Evaluation of Risk) was developed by the US EPA Mid-Continent Ecology Division, Duluth, MN (MED-Duluth). This software tool represents an integration of the AQUIRE (AQUatic toxicity Information REtrieval) toxic effects database and the QSAR expert system. When experimental data are not available, estimation methods are used for the endpoint.

SPARC (SPARC Performs Automatic Reasoning in Chemistry) developed by Karickhoff *et al* (1991), estimates physical chemical properties using linear free energy relationships and molecular orbital theory. SPARC is supported by the University of Georgia.

PhysChemBatch is a commercially available software package provided by Advanced Chemistry Development (ACD) Laboratories that is used by pharmaceutical companies for property estimation of substances included in large data sets.

PREDICT is another commercially available software program that is marketed by Dragon Technology, Inc. (PREDICT, 2002). This program calculates thermodynamic and transport properties that are of specific interest to chemical engineers.

A comparison of general attributes associated with these software tools is provided in Table C.1. It should be noted that all of the above programs have associated databases of measured physical-chemical property values. Thus for selected substances, model predictions can be compared to experimental data when available. Both ASTER and SPARC are available on-line via the internet while the other programs are designed as stand alone applications for routine use on a personal computer. A unique aspect of PhysChemBatch is that uncertainty estimates are provided with each endpoint prediction.

Table C.1: Comparison of QSPR software for physico-chemical property estimation

Attribute	EPIWIN	ASTER	SPARC	ACD	PREDICT
Transparency	Good	Fair	Poor	Fair	Good
Allows modifications	No	No	No	No	No
Availability of the training set	Yes	No	No	No	No
Documentation of the methods	Yes, user manual	Yes, user manual	Limited	Yes	Yes, user manual
Endpoint validation	Good	Limited	Limited, good for K_{ow}, pKa	Limited, good for K_{ow}	Uncertain
Prediction uncertainty	General guidance	No	No	Yes	General guidance
User input	CAS #, SMILES,	CAS #, SMILES MOL file	SMILES	MOL file, application	Application editor editor
User access	PC program	On-line web version	On-line web version	PC program	PC program
User training	Yes	No	No	Yes	No
Batch mode operation	Yes	No	No	Yes	No
Cost	Free	Free	Free	~5 K	~0.2 K
Who supports	Syracuse Research Corporation www.syrres.com	University of Georgia http://ibmlc2 chem.uga.edu/ sparc/	US EPA www.epa.gov/m ed/databases/as ter.htm	Analytical Chemistry http://www.acdla bs.com	Dragon Technology Inc. http://mwsoft ware.com/ dragon/

Another software package discussed in the ECETOC (1998) QSAR review for PC property estimation is CHEMEST (Technical Database Services (TDS), Inc (http://www.tds-tds.com)). However, TDS no longer appear to support this product. They do however offer an on-line service (NUMERICA) on a fee basis that includes a physical property calculation engine for estimating a variety of thermodynamic properties including melting point, boiling point, density, vapour pressure, aqueous solubility, flash point, autoignition temperature and other explosive properties. This program can also be purchased as stand alone personal computer application for approximately 3000 Euro. Very limited documentation is available for this program and hence it is not discussed further. A future software system that is currently under development by Schüürmann *et al* (1997) is ChemProp. This program estimates a variety of physical chemical properties including boiling point, vapour pressure, water solubility, octanol water partition coefficient, Henry's Law constant based on a suite of QSPR methods. However, owing to limited availability and experience in practical use at present this tool is simply mentioned here for future consideration.

Additional software is available for predicting specific PC endpoints, most notably log K_{ow}. These programs will be mentioned below under the discussions for specific endpoints.

C.2 Melting point

Melting point is defined as the temperature corresponding to one atmosphere of pressure at which the solid and liquid phases of a substance are at equilibrium. Melting point is one of the simplest properties to measure experimentally hence estimation methods are often not required. A review of QSPR methods for melting point estimation is provided by Tesconi and Yalkowsky (2000).

The MPBPWINv1.4 module in EPIWIN estimates melting point by two different methods. The Gold and Ogle (1969) approach and the Joback group contribution method (Joback 1982; Joback and Reid, 1987). A further explanation of these methods is provided in the EPIWIN user manual (US EPA, 1999). Tesconi and Yalkowsky (2000) recommend that the Joback and Reid method offers the greatest general applicability. However owing to theoretical limitations, this approach is less accurate than provided by other methods that are more narrowly applicable to specific substance classes. Moreover, such methods do not appear to be presently coded into widely available software.

The final step in the MPBPWINv1.4 algorithm includes an analysis of the predicted melting points derived from both equations. If the difference is less than 30°K a mean value is reported. If the difference is greater than 30°K, a weighting factor based on molecular structure is applied. This correction has been shown to improve significantly overall prediction accuracy.

Two validation studies with large data sets containing diverse compounds have been reported to assess the accuracy of MPVPWINv1.4 (US EPA, 1999). Summary statistics of model performance obtained in these independent tests were similar:

n=666, r^2= 0.73, sd=59°K, me=45°K
n=1379; r^2= 0.71, sd=58°K, me=44°K

These results provide an indication of the expected accuracy for this QSPR.

C.3 Boiling point

The normal boiling point is defined as the temperature at which the liquid vapour pressure of the substance equals one atmosphere. Often the boiling point can be used to estimate other properties that are more difficult to measure (e.g. vapour pressure) hence like the melting point experimental data are preferred to model predictions. For thermally unstable compounds, experimental measurements at reduced pressure can be used to estimate reliably the (hypothetical) normal boiling point. This approach is included as one of the options included in the PREDICT program for boiling point estimation (PREDICT, 2002.). In the absence of experimental data, estimation methods are available and have been reviewed by ECETOC (1998a) and Lyman (2000). Two general approaches are used: correlation equations using other measured substance properties (e.g. molecular weight, melting point) or group contribution methods.

The MPBPWINv1.4 module included within EPIWIN, estimates boiling point using an adaptation of the Stein and Brown (1994) group contribution method. The performance of this method has been evaluated during training and validation (US EPA, 1999):

Training:

n=4426, sd=24.6°K, me=15.5°K (or 3.2%)

Validation:

n=6584, sd=38.1°K, me=20.4°K (or 4.3%)

ASTER determines the boiling point using the method of Meissner (1949). This method is generally limited to organic compounds containing C, H, N, O, S and halides and is reported to have a mean error of 2% in K and a maximum error of 8% in K within this domain of applicability (ECETOC, 1998).

The estimation method included within PhysChemBatch is not known. However, ACD website states that predictions are normally accurate to within +/- 5°C for organic compounds with fewer than two polar functional groups and within +/- 10°C for structures with more polar groups. The applicability domain is limited to organic compounds containing H, B, C, N, O, F, Cl, Br, I, Si, P, S, Ge, As, Se, Sn, or Pb. Further details on model validation are posted on the website (http://www.acd labs. com/products/phys_chem_lab/bp/vs_experts.html#BoilingPoint).

The PREDICT program also includes two different group contribution methods for boiling point prediction (PREDICT, 2002). Based on available literature studies evaluating these methods relative errors in the range of 3-10% in K are suggested.

Further details on the method used in SPARC for boiling point estimation are discussed by Hilal *et al* (1994). For solids, this program requires that the melting point be provided as input. To our knowledge, model performance has not been reported.

C.4 Density

Density is the mass of a substance within a unit volume at a specified temperature and pressure and is usually expressed in units of g/cm^3. The density of a substance can be an important consideration in elucidating environmental fate in a spill situation since this parameter will indicate the tendency for the substance to sink or float upon release to the aquatic or marine environment. Often experimental measurements for the density of the substance will be available which probably accounts for the general lack of attention devoted to QSPRs for this endpoint in the environmental literature. As shown in Table 1, several programs include methods for density estimation. However, owing to the limited documentation and/or validation of these methods, coupled with the modest role such QSPRs are likely to serve in practice, no further review is provided other than noting that estimation techniques are sometimes included for this endpoint in available software programs.

C.5 Vapour pressure

Vapour pressure is defined as the pressure of a substance that is in equilibrium with its liquid or solid phase at a given temperature. The vapour pressure indicates the tendency for a substance to escape to the air and is a critical input for emission estimation (to air) and exposure modelling (EC, 1996, 2003). Using the ideal gas law, vapour pressure for a substance can be converted in concentration units to provide an indication of the solubility limit in air.

For substances with boiling points below 100°C, experimental data on vapour pressure is often available. However, for higher boiling point substances this property may not be known or only poorly characterised. Reviews of QSPR methods for vapour pressure estimation are provided by ECETOC (1998), Nenzda (1998) and Sage and Sage (2000). The latter authors recommend that the Antoine equation be applied for liquids that boil below 200°C and the Grain-Watson method for higher boiling point liquids as well as solids. A general rule of thumb is that relative error in prediction increases as the vapour pressure of the substance decreases.

MPBPWINv1.4 estimates vapour pressure by three different QSPRs: Antoine, modified Grain-Watson and Mackay methods (US EPA, 1999). The Mackay method has a limited domain of applicability to aliphatic and aromatic hydrocarbons and halogenated compounds. This model calculates vapour pressure using all three methods and provides an arithmetic mean of the Antoine and Grain-Watson models for liquids and gases.

For solids, the modified Grain-Watson model is applied with a correction factor to convert the predicted super-cooled liquid phase vapour pressure to that of a solid. A validation study examining model performance has been reported using a data set of measured vapour pressures for 805 compounds. In this analysis, melting and boiling points that were required as inputs for vapour pressure estimation were calculated using the QSPRs described in Section 3.1 so that the maximum error associated the EPIWIN algorithms could be assessed. Summary statistics on model performance for vapour pressure prediction expressed on a log scale were:

n=805, r^2= 0.941, sd=0.717°K, me=0.476°K

While these results suggest reasonable performance, especially for screening purposes, the observed error can be significantly reduced by using measured rather than calculated melting and boiling points. This alternative calculation option provides the most accurate method for vapour pressure determination and is available by directly inputting experimental values for melting and boiling point for the substance of interest into the data entry screen of MPVPWINv1.4. ASTER also uses the modified Grain-Watson method and thus is expected to give similar results to MPVPWIN.

PhysChemBatch uses the output from the QSPR included in this program for boiling point (Section 3.1) as input for vapour pressure prediction. Consequently, the applicability domain is the same as that described for the boiling point QSPR. Model predictions are compared to experimental vapour pressure data at different temperatures in tabular from for a few compounds (http://www.acdlabs.com/download/publ/ vapor_press.pdf). However, no statistics to characterise model performance are provided.

PREDICT estimates vapour pressure based on four methods, two of which are applicable to nonpolar compounds (Pitzer and Reidel methods) and two which are applicable to polar substances (Gomez-Nieto-Thodos and Halm and Steil methods). Further details on the theory, limitations and accuracy of these algorithms included in PREDICT are provided in the user manual (PREDICT, 2002). General information on vapour pressure estimation by SPARC is provided by Karickhoff *et al* (1991). To our knowledge, performance of the SPARC model has not been evaluated for this endpoint.

With the exception of ASTER, all of the above programs also enable the temperature-dependence of the vapour pressure to be calculated which is an important consideration in environmental fate modelling.

C.6 Aqueous solubility

Aqueous solubility defines the maximum amount of a substance that can freely dissolved in water at a given temperature (typically 25°C). Water solubility is a critical consideration in environmental classification, emission estimation, environmental fate modelling. and aquatic toxicity test interpretation. A number of QSPRs are available for the prediction of aqueous solubility. Reviews on this topic are provided by Lyman *et al* (1990), Yalkowsky and Banerjee (1992), Howard and Meylan (1997); Nenzda (1998), ECETOC (1998) and Mackay (2000). Estimation methods can be divided into three general classes (Mackay, 2000):

- Correlation with the octanol-water partition coefficient;
- group contribution methods or topological indices;
- correlation with other measured or calculated properties (such as HPLC retention time).

Mackay (2000) recommends the correlation with octanol-water partition coefficient method and the AQUAFAC group contribution method for routine estimation of water solubility. However, only the former method is presently available in user-friendly software.

A description of the QSPR (WSKOW) method used for water solubility prediction within EPIWIN is provided by Meylan *et al* (1996). This method includes two equations that involve log K_{ow} as input:

$\log S = 0.796 - 0.854 \log K_{ow} - 0.00728\ MW$ + corrections and

$\log S = 0.693 - 0.96 \log K_{ow} - 0.0092\ (Tm - 25) - 0.00314\ MW$ + corrections

The first is a general equation and the second equation is applicable to solids and is used when the melting point of the compound is available. WSKOW was developed from a data set containing 1,450 compounds of known water solubility, log K_{ow} and melting point. The model uses correction factors based on 15 structural types. The performance of this method for aqueous solubility prediction (expressed on a log scale) has been evaluated during training and validation (Meylan *et al*, 1996):

Training:

(melting point and log K_{ow} measured)

n=1450, r^2= 0.97, sd=0.409, me=0.313

(melting point measured and log K_{ow} estimated)

n=1450, r^2= 0.95, sd=0.510, me=0.378

(melting point estimated and log K_{ow} measured)

n=85, r^2= 0.86, sd=0.961, me=0.714

Validation:

(melting point measured and log K_{ow} estimated)

n=817, r^2= 0.90, sd=0.615, me=0.480

Use of predicted rather than measured values for log K_{ow} and melting point were thus found to reduce the overall accuracy of the estimation method. Thus, whenever possible experimental data on these properties should be used as the basis for aqueous solubility estimation.

The algorithm included in PhysChemBatch (ACD) for water solubility estimation are described by Jouravela (2002). This approach is based on an extension of the general solubility equation proposed by Yalkowsky to broad classes of organic compounds. The use of class-specific correlations involving multiple dependent variables calculated from molecular structure are reported to improve the predicted accuracy of this method and reduce sensitivity to melting point considerations. A validation study of model performance for aqueous solubility prediction expressed on a log scale yielded the following statistics:

n=2094, r^2= 0.87, sd=0.75 (me not provided)

Additional statistics characterising model performance for specific compound classes are posted on the website (http://www.acdlabs.com/products/phys_chem_lab/aqsol/exp.html). A unique feature of this model is that ionic species are included in the applicability domain such that pH-dependence of solubility can be readily assessed.

Limited information is available on aqueous solubility prediction within SPARC other than general principles that are used in the model construct (Hilal *et al* 1994). Letinski *et al* (2002) recently compared experimental slow-stir water solubility measurements for several classes of hydrophobic liquids (alcohols, phthalates, adipates) to aqueous solubility predictions obtained from WSKOW and SPARC. While the number of chemicals examined was very limited (n=11) SPARC predictions were reported to provide better overall agreement with measured values than WSKOW estimates.

In addition to the software packages listed above, five additional programs for estimating water solubility (and log K_{ow}) have been identified (Dearden *et al*, 2002). SLIPPER (http://www.ipac.ac.ru/qsar/prog/4/struct4.htm) estimates both aqueous solubility and log K_{ow} as function of pH using a variety of molecular descriptors. Structures are input to the program either as SDF files or via a program editor. Another estimation program is ABSOLV (http://www.sirius-analytical.com/absolv.htm). Like SPARC, this model uses linear free energy relationships for solubility and log K_{ow} prediction (Abraham and Le, 1999). QikProp 2.0 (http://www.schrodinger.com/Products/qikprop.html) is another software tool used in drug design that estimates both aqueous solubility and log K_{ow} based on 2-D and 3-D molecular descriptors. Another program used in the pharmaceutical field for solubility and log K_{ow} prediction is the C2.ADME module of the Cerius^2 software (http://www.accelrys.com/cerius2/c2adme.html).

Interactive Analysis also markets programs (IALogW and IALogP) for water solubility and log K_{ow} estimation (http://www.logp.com). These models are based on a neural network analysis of molecular connectivity indices (Parham *et al*, 2000) using either SMILES or MOL files as input.

These five programs are all available as stand alone packages for use on a personal computer and can be run in a batch mode for high throughput processing of large substance data sets. However, comparative cost information is not available. An on-line web version of the IALogW program is also available for free.

In a recent study, Dearden *et al* (2002) evaluated model performance of all of the above software packages using a validation set of 113 compounds compiled by Abraham and Le (1999). Results from this exercise are reproduced in Table C.2.

Table C.2: Comparison of QSPR model performance for log solubility (mol/l) prediction

Model	% within 0.5 log unit of data	% within 0.5 1.0 log unit of data	% above 1.0 log unit of data	R^2	SE
IA	75.0	17.0	8.0	0.943	0.487
ACD	72.6	18.6	8.8	0.940	0.498
WSKOW	69.9	18.6	11.5	0.923	0.562
SPARC	68.1	18.6	13.3	0.854	0.779
ABSOLV	61.9	25.7	12.4	0.888	0.680
QuikProp	55.7	31.0	13.3	0.867	0.742
SLIPPER	50.4	23.0	26.6	0.766	0.984
C2.ADME	48.7	31.0	20.3	0.844	0.803

Differences in prediction accuracy for liquids versus solids was not addressed and it was unclear if experimental or measured inputs were used in QSPRs that required log K_{ow} and melting point as model inputs. Nevertheless, this preliminary analysis presents a very useful indication of expected accuracy in aqueous solubility prediction by commercial software.

C.7 Octanol-water partition coefficient

The octanol-water partition coefficient represents the mass equilibrium ratio of a substance between octanol and water at a fixed temperature (typically 25°C). This parameter is used extensively in environmental classification, priority setting, exposure modelling and toxicity prediction. Many QSPRs are available for the prediction of the log K_{ow}. Recent reviews on this topic are provided by Mannhold and Dross (1996); Meylan and *et al* (1996); ECETOC (1998); Petrauskas and Kolovanov (2000); Leo (2000) and Chen *et al* (2002). Estimation methods can be divided into fragment based and atom based. Fragment-based methods are regarded as providing the best prediction accuracy (Mannhold and Dross, 1996).

The KOWWIN module within EPIWIN is a fragment-based approach. This QSPR includes 144 atom/fragment contributions and a further 248 structural correction factors to account for interaction between these fragments. The performance KOWWIN for log K_{ow} prediction is provided in the users-manual:

KOWWIN v1.63

Total: n=12805; r^2=0.95; sd=0.435; me=0.316

Training: n=2474; r^2=0.981; sd=0.22; me=0.16

Validation: n=10331 r^2=0.94; sd=0.47; me=0.35

The estimation algorithm used in the well known CLOGP (Leo, 1993) program is also a group/fragment contribution method, but uses a constructionist approach. Fragments and the fragment interaction correction factors are evaluated from the simplest molecules in which they occur. An evaluation of model performance of CLOGP is provided in the users-manual for KOWWIN:

CLOGP for Windows (v1.0)

Total: n=11735(a); r^2=0.91; sd=0.59; me=0.384

In addition to the software already identified in section 3.16 which estimate not only aqueous solubility but also log K_{ow}, additional programs are KlogP (http://www.multicase.com/), PrologP (http://www.compudrug.com), ALOGPS and XLOGP (Tetko *et al*, 2001; www.lnh.unil.ch/~iteko/logp); AUTOLOGP (Devillers *et al*, 1995a). In addition QSRPs for log K_{ow} prediction are included in the DEREK and TOPKAT toxicity estimation models (see Section 3.3).

Further work comparing the performance of current software models is limited. Leo (2000) reported that ACD/LogP, KOWWIN and CLOGP all performed well on the five benchmark chemicals examined. In a recent study, Dearden *et al* (2002) provided a comparison of the prediction accuracy obtained by ten commercially available programs using the same validation data set used by Mannhold and Dross (1996). Results from this study are reproduced in Table C.3.

Table C.3: Comparison of QSPR model performance for log Kow prediction

Model	% within 0.5 log unit of data	% within 0.5 1.0 log unit of data	% above 1.0 log unit of data	R^2	SE
ACD	93.5	5.8	0.7	0.965	0.271
IA	93.5	5.1	1.4	0.958	0.297
KOWWIN	89.1	7.3	3.6	0.947	0.335
SPARC	88.5	9.0	2.5	0.941	0.330
CLOGP	88.4	10.9	0.7	0.961	0.287
KLOGP	81.9	15.9	2.2	0.929	0.388
ProLogP	81.2	15.2	3.6	0.912	0.431
SLIPPER	81.1	10.9	8.0	0.847	0.568
ABSOLV	67.4	17.4	15.2	0.797	0.654
QuikProp	50.0	34.1	15.9	0.871	0.521

It is important to point out that the above methods generally apply to K_{ow} Prediction for surface active compounds. For this class of compounds, special treatment is required (Roberts, 2002).

C.8 Henry's Law constant

The Henry's Law constant (HLC) indicates the tendency of a substance to partition to air at a specified temperature and is related to the air-water partition coefficient (K_{AW}) by the ideal gas law:

$$K_{AW} = HLC / RT$$

where:

K_{AW} = the mass equilibrium ratio of a substance between air and water at a given temperature

HLC = Henry's Law constant (Pa m^3 mol^{-1})

R = ideal law gas constant = 8.314 Pa m^3 mol^{-1} K^{-1})

T = temperature in degrees K

Recent reviews describing QSPR methods for HLC estimation are provided by Dearden *et al* (1997) and Mackay *et al* (2000). Estimation methods can be divided into the following general approaches:

- Vapour pressure / solubility ratio;
- fragment constant;
- connectivity indices;
- linear solvation energy relationship;
- other (e.g. combination of above approaches, class-specific correlation to molecular properties such as molar volume).

The first approach is generally best applied to experimental estimates of vapour pressure and water solubility, rather than independent QSPR predictions, owing to the potential compounding of estimation errors. This method cannot be applied to substances that are miscible with water or have aqueous solubilities above 1 mol/l. In a recent comparison exercise by Brennan *et al* (1998) and cited by Mackay *et al* (2000), the group contribution and the molecular connectivity index methods yielded the best predictions.

Three of the commercial software packages identified in Table 1 enable HLC estimation. It should be noted that both the algorithms in EPIWIN and SPARC allow HLC to be predicted as a function of temperature that is an important consideration in environmental fate modelling.

The original algorithm reported by Meylan and Howard (1991) uses two fragment-based methods to provide independent estimates of the HLC. An evaluation study of model performance for log HLC prediction yielded the following results:

HENRYWIN v1.0

Training (w/o class specific correction factors): n=345; r^2=0.94 sd=0.45; me=0.30

Training (with class specific correction factors): n=345; r^2=0.97 sd=0.34; me=0.21

Validation: (with class specific correction factors) n=74; r^2=0.96 sd=0.46; me=0.31

This algorithm has since been updated by HERYWIN 3.0 to include additional fragment constants and class-specific correction factors. This version also includes an experimental HLC database for 1650 compounds and enables the user to estimate the HLC of one compound based on experimental data for a structurally related analogue. An evaluation of v3.0 model performance has not been reported but directionally the updated algorithm is expected to be more reliable than previous versions.

SPARC estimates HLC using linear energy solvation relationships. It is unclear from the ASTER manual what approach is used in this package. No information was found on QSPR model performance for HLC estimation with either of these programs.

C.9 Acid dissociation constant (pKa)

The acid dissociation constant expresses the ratio of ionised and unionised forms of a substance in water at equilibrium:

$$KA = [A^-] [H^+] / [HA]$$

Taking logarithms of both sides of the above equation upon rearrangement yields Henderson- Hasselbach equation:

$$pH = pKa - [A] / [HA]$$

where the pH and pKa denote the negative logarithm of the hydrogen ion and acid dissociation constant, respectively. Practically the pKa can thus be interpreted as the pH at which the compound exists equally in dissociated and undissociated forms. This endpoint is estimated by SPARC, ASTER and ACD and is one of the few physical chemical property endpoints not included in EPIWIN.

The principles for pKa estimation in SPARC has been described by Hilal and Karickhoff (1995) and Hilal *et al* (1999). While theoretically complex, this algorithm was found to perform remarkably well as indicated by the following statistics for model prediction:

n=4338; r^2=0.99; sd=0.37

The ACD program estimates pKa using a fragment based approach based on an internal training set of experimental pKas of almost 16,000 compounds. Accuracy of the algorithm used is stated to be better than +/- 0.5 pKa units. This program also allows the user to perform structure and substructure searches of the experimental pKa database. Specific limitations on the applicability domain for pKa estimation are detailed on the ACD website. No information was available describing the approach or performance of the algorithm for pKa estimation included in ASTER.

C.10 Other properties

A review of additional physical chemical endpoints was beyond scope of this present exercise. However, two general types of properties that warrant further investigation are surface tension and explosivity-related properties. These properties are included in IUCLID and may represent important considerations in product risk assessment. While such endpoints are not included at present in commercially available software packages used in the health and environmental sciences, recent QSPR research on these endpoints offer promise for future inclusion in next generation software.

APPENDIX D: PREDICTION OF ENVIRONMENTAL ENDPOINTS

D.1 Software evaluated by the Task Force

SRC EPIWIN

The EPIWIN suite of models is available as a free download from the Syracuse Research Corporation. This program provides a convenient user interface to several model algorithms to fate and distribution prediction.

TOPKAT (Toxicity Prediction by Komputer Assisted Technology)

TOPKAT is a computerised statistical program developed by Enslein and co-workers (Enslein, 1984; Enslein *et al*, 1994). The program employs principles of QSARs to predict aquatic toxicity, biodegradability and various toxicological endpoints. Each endpoint is in a separate module that can be purchased separately. The QSAR equations are either linear multiple regression equations (for continuous toxicity metrics such as LD_{50}, LC_{50} and LOAEL) or two-group linear discriminant functions (for dichotomous endpoints such as mutagenicity, carcinogenicity, teratogenicity). The regression models produce toxicity estimates in weight/weight or weight/volume units, whereas the discriminant models output the probability (between 0 and 1) that the submitted structure is associated with a positive endpoint value.

Each model in the program is associated with a respective database and is developed from experimental toxicity data and numerical descriptors of molecular structures of single organic compounds in the database. For each model, there is a set of structure descriptors which are related to the chosen endpoint and influence the estimate and which are also important in determining if the submitted structure is within the structure space spanned by the database compounds (optimum prediction space OPS). Structural descriptors, e.g. electronic, connectivity, shape and substructural descriptors are selected from a library of about 3000 molecular fragments. The chemicals in the database serve as structural analogues to see similarities between the chemical in question and the library.

Structures are entered into the program by SMILES notation and subsequently, the desired module/endpoint is selected. TOPKAT screens the database, determining, whether the molecule is 'known' and provides the prediction of the toxicity.

MultiCASE (Multiple Computer Automated Structure Evaluation)

MultiCASE (Klopman, 1984, 1992) selects its own descriptors automatically from a 'learning set' composed of active and inactive molecules. The descriptors are readily recognisable single, continuous structural fragments that are embedded in the complete molecule. The descriptors consist of either activation (biophore) or inactivating (biophobe) fragments. Each of these fragments is associated with a confidence level and a probability of activity which is derived from the distribution of these biophores and biophobes among active and inactive molecules.

Upon completion of these analyses, MultiCASE selects the most important of these fragments as a biophore, i.e. the functionality that is responsible for the experimentally observed activity of the molecules that contain it. MultiCASE will then use the molecules containing this biophore as a 'learning set' to identify the chemical properties (i.e. structural fragments) or physical chemical properties (i.e. log P_{ow}, water solubility, quantum mechanical parameters such as HOMO and LUMO) that modulate the activity of the initially identified biophore. This will result in a QSAR equation for this subset of molecules. If the data set is congeneric, then the single biophore and associated modulators may explain the activity of the whole training set. If this is not the case (i.e. a residue of molecules exists which is not explained by the current biophore and its modulators), the program will remove from consideration the molecules already explained by the previous biophore and will search for the next biophore and associated modulators. The process is continued until the activity of all of the molecules of the 'learning set' has been explained.

The resulting list of biophores is then used to predict the activity of yet untested molecules. This process determines whether an unknown molecule contains a biophore. If the molecule has been used for the training set the program will flag this molecule accordingly. If the molecule does not contain a known biophore, it will be predicted to be inactive unless it contains a group that chemically resembles one of the biophores, in which case it will flagged. When the molecule contains a biophore, the presence of the modulators for that biophore will be investigated. MultiCASE will then make qualitative as well as quantitative predictions of the activity of the unknown molecule.

While biophores are the determining structures, the modulators may determine whether and to what extent the biological potential of the chemical is expressed.

The distribution of descriptors among active and inactive molecules is analysed for statistical significance. If the atoms at both ends of the distance descriptor are all the same, including the number of attached hydrogenous, the biophore is designated an 'exact' descriptor. Various atom groupings are also investigated, e.g. hydrogen bond acceptors, hydrogen bond donors and halogens.

MultiCASE is especially appropriate to include proprietary learning data by allowing either addition to a pre-existing (possibly commercial) data set or by replacing it entirely. In some cases, a discrepancy between a predicted negative result and a positive test result may indicate that the test substance had been contaminated by a (strongly) positive by-product.

The program has several helpful features incorporated. Among them it is possible to depict within the molecular structure under examination the underlying biophore which has been used for an activity assignment. Under this setting the system is in an expert mode where the trained investigator can confirm or falsify the significance of the detected substructure according to his own experience.

The program is set up in modules that can be acquired separately. As the program allows the simultaneous prediction of several endpoints, use of a series of databases might be advantageous, depending on their quality and validation.

In trying to interpret the results of the program, it is important to keep in mind the fact that the program makes its predictions on the basis of its knowledge of the molecules used to train it. Any warning indicates a weakness of the support for the predictions. The presumption of inactivity is based on the absence of a recognised biophore. Biases in the 'learning sets' (especially in commercial ones) with respect to the underlying chemistry is prone to either give false negative results or warnings of unknown functionalities, i.e. structural groups not represented in the training database. These have to be taken seriously because of the possibility that the undocumented functionality may lead to misinterpretation.

It is also important to take into consideration the probability of activity and the confidence level. A high confidence level is achieved if the predictions are based on a biophore (fragment) present in many of the molecules of the database. MultiCASE offers the possibility to take metabolism into account by linking with META program.

ECOSAR

(Q)SARs for aquatic toxicity to fish, aquatic invertebrates and algae have been developed and used by US EPA since 1979. These (Q)SARs have been incorporated into a software program (ECOSAR) available free from the US EPA website at www.epa.gov/opptintr/newchms (click on the ECOSAR button). ECOSAR is also now included as a module within the EPIWIN program as discussed above.

ASTER

A brief description of this software tool has already been provided in Section 3.1.

CATABOL

CATABOL is a hybrid of a knowledge-based expert system for predicting biotransformation pathway, working in tandem with a model that calculates probabilities of the individual transformations and overall BOD and/or CO_2 extent (Jaworska *et al*, 2002).

OASIS

OASIS is an expert system for prediction of acute aquatic fish toxicity of noncongeneric chemicals. It is based on the Duluth database of acute toxicities to fathead minnow for about 660 organic chemicals. The program has a build in expert system that chooses the most appropriate QSAR for a queried chemical and shows on a graph how the queried chemical is positioned versus chemicals used in the training set.

D.2 Distribution and fate

D.2.1 Organic carbon - water partition coefficients

Most sorption of hydrophobic organic chemicals in soil, sediment or water is to the natural organic matter present in these media. Partition coefficients between organic carbon (K_{oc}) or organic matter (K_{ow}) and water are used to describe soil and sediment sorption, whereas partition coefficients between dissolved organic carbon (DOC) and water (K_{DOC}) are used to describe sorption in aqueous media.

Numerous measurements of K_{oc}, K_{ow} and K_{DOC} have been reported in the literature and many correlations have been derived using various molecular descriptors the most common of which is K_{ow} (Sabljic and Güsten, 1995; Meylan *et al*, 1992; Gawlik *et al*, 1997). It was not possible to recommend any single correlation as being preferred universally for all compounds. Selection among these correlations is complex and no automated system is currently available. Meylan *et al* (1992) proposed a molecular topology/fragment contribution method for predicting soil sorption coefficients which is included in the EPIWIN software. In ASTER the adsorption coefficient estimation is estimated using the method described by Lyman *et al* (1982). Uncertainty in these relationships is, in part, caused by the variability in structure and composition of particulate or dissolved organic carbon (Seth *et al*, 1999; Burkhard, 2000) which is not accounted for by the hydrophobicity parameter.

Examining the existing K_{oc} and K_{ow} database and the K_{DOC} database for nonionic organic chemicals sorbed to Aldrich humic acid (Burkhard, 2000), it is suggested that experimental difficulties and constraints in measurements further account for the wide variability in reported values. For assessing experimental data, Seth *et al* (1999) have suggested calculating K_{oc}/K_{ow} ratio's and then carefully scrutinising the validity of measured values

above 1.0 and below 0.1. Based on theoretical considerations governing soil sorption and examining the existing database, Seth *et al* (1999) concluded that the preferred approach is to correlate the quantity log (K_{oc}/K_{ow}) with a molecular property or with log K_{ow}. For estimating K_{oc} for hydrophobic chemicals, they propose as a rule of thumb the equation $K_{oc} = 0.35*K_{ow}$. The authors indicate that this prediction is expected to be within a factor of 2.5 of the actual value. Burkhard (2000) examined K_{DOC} data based upon naturally occurring DOC and Aldrich humic acid for organic chemicals. Based on this study a predictive equation for the combined database $K_{DOC} = 0.11\ K_{ow}$ was reported. The prediction was reported to be within a factor of 14 with a 95% confidence. Consideration of only naturally occurring DOC gave a different equation ($K_{DOC} = 0.08\ K_{ow}$) with 95% confidence limits of within a factor of 20.

D.2.2 Abiotic degradation

Aquatic photolysis

A number of photolytic transformation processes are possible in water. Transformations which are caused by light absorption of a given substance are called direct photolysis. Indirect photolysis includes a number of different mechanisms, where it is not the substance of interest which absorbs light, but some other chemical moiety present in the system which may lead to either an increase or decrease in reactivity. These processes involve either energy transfer from an excited species, such as humic acids, to the substance of interest (sensitised photolysis) or chemical reactions of the substance of interest with highly reactive species (e.g. hydroxy-radicals, peroxy-radicals and singlet oxygen) that are formed in the presence of light (ECETOC, 1998). A recent review of photoreactions in surface waters is provided by Mill (2000). Chemical classes known to be susceptible to direct photolysis include polyaromatic hydrocarbons, aromatic amines, phenols, alkyl sulphides, azodyes, nitroalkanes, ketones and aldehydes. Indirect sensitised photolysis is restricted to electrophilic compounds that are also susceptible to direct photolysis. However, indirect photolysis that results from reaction with reactive species such as hydroxyl radical is a relevant abiotic degradation mechanism for many organic chemicals and plays a key role in the natural breakdown of colloidal organic matter. In fact according to Mill (2000) the indirect photolysis rate for almost any compound in freshwater located in the mid-latitudes with more than 1 mg/l nitrate (which upon photolysis generates hydroxyl radical) will be 50 days in summer.

Currently, no general (Q)SAR for predicting either direct or indirect photolysis has been identified within commercially available software, probably due to the complexity of this fate process, coupled with the lack of experimental data and standardised test methodology for this endpoint. However a few class specific methods for direct photolysis have been reported. For example, Peijnenburg *et al* (1992) developed a QSAR for application to mono- and disubstituted meta-halobenzenes and halophenols.

The model uses as predictors, steric factors of the substituents, as well as the bond strength of the C-halogen bond, since cleavage of this bond is the rate-limiting step of the reaction. The applicability of the approach was extended to multiple-substituted halobenzene derivatives (Stegeman *et al*, 1993). In addition to the above specified molecular descriptors, the authors used the inductive constants of the substituents to predict the quantum yield. Mill (2000) further discusses the current state of science in QSAR development for aqueous photolysis prediction.

Hydrolysis

Reactions of a chemical with a water molecule, a hydroxide or a hydronium ion are commonly called hydrolysis. In a hydrolysis reaction, the compounds are transformed into more polar products. These typically have different properties with respect to fate and effects than the parent compound. When considering environmental fate, the increased polarity of hydrolysis products will in general increase their aqueous solubility and therefore decrease the fractions residing in the adsorbed and gas phases.

The few QSARs available for the prediction of hydrolysis tend to be based on limited data sets and cover a narrow range of chemical classes. All models are based on the application of linear free energy relationships and generally use the Hammett or Taft equations or correlations with the pKa of the leaving group. A further review of theory and methods for generating experimental data and developing estimation methods for this endpoint is provided by Wolfe and Jeffers (2000).

The HYDROWIN method (Meylan and Howard, 1993a), (esc.syrres.com) estimates hydrolysis rates for selected chemical classes including esters, carbamates, epoxides, halomethanes and certain alkyl halides. Rate constants based on chemical structure are estimated from regression equations derived from experimental hydrolysis data established for each chemical class and for acid- and base-catalysed hydrolysis. HYDROWIN estimates acid- and base-catalyzed rate constants which are then used to calculate half-lives and selected pHs. Since the neutral hydrolysis rate is not considered, this approach may overestimate the total hydrolysis rate for certain compounds (e.g. epoxides). The equations rely on Hammett and Taft constants, steric factors and, for some equations, on further fragment specific parameters. The method for the estimation of hydrolysis rates in ASTER uses a similar approach (Harris, 1990). No statistical assessment of model performance was available for either HYDROWIN or the ASTER hydrolysis QSAR.

Atmospheric oxidation

Organic substances emitted into the troposphere may be removed by direct photolysis and by chemical reactions with a number of reactive intermediate species, including OH, HO_2 and NO_3 radicals and ozone (Atkinson, 1988). The reaction with photochemically generated hydroxyl radicals during daylight hours is thought to be the dominant atmospheric removal process for the majority of organic substances (Atkinson, 1989).

The rate at which organic compounds are transformed by indirect photodegradation depends on the rate constant of reaction of the compound with the sensitiser (e.g. the hydroxyl radical) and on the concentration of both reactants. Usually the hydroxyl radical concentration is assumed to have an approximately constant value and the reaction is described by pseudo-first-order kinetics. A recent review on this subject is provided by Atkinson (2000).

The AOPWIN program (Meylan and Howard 1993b, 1994), (esc.syrres.com) is used to estimate the atmospheric photooxidation half-life of organic substances. The estimation methods used by AOPWIN are based on the methods developed by Atkinson and co-workers (Atkinson, 1985; Atkinson, 1987, 1988; Kwok and Atkinson, 1995). Atmospheric ozonolysis is also of importance for unsaturated compounds (Atkinson and Carter, 1984) and an estimation of the rate constants for gas-phase reactions between ozone and olefinic/acetylenic compounds is included in the program. The rate constants estimated by the program are then used to calculate atmospheric half-lives for organic compounds based on average atmospheric concentrations of hydroxyl radicals and ozone. Based on a comparison of predicted and experimental rate constants for 647 different chemicals, as presented in the AOPWIN users manual, over 95% percent of the estimated values were within a factor of three of the measured results.

D.2.3 Biodegradation

Biodegradability can be defined as the molecular degradation of a substance, resulting from the complex action of micro-organisms. Biodegradation is one of the most important processes determining the fate of organic chemicals in the environment. Hence, biodegradation rates play an important role in the estimation of the fate of organic chemicals in the environment.

There are a number of different tests designed to measure biodegradation. However, one shortcoming of tests used in regulatory decision-making is that they measure extent of biodegradation, not rates. Many of the tests are carried out for regulatory purposes and give rise to a variety of endpoints. Biodegradation tests commonly used to assess ultimate biodegradability in a regulatory context differ in measured endpoint, source and concentration of inoculum and test substance concentration used.

For example, the ready biodegradability test yields a value (frequently expressed as % biodegradation) and a term, (not-) readily biodegradable. The former indicates the extent to which the substance degraded, while the latter is a legal or regulatory term that indicates whether a chemical passes or fails the OECD ready biodegradability test (OECD 301A-F). The test substance is the only source of carbon and conditions for biodegradation are stringent. The test can produce false negatives with respect to biodegradation potential in the environment.

Many factors reflecting the extreme difference in biodegradation mechanisms to the very specific environmental conditions of each phenomenon affect the biodegradability of a substance in the environment, e.g. structural features such as molecular weight, types of bonds and substitutions (Alexander, 1981; Kelcka, 1985), and environmental factors, including microbial activity and growth as determined by temperature, pH, availability of nutrients, moisture level and residence time of the microbial population in the environmental compartment of interest. Processes such as microbial adaptation and co-metabolism add to the complexity of biodegradation (ECETOC, 2003a).

Lack of uniform endpoints, substrate to biomass ratio and time allowed for acclimation across the tests are responsible for the limited size of available training sets (compared to those used in toxicology) for QSAR development.

Within a specific test, intra- and inter-laboratory variability in endpoint measured add to the difficulties in selecting a training set. Discrepancies and large variability can be observed in the results of specific standard biodegradation tests carried out at different laboratories or within a single laboratory (King and Painter, 1983; Kitano and Takatsuki, 1988). Although biodegradation tests have been standardised by the OECD, a 20-% deviation is considered acceptable when a test procedure is carried out by the same laboratory (OECD, 1993). This has limited extensive development of QSARs for biodegradation and biodegradation kinetics. There are several reviews on biodegradability modelling that discuss in more detail the above points (Jaworska *et al*, 2002).

SRC BIOWIN

The BIOWIN program consists of four modules that predict primary and ultimate biodegradability using linear and nonlinear regression and two expert judgment modules. The approach has been extensively documented in the literature (Boethling and Sabljic, 1989; Boethling *et al*, 1994; Howard *et al*, 1987, 1992). BIOWIN was recently extended to model biodegradability results from tests according to the MITI.

BIOWIN relies on a group contribution approach where each fragment in the training set gets a specific weight. In addition to the fragment's weight, molecular weight of the chemical is used as a descriptor. Over the years the linear and nonlinear BIOWIN probability models have come into fairly widespread use in chemical screening activities. More recently the survey models have been proposed for applications requiring screening-level estimation of biodegradation half-lives, including the US EPA's Waste Minimization Prioritization Tool (WMPT; http://www.epa.gov/epaoswer/hazwaste/minimize/) and multimedia modelling to identify substances with potential PBT (Persistent, Bioaccumulative and Toxic) properties. BIOWIN is part of the US EPA's PBT Profiler (http://www.epa.gov/pbt/toolbox.htm) that has become a freely accessible web-based program. BIOWIN is a simple and very well working model.

MultiCASE

The software is described in Section 3 and Appendix D.

Aerobic biodegradation. META expert system (Klopman *et al*, 1997) simulating aerobic biodegradation pathways works in conjunction with MultiCASE. The META system relies on MultiCASE statistics to set the hierarchy and on mechanistic understanding for development of the rules. The META expert system is built from 70 transformations that match 13 biophores found earlier in MultiCASE analysis (Klopman and Tu, 1997). The hierarchy is based on weights of biophores in the *source* substance as calculated by MultiCASE and ascribing the same weights to associated transformations.

Anaerobic biodegradation. The MultiCASE approach has also been used to model anaerobic aquatic biodegradation rates (Rorije *et al*, 1998). Anaerobic biodegradation data are very scarce in comparison to data for aerobic aquatic biodegradation. After removing low quality data, only 79 chemicals from the Environmental Fate Database remained. In that data set, 45 chemicals were active (i.e. biodegradable), 16 were inactive and 18 had marginal activity. MultiCASE identified aromatic and aliphatic thiol, methoxy, alcohol and carboxylic ester groups as anaerobic biophores, but no significant biophobes were found, most likely owing to the limited number of negatives for biodegradation in the data set. The relevance of all identified biophores was confirmed because all were shown to be associated with known metabolic transformations. In the MultiCASE methodology, substances with one or more biophores are judged biodegradable unless the molecule contains one or more biophobes. Therefore, inability to clearly identify biophobes should influence model accuracy for the non-biodegradable structures and this was in fact observed. The difficulties in finding biophobes (i.e. structures inhibiting biodegradation), negatively influenced model performance. The model did not handle well chemicals containing both easy and difficult to degrade substructures and tended to predict that such chemicals were biodegradable. As with the new BIOWIN MITI, the model was better at predicting biodegradable structures (89% correct) than non-biodegradable (64% correct).

CATABOL

The novel features of CATABOL, compared to other biodegradation prediction software are:

- Capability to predict biodegradation pathway;
- biodegradation extent assessment based on the entire pathway and not, as with all other models, the parent structure alone;
- accounting for adjacent fragments' effects before executing each transformation step;
- on-line documentation of every transformation.

CATABOL is therefore a research as well as a predictive tool and a depository of knowledge, as it brings together information on individual transformations and pathways. Currently CATABOL contains over 600 transformations, that often encompass more than one possible biodegradation step, thus improving the speed of predictions. Before executing the transformation of a target fragment, adjacent fragments are checked for inhibiting fragments. These inhibiting fragments may lower or completely prevent the probability of the execution of the transformation. There are three or four inhibiting fragments per transformation and therefore over 2000 combinations of transformations and inhibiting fragments in the system. There are two types of transformations: spontaneous and catabolic. Spontaneous transformations may be biotic or abiotic; catabolic transformations describe only biotic processes. The hierarchy of transformations is set according to descending probabilities of individual transformations that are derived from training data set. There are 2 transformation tables that differ in the order of the transformations owing to their differing probabilities. The first transformation table was calibrated with MITI test data, the second was calibrated with OECD 301B test data. Currently CATABOL generates one, the most probable, pathway.

TOPKAT

The Aerobic Biodegradability Module of the TOPKAT® package is a single, self-contained module, consisting of four structurally based sub-models. This module comprises a statistically significant and cross-validated quantitative structure-toxicity relationship (QSTR) model applicable to a specific class of chemicals and the data from which these models were derived. (MITI) I test database was used to develop these models. The 4 discriminant models compute the probability of a submitted structure of being capable of aerobic biodegradation (probability greater than 0.7) or incapable of being degraded aerobically (probability below 0.3). Probability values between 0.3 and 0.7 refer to an indeterminate region in which decisions should not be made except in special circumstances or under further analytical assessments.

Table D.1: Performance statistics as correct prediction of BIOWIN–MITI, MultiCASE/META and CATABOL models (from Jaworska et al, 2002) for MITI database[1]

	MITI-BIOWIN	META/ Nonlinear method	CATABOL MultiCASE
Predictions training set			
Not readily biodegradable (NRB):	85%	Not available	86%
Readily biodegradable (RB)	80%		91%
Total	83%		87%
Predictions validation set[2]			
NRB: Correct prediction	82%	80%	82%
RB: Correct prediction	77%	73%	91%
Total : Correct prediction	81%	77%	83%

[1] Comparative statistics for TOPKAT were not derived because the program was not available to the TF

[2] For MITI-BIOWIN, external validation with 295 MITI chemicals not used in training. For the remaining models, results are averages from 4 cross-validations leaving out 25% of the data used in training.

D.2.4. Bioconcentration

The purpose of the estimation of bioconcentration is to assess whether there is any potential for a chemical to accumulate in organisms and hence, for further transfer up a food chain. The process recommended for assessing the potential impact of bioconcentration and bioaccumulation is a step wise approach and is fully described in ECETOC (1995). Other reviews include Connell (1987), Bysshe (1990), Kristensen and Tyle (1990) and Gobas and Morrison (2000).

Points of note from these reviews include:

- The bioconcentration factor (BCF) is defined as the concentration in fish divided by the concentration in water at steady-state;
- in most cases the QSAR predicted BCF will not take account of factors which will lead to reduced bioconcentration, e.g. metabolism, nor will they account for other mechanisms that may lead to increased bioconcentration, e.g. facilitated uptake;
- when assessing QSARs for bioconcentration it is important to ensure parent compound BCFs are used;
- for the purposes of QSARs and to account for much of the variability noted in bioconcentration experiments, the BCF should be expressed in relation to the fish lipid content.

EU TGD and EUSES

Chapter 4 of the Technical Guidance Document contains a section on the QSARs available for bioconcentration (EC, 1996, 2003). The recommended model for log K_{ow} up to 6 is that described by Veith *et al* (1979), while for chemicals with log K_{ow} greater than 6, a parabolic equation, re-calculated from that described by Connell and Hawker (1998), is recommended. The models described are implicitly used in the EUSES system for the assessment of secondary poisoning potential.

The linear QSAR model described by Veith *et al* (1979):

$$\log BCF = -0.7 + 0.85 * \log K_{ow}$$

n=55; $r^2 = 0.86$; domain is mainly nonpolar and polar aromatics plus some surfactants

This QSAR has a number of issues that should be carefully assessed when considering its use. These include:

- The log K_{ow} values used include those based on estimations using HPLC. This may be acceptable for those chemicals that fitted the domain of the method used, however, in a number of cases the values were obtained by extrapolation. Thus the applicability of the method is unknown and the accuracy of the extrapolation uncertain;
- a number of the bioconcentration values used were from alternative sources and probably should be examined independently of the experimental values used;
- the data for BCFs in rainbow trout have been adjusted by the addition of 0.5 to allow for the difference between trout and bluegill. This was based on a bioconcentration experiment carried out on 1,2,4-trichlorobenzene and hexachlorobenzene to fathead minnow, green sunfish and rainbow trout. The trout bioconcentration values obtained were, for these two chemicals, approximately one third the value of those in the other two species;
- three chemicals were omitted because they gave low BCFs. This is an example of exclusion of chemicals from a domain without sufficient reasoning. The assessment of unknown chemicals becomes increasingly uncertain when such exclusions are not described, as it is also uncertain whether or not the unknown chemical fits the domain of the QSAR.

Mackay (1982) critically evaluated the Veith data, either altering the log K_{ow} to more accurate values or removing unsuitable chemicals, e.g. surfactants. This is clearly an example of changing or reducing the domain of the chemicals to which the QSAR applies. Further results were added, applying to chemicals that fitted the re-defined domain, mainly restricted to nonpolar aromatics. Based on this work, the regression was slightly improved and the slope set to 1.

The original Connell and Hawker (1998) model was recalculated as a parabolic equation and presented in the EU TGD (EC, 1996, 2003) for substances with log K_{ow} > 6:

$$\log BCF = -0.20 (\log K_{ow})^2 - 2.74 (\log K_{ow}) - 4.72$$

$n=43$; $r^2 = 0.78$

The model is based on experimental data from the literature for several fish species (n=45) and considers persistent chemicals, mainly chlorinated hydrocarbons, with a log K_{ow} of 3.4 – 9.8.

It is well established that a linear model of bioconcentration is inaccurate when predictions are based on chemicals with log K_{ow}s greater than 6 (Bintein *et al*, 1993). Hence, Bintein *et al* (1993) developed a non-linear QSAR model based on a training set of 154 chemicals and tested it with a small set of 29 chemicals:

$$\log BCF = 0.91 \log K_{ow} - 1.975 \log (6.8 \times 10^{-7} K_{ow} + 1) - 0.786$$

This model does give more accurate assessments of the potential for molecules with high log K_{ow} to bioconcentrate. In a recent study carried out by Devillers *et al* (1995b), this model gave the most realistic results for BCFs predicted for compounds with log K_{ow} above 6 and an assessment of the mean square errors showed these to be lower than the other models assessed in that study (Veith *et al*, 1979; 1980; Mackay, 1982; Isnard and Lambert, 1988; Connell and Hawker, 1998).

Performance of the above methods were assessed by Parkerton (2001) on the basis of experimental data for 35 substances from diverse chemical classes, many of which were known or expected to be significantly biotransformed by fish. Consequently BCF predictions based on this model were expected to be overly conservative as demonstrated in this analysis.

US EPA ASTER

ASTER (ASsessment Tools for the Evaluation of Risk) was developed by the US EPA Mid-Continent Ecology Division, Duluth, MN (MED-Duluth) and represents an integration of the AQUIRE (AQUatic toxicity Information REtrieval) toxic effects database and the QSAR expert system. When no experimental data are available, estimation methods are used for the endpoint.

The BCF used by the US EPA ASTER program is based on that described by Veith and Kosian (1983), e.g. a linear relationship between log K_{ow} and the BCF.

$\log BCF = -0.4 + 0.79^* \log K_{ow}$

n=122; $r^2 = 0.86$, general domain – non metabolised substances

Veith and Kosian (1983), after carrying out a wide sweep of available data from the literature, developed this very general QSAR, covering a wide range of different fish species. A number of su-sets are described and the changes to the correlations are discussed in detail.

SRC: BCFWIN

In order to address the outages in applicability for published QSARs, Syracuse Research Corporation (SRC) recently developed in cooperation with the US EPA a computer program (BCFWIN) that reviews all available BCF-data in the literature, thus allowing the estimation of BCF-values for a wider range of chemicals. This program estimates the BCF of an organic compound using the compound's log octanol-water partition coefficient (K_{ow}) and uses correction factors to take into account that certain structural and molecular factors influence bioaccumulation or stericly hinder uptake. The estimation methodology used by BCFWIN is described more completely in 2 documents (US EPA, 1997; Meylan *et al*, 1999).

The BCFWIN method classifies a compound as ionic or non-ionic. Ionic compounds include carboxylic acids, sulphonic acids and salts of sulphonic acids and charged nitrogen compounds (nitrogen with a +5 valence such as quaternary ammonium compounds); all other compounds are classified as non-ionic.

Non-ionic compounds are predicted by the following relationships:

$\log K_{ow} < 1.0$	$\log BCF = 0.50$
$\log K_{ow}$ 1.0 to 7.0	$\log BCF = 0.77 \log K_{ow} - 0.70 + \text{Sum F(i)}$
$\log K_{ow}$ 7.0 to 10.5	$\log BCF = -1.37 \log K_{ow} + 14.4 + \text{Sum F(i)}$
$\log K_{ow} > 10.5$	$\log BCF = 0.50$

(where Sum F(i) is the summation of structural correction factors)

Ionic compounds are predicted as follows:

$\log K_{ow} < 5.0$	$\log BCF = 0.50$
$\log K_{ow}$ 5.0 to 6.0	$\log BCF = 0.75$
$\log K_{ow}$ 6.0 to 7.0	$\log BCF = 1.75$
$\log K_{ow}$ 7.0 to 9.0	$\log BCF = 1.00$
$\log K_{ow} > 9.0$	$\log BCF = 0.50$

The methodology introduced a special treatment for organometals (tin and mercury), long chain alkyls and aromatic azo-compounds.

BCFWIN uses three data entry fields as follows:

SMILES: Structure of the compound as a SMILES notation (required); maximum of 360 characters allowed.
NAME: Name or ID of the compound; optional (not required for estimation).
log K_{ow}: Known value for the log octanol-water partition coefficient; it is optional (not required for estimation). When this value is entered here, it is used to estimate the BCF rather than the estimated K_{ow} or a K_{ow} retrieved from the experimental database.

BCFWIN estimates a log K_{ow} for every SMILES notation by using the estimation engine from SRCs LOGKOW (KOWWIN for Windows) program. BCFWIN also automatically retrieves experimental log K_{ow} values from a database containing more than 11500 organic compounds with reliably measured values. When a SMILES structure matches a database structure (via an exact atom-to-atom connection match), the experimental log K_{ow} value is retrieved and used to predict BCF rather than the estimated value.

Parkerton (2001) performed an external assessment of the BCFWIN system. Fish BCFs were compiled from primary literature and expressed on a wet weight basis for parent substance. Radiotracer studies reflecting bioconcentration of both parent substance + metabolites were specifically excluded. BCF data were identified for 50 substances. Chemical classes represented include:

- Hydrocarbons (alkanes, cyclics, monoaromatics, diaromatics and polyaromatics);
- chlorinated organics;
- phthalate esters;
- fragrance compounds (e.g. nitromusks);
- alkyl phenols;
- surfactants (alkyl sulphonates, alkyl ethoxylates, soap).

Many of these classes are known or expected to be significantly biotransformed by fish.

The results of this analysis indicate that BCFWIN provides a reasonable predictor of BCF. While the average residual error across this data set did not suggest a systematic bias in BCF predictions, clear biases were apparent for specific chemical classes. The analysis also indicates that even chemical classes that are perceived to be highly bioaccumulative (e.g. chlorinated organic chemicals) may have limited bioaccumulation potential owing to mitigating processes (e.g. biotransformation, bioavailability constraints).

Results of this exercise suggest that broad application of BCFWIN across chemical classes would not provide a sufficiently reliable tool for correctly characterising bioaccumulation potential in many regulatory decision-making contexts. Application of the current version of this model could lead to both false positive and negative conclusions depending on the BCF cut-off selected as illustrated for this data set (Table D.2).

Table D.2: Probabilities for correct and incorrect predictions by BCFWIN of bioaccumulation potential for different BCF cut-off triggers

	BCF cut-off = 500	**BCF cut-off = 5000**
Correct	0.7	0.82
False positive[1]	0.16	0.18
False negative[2]	0.14	0.0

[1] Predicted BCF > cut-off while measured BCF is < cut-off

[2] Predicted BCF < cut-off while measured BCF is > cut-off

For the purpose of classification and labelling in which a cut-off of 500 is used, BCFWIN is shown to misclassify 30% of the test substances included in this data set. However, if a higher BCF cut-off of 5000 is adopted the overall precision of the method increases to 82%. Only false positives are identified indicating BCFWIN shows promise as a possible screening tool for PBT assessments where a cut-off value of 5000 has been proposed. Further validation with a larger data set is needed before definitive conclusions can be drawn.

The BCFWIN method developed by SRC provides a considerably better fit to the data set of recommended BCF values than other methods (Table D.3). This is shown by the higher correlation coefficient (higher r^2-value), but more importantly, a much lower standard deviation (SD) and mean error (ME) when compared with other methods (Meylan *et al*, 1999). A similar conclusion was drawn by Parkerton (2001) in his external validation of the Danish EPA QSAR system that uses multiple approaches to BCF prediction.

Table D.3: Statistics for estimated versus measured BCF (Meylan et al, 1999)

Statistic	New method	Veith and Kosian (1983)	Bintein *et al* (1993))
Nonionic compounds			
n	610	610	610
r^2	0.73	0.31	0.52
SD	0.67	1.56	1.09
ME	0.48	1.04	0.83
Ionic compounds			
n	84	84	84
r^2	0.62	0.19	0.21
SD	0.41	20.3	1.90
ME	0.31	1.69	1.60
All compounds			
n	694	694	694
r^2	0.74	0.32	0.58
SD	0.65	1.62	1.25
ME	0.47	1.12	0.94

Danish EPA bioconcentration assessment (Parkerton, 2001)

A preliminary assessment of the Danish QSAR system was performed using the fish bioconcentration factor (BCF) as an indication of bioaccumulation potential. The BCF is a laboratory-derived measurement that reflects the steady-state ratio of the substance to that in aqueous exposure media. Fish BCFs were compiled from primary literature and expressed on a wet weight basis for parent substance. Radiotracer studies reflecting bioconcentration of both parent substance + metabolites were specifically excluded. BCF data were identified for 50 substances and are summarised in Table D.4 along with corresponding SMILES notations and CAS Nos (if available). Chemical classes represented include:

- Hydrocarbons (alkanes, cyclics, monoaromatics, diaromatics and polyaromatics);
- chlorinated organics;
- phthalate esters;
- fragrance compounds (e.g. nitromusks);
- alkyl phenols;
- surfactants (alkyl sulphonates, alkyl ethoxylates, soap).

Many of these classes are known or expected to be significantly biotransformed by fish. Consequently BCF predictions based on historical models (i.e. correlations with octanol-water partition coefficient) are expected to be overly conservative.

Three models for predicting fish BCF are considered in the Danish QSAR system:

- Method I - Meylan and Howard (1995) - this algorithm is included in the EPIWIN software that is commercially available from Syracuse Research Corporation;
- method II - Bintein *et al* (1993);
- method III - Connell and Hawker (1998).

The first method is expected to be the most reliable since it is based on an empirical model that was 'trained' using an extensive data set of measured BCFs for a wide range of organic chemicals which includes class-specific corrections. Methods II and III are based on simple parabolic correlations with the octanol-water partition coefficient. These latter two models are not applicable for surface-active agents since the octanol-water partition coefficient is not an appropriate descriptor of partitioning behaviour.

DEPA provided fish BCF predictions for all substances using Method I and for 36 of the substances using Method III included in this comparison exercise; no predictions were provided for using Method II. To assess the extent of agreement between model predictions and measured data, the log residual error was calculated for each substance as follows:

$$\log Resid_{error} = \log BCF_{predicted} - \log BCF_{measured}$$

Measured and predicted log BCFs and calculated residual error for Methods I and III are summarised in Table D.5. Residual error statistics for each chemical class are summarised in Table D.6.

As expected, the results of this analysis indicate that Model I provides a more reliable predictor of BCF than Method III; on average Method III overestimates measured BCFs by a factor of five. While the average residual error for Model I across this data set does not suggest a systematic bias in BCF predictions, clear biases were apparent for specific chemical classes. For example, Method I overestimates measured BCFs on average by more than a factor of five (log residual error >0.7) for phthalate esters, chlorinated toluenes, nitromusks and alkyl phenols. In contrast, this model significantly underestimates measured BCFs for alkyl sulphonates and sodium laurate (cf. Tables D.2 and D.3). In the case of the non-surface active substances, residual error is sometimes shown to increase (e.g. mono- and diaromatic hydrocarbons) or decrease (e.g. phthalates) as a function of log K_{ow} (Figure D.1). Directionally, negative residual errors can increase (i.e. sulphonates) or decrease (ethoxylates) with increasing alkyl chain length of the surfactant molecule (Table D.5).

The results suggest that broad application of Method I across chemical classes would not provide a sufficiently reliable tool for correctly characterising bioaccumulation potential in regulatory decision-making contexts. Application of the current version of this model could lead to both false positive and negative conclusions. This analysis also indicates that even chemical classes that are perceived to be highly bioaccumulative (e.g. chlorinated organic chemicals) may have limited bioaccumulation potential owing to mitigating processes (e.g. biotransformation, bioavailability constraints). While recognising the current limitations of Method I, this model is clearly preferable to Method III. While Method II was not investigated in the present study it is expected to yield results that are fundamentally similar to Method III.

Figure D.1: Dependence of residual error in the Syracuse Research Corporation BCF QSAR model as a function of the octanol-water partition coefficient for non-surface active substances

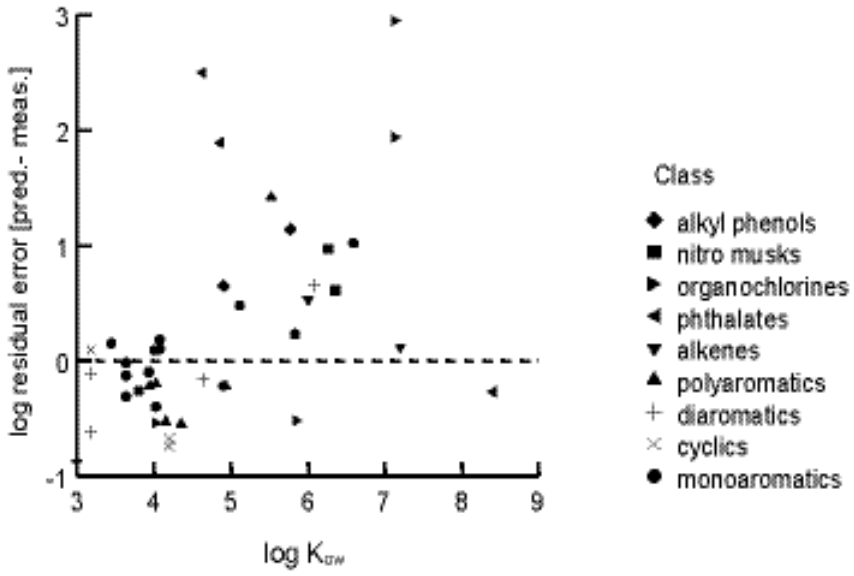


Table D.4: Summary of measured bioconcentration factors (BCF) for fish[1]

Compound		CAS No.	SMILES	(l/kg wet)	REFERENCE
Methylstyrene	Monoaromatic hydrocarbon	98-83-9	C(C(=C)C)(cccc1)c1	70	MITI, 1992
1,2,4-Trimethylbenzene	Monoaromatic hydrocarbon	95-63-6	C(ccc(c1C)C)(c1)C	137	MITI, 1992
1,3,5-Trimethylbenzene	Monoaromatic hydrocarbon	108-67-8	C(cc(cc1C)C)(c1)C	184	MITI, 1992
1,2,3-Trimethylbenzene	Monoaromatic hydrocarbon	526-73-8	C(c(c(cc1)C)C)(c1)C	186	MITI, 1992
Ethylstyrene	Monoaromatic hydrocarbon	28106-30-1	C=Cc1ccc(cc1)CC	459	MITI, 1992
m-Cymene	Monoaromatic hydrocarbon	535-77-3	C(cccc1C(C)C)(c1)C	518	MITI, 1992
p-Diethylbenzene	Monoaromatic hydrocarbon	105-05-5	C(ccc(c1)CC)(c1)CC	477	MITI, 1992
m-Diethylbenzene	Monoaromatic hydrocarbon	141-93-5	C(cccc1CC)(c1)CC	568	MITI, 1992
Divinylbenzene	Monoaromatic hydrocarbon	1321-74-0	C=Cc1ccccc1C=C	321	MITI, 1992
Di-isopropylbenzene	Polyaromatic hydrocarbon	25321-09-9	C(ccc1C(C)C)cc1C(C)C	2157	MITI, 1992
Hexachlorobenzene	Chlorinated organic	118-74-1	C(c(c(c(c1Cl)Cl)Cl)Cl)(c1Cl)Cl	18500	MITI, 1992
Diethylbiphenyl	Monoaromatic hydrocarbon	28575-17-9	CCc2cc(ccc2)c1cc(ccc1)CC	4000	MITI, 1992
Diphenylmethane	Monoaromatic hydrocarbon	101-81-5	C(cccc1)(c1)Cc(cccc2)c2	832	MITI, 1992
Naphthalene	Diaromatic hydrocarbon	91-20-3	C(c(ccc1)ccc2)(c1)c2	93	MITI, 1992
Tetrahydronaphthalene	Diaromatic hydrocarbon	26545-58-4	C(c(CCC1)ccc2)(C1)c2	260	MITI, 1992
Decalin (cis)	Diaromatic hydrocarbon	91-17-8	C(C(CCC1)CCC2)(C1)C2	1727	MITI, 1992
Decalin (trans)	Diaromatic hydrocarbon	91-17-8	C(C(CCC1)CCC2)(C1)C2	2008	MITI, 1992
Dibenzyltoluene	Monoaromatic hydrocarbon	26898-17-9	Cc1cc(ccc1Cc2ccccc2)Cc3ccccc3	2536	MITI, 1992
Fluorene	Polyaromatic hydrocarbon	86-73-7	C(c(c(c1ccc2)c2)ccc3)(c3)C1	567	MITI, 1992
Acenaphthylene	Polyaromatic hydrocarbon	208-96-8	C1ccc2cccc3c2c1C=C3	378	MITI, 1992
Acenapthene	Polyaromatic hydrocarbon	83-32-9	C(c(ccc1)ccc2)(c1CC3)c23	753	MITI, 1992
Anthracene	Polyaromatic hydrocarbon	120-12-7	C(c(ccc1)cc(c2ccc3)c3)(c1)c2	2023	MITI, 1992
Phenyl-xylylmethane	Monoaromatic hydrocarbon	13540-50-6	C1ccccc1Cc2cc(C)ccc2C	1605	MITI, 1992
Isopropylnaphthalene	Diaromatic hydrocarbon	29253-36-9	CC(c1cccc2c1cccc2)C	1111	MITI, 1992
Di-isopropylnaphthalene	Diaromatic hydrocarbon	38640-62-9	CC(c1c2cc(ccc2ccc1)C(C)C)C	2368	MITI, 1992
Cyclohexane	Cyclic hydrocarbon	110-82-7	C(CCCC1)C1	75	MITI, 1992

Table D.4: Summary of measured bioconcentration factors (BCF) for fish[1] (cont'd)

Compound		CAS No.	SMILES	(l/kg wet)	REFERENCE
Naphthalene	Diaromatic hydrocarbon	91-20-3	C(c(ccc1)ccc2)(c1)c2	302	de Maagd, 1997
Anthracene	Polyaromatic hydrocarbon	120-12-7	C(c(ccc1)cc(c2ccc3)c3)(c1)c2	2024	de Maagd, 1997
Benzo(a)anthracene	Polyaromatic hydrocarbon	56-55-3	C(c(c(c(c1)ccc2)c2)cc(c3ccc4)c4)(c1)c3	230	de Maagd, 1997
Fluoranthene	Polyaromatic hydrocarbon	206-44-0	C(c(ccc1)ccc2)(c1c(c3ccc4)c4)c23	3388	de Maagd, 1997
Dodecane	Alkane		CCCCCCCCCCCC	240	Tolls and van Dijk, 2001
2,2,4,6,6-Pentamethylheptane	Alkane		CC(C)(C)CC(C)CC(C)(C)C	2190	Tolls and van Dijk, 2001
Butylbenzyl phthalate	Phthalate	85-68-7	O=C(OCc(cccc1)c1)c(c(ccc2)C(=O)OCCCC)c2	12	Carr *et al*, 1997
Di-n-butyl phthalate	Phthalate	84-74-2	O=C(OCCCC)c(c(ccc1)C(=O)OCCCC)c1	2	Hüls, 1997
Di-2-ethylhexyl phthalate	Phthalate	117-81-7	O=C(OCC(CCCC)CC)c(c(ccc1)C(=O)OCC(CCCC)CC)c1	616	Mayer *et al*, 1976
2,3',4,4',5-Me tetrachlorotoluene	Chlorinated organic		C1c(CL)cc(CL)c(C)c1Cc2cc(CL)c(CL)cc2	479	van Haelst *et al*, 1996
2,3',4,4',3Me tetrachlorotoluene	Chlorinated organic		C1c(CL)c(C)c(CL)cc1Cc2cc(CL)c(CL)cc2	47	van Haelst *et al*, 1996
7-Acetyl-1,1,3,4,4,6-hexamethyl-1,2,3,4-tetrahydronapthalene	Fragrance compound	1506-02-1	O=C(c(c(cc(c1C(CC2C)(C)C)C2(C)C)C)c1)C	597	Balk and Ford, 1999
1,3,4,6,7,8-Hexahydro-4,6,6,7,8,8-hexamethylcyclopenta-gamma-2-benzopyran	Frangrance compound	1222-05-5	O(CC(c(c1cc(c2C(C3C)(C)C)C3(C)C)c2)C)C1	1584	Balk and Ford, 1999
Pentachloraniline	Chlorinated organic		CLc1Nc(CL)c(CL)c(CL)c1(CL)	363	de Wolf *et al*, 1993

Table D.4: Summary of measured bioconcentration factors (BCF) for fish[1] (cont'd)

Compound		CAS No.	SMILES	(l/kg wet)	REFERENCE
Musk xylene	Fragrance compound	81-15-2	O=N(=O)c(c(c(N(=O)=O)c(c1N(=O)=O)C(C)(C)C)C)c1C	4400	Rimkus et al, 1997
4-Tert-octyl phenol	Alkylphenol		C(C)(C)(C)CCCc1ccc(O)cc1	261	Tsuda et al, 2001
Sodium laurate	Surfactant	629-25-4	[Na]OC(=O)CCCCCCCCCCC	255	van Egmond et al, 1999
C13-2-n-(p-sulphophenylalkane)LAS	Surfactant		CC(c1ccc(S(=O)(=O)(O))cc1)CCCCCCCCCCC	670	Tolls et al, 1997
C12-2-n-(p-sulphophenylalkane)LAS	Surfactant		CC(c1ccc(S(=O)(=O)(O))cc1)CCCCCCCCCC	131	Tolls et al, 1997
C10-2-n-(p-sulphophenylalkane)LAS	Surfactant		CC(c1ccc(S(=O)(=O)(O))cc1)CCCCCCCC	4	Tolls et al, 1997
C11-5-n-(p-sulphophenylalkane)LAS	Surfactant		CCCCC(c1ccc(S(=O)(=O)(O))cc1)CCCCCC	8	Tolls et al, 1997
C13-5-n-(p-sulphophenylalkane)LAS	Surfactant		CCCCC(c1ccc(S(=O)(=O)(O))cc1)CCCCCCCC	34	Tolls et al, 1997
Alcohol ethoxylate - C14EO8	Surfactant		CCCCCCCCCCCCCCOCCOCCOCCOCCOCCOCCOCCOCCO	96	Tolls et al, 2000
Alcohol ethoxylate - C13EO8	Surfactant		CCCCCCCCCCCCCOCCOCCOCCOCCOCCOCCOCCOCCO	32	Tolls et al, 2000
Alcohol ethoxylate - C12EO8	Surfactant		CCCCCCCCCCCCOCCOCCOCCOCCOCCOCCOCCOCCO	12	Tolls et al, 2000
Nonylphenol	Alkylphenol		CCCCCCCCCc1ccc(O)cc1	452	Staples et al, 1998

[1] BCF values reflect steady-state conditions for the parent substance.

Table D.5: Measured and predicted log BCFs and model errors

Class	Chemical	Log BCF Measured	Log BCF Method I	Log BCF Method III	Log K_{ow}	Log Resid Method I	Log Resid Method III
1	Methyl styrene	1.85	2.00	2.34	3.44	0.15	0.49
1	1,2,4 trimethylbenzene	2.14	2.12	2.59	3.63	-0.02	0.45
1	1,3,5 trimethylbenzene	2.26	1.95	2.59	3.63	-0.31	0.33
1	1,2,3 trimethylbenzene	2.27	2.14	2.59	3.63	-0.13	0.32
1	Divinylbenzene	2.51	2.25	2.80	3.80	-0.26	0.29
1	Ethylstyrene	2.66	2.56	2.96	3.93	-0.10	0.30
1	m-Cymene	2.71	2.80	3.04	4.00	0.09	0.33
1	Diphenylmethane	2.92	2.52	3.06	4.02	-0.40	0.14
1	p-Diethylbenzene	2.68	2.86	3.12	4.07	0.18	0.44
1	m-Diethylbenzene	2.75	2.85	3.12	4.07	0.10	0.37
1	Di-isopropylbenzene	3.33	3.11	3.90	4.90	-0.22	0.57
1	Phenyl-xylylmethane	3.21	3.69	4.06	5.11	0.48	0.85
1	Diethylbiphenyl	3.60	3.83	4.46	5.83	0.23	0.86
1	Dibenzyltoluene	3.40	4.42	4.65	6.59	1.02	1.25
2	Cyclohexane	1.88	1.97	1.97	3.18	0.09	0.09
2	Decalin (cis)	3.24	2.56	3.26	4.20	-0.68	0.02
2	Decalin (trans)	3.30	2.56	3.26	4.20	-0.74	-0.04
3	Naphthalene	1.97	1.86	1.96	3.17	-0.11	-0.01
3	Naphthalene	2.48	1.86	1.96	3.17	-0.62	-0.52
3	Isopropylnaphthalene	3.05	2.89	3.68	4.63	-0.16	0.63
3	Di-isopropylnaphthalene	3.37	4.02	4.55	6.08	0.65	1.18
4	Acenaphthylene	2.58	2.36	2.97	3.94	-0.22	0.39
4	Fluorene	2.75	2.55	3.06	4.02	-0.20	0.31
4	Acenapthene	2.88	2.35	3.21	4.15	-0.53	0.33
4	Anthracene	3.31	2.76	3.41	4.35	-0.55	0.10
4	Fluoranthene	3.53	3.31	3.93	4.93	-0.22	0.40
4	Benzo(a)anthracene	2.36	3.78	4.31	5.52	1.42	1.95
5	Dodecane	2.38	2.50	NR	7.20	0.12	NC
5	2,2,4,6,6 pentamethylheptane	3.34	3.87	NR	6.00	0.53	NC
6	Di-n-butyl phthalate	0.30	2.80	3.66	4.61	2.50	3.36
6	Butylbenzyl phthalate	1.08	2.97	3.86	4.84	1.89	2.78
6	Di-2-ethylhexyl phthalate	2.79	2.52	4.19	8.39	-0.27	1.40
7	2,3',4,4',5-Me tetrachlorotoluene	2.68	4.62	NR	7.14	1.94	NC
7	2,3',4,4',3-Me tetrachlorotoluene	1.67	4.62	NR	7.14	2.95	NC
7	Pentachloraniline	2.56	2.02	NR	4.03	-0.54	NC
7	Hexachlorobenzene	4.27	3.75	4.47	5.86	-0.52	0.20
8	1,3,4,6,7,8-Hexahydro-4,6,6,7XXX	3.20	4.17	4.59	6.26	0.97	1.39
8	7-acetyl-1,1,3,4,4,6-hexamethXXX	2.78	3.39	4.61	6.35	0.61	1.83
9	4-tert-octylphenol	2.42	3.07	NR	4.90	0.65	NC
9	Nonylphenol	2.66	3.80	NR	5.77	1.14	NC
10	C13-2-n-(p-sulphophenylalkane)	2.83	0.78	NR	NC	-2.05	NC
10	C12-2-n-(p-sulphophenylalkane)	2.12	0.48	NR	NC	-1.64	NC
10	C10-2-n-(p-sulphophenylalkane)	0.60	0.48	NR	NC	-0.12	NC
10	C11-5-n-(p-sulphophenylalkane)	0.90	0.48	NR	NC	-0.42	NC
10	C13-5-n-(p-sulphophenylalkane)	1.53	0.78	NR	NC	-0.75	NC
11	Alcohol ethoxylate - C14EO8	1.98	2.04	NR	NC	0.06	NC
11	Alcohol ethoxylate - C13EO8	1.51	1.08	NR	NC	-0.43	NC
11	Alcohol ethoxylate - C12EO8	1.08	0.30	NR	NC	-0.78	NC
12	Sodium laurate	2.41	0.48	3.98	NC	-1.93	1.57

Classes: 1 = monoaromatic hydrocarbons; 2 = cyclic hydrocarbons; 3 = diaromatic hydrocarbons; 4 = polyaromatic hydrocarbons; 5= alkanes; 6 = phthalates; 7 = misc. organochlorines; 8 = fragrance compounds; 9 = alkyl phenols 10 = alkyl sulphonates; 11 =alkyl ethoxylates; 12 = soap

Table D.6: Residual model error by chemical class

Class	Method	Average Resid	Std Resid	Min Resid	Max Resid	Number
Monoaromatics	I	0.06	0.37	-0.40	1.02	14
Hydrocarbons	III	0.50	0.30	0.14	1.25	14
Cyclics	I	-0.44	0.46	-0.74	0.09	3
Hydrocarbons	III	0.02	0.07	-0.04	0.09	3
Diaromatics	I	-0.06	0.53	-0.62	0.65	4
Hydrocarbons	III	0.32	0.74	-0.52	1.18	4
Polyaromatics	I	-0.12	0.70	-0.55	1.42	7
Hydrocarbons	III	0.51	0.65	0.10	1.95	7
Alkanes	I	0.33	1.27	0.12	0.53	2
Phthalates	I	1.37	1.46	-0.27	2.50	3
Esters	III	2.51	1.01	1.40	3.36	3
Misc. Organo-	I	0.96	1.77	-0.54	2.95	4
Chlorines	III	0.20				1
Nitromusks	I	0.79	0.25	0.61	0.97	2
Fragances	III	1.61	0.31	1.39	1.83	2
Alkylphenols	I	0.90	0.35	0.65	1.14	2
Alkyl sulphonates	I	-1.00	0.82	-2.05	-0.12	5
Alkyl ethoxylates	I	-0.38	0.42	-0.78	0.06	3
Sodium laurate	I	-1.93				1
	III	1.57				1
All Substances	I	0.05	0.96	-2.05	2.95	50
	III	0.70	0.82	-0.52	3.36	35

D.2.5 Terrestrial bioaccumulation

A detailed review of (Q)SARs for predicting bioaccumulation in the terrestrial environment is beyond the scope of this present effort. However, it is important to point out that a number of terrestrial bioaccumulation endpoints are often required as input to substance risk assessments to characterise indirect exposure via the foodchain (e.g. potential for secondary poisoning). Key endpoints include:

- Plant-air partition coefficient;
- plant-soil pore water or plant-soil partition coefficient;
- worm-pore water or worm-soil partition coefficient;
- milk-meat transfer factor.

Available commercial software does not presently include prediction of terrestrial bioaccumulation, this endpoint is currently the focus of considerable research that should provide test methods and experimental data to support future QSAR development. For example, recent work has shown that the plant-air partition coefficient can be predicted based on species-specific correlations with the octanol-air partition coefficient (K_{oa}). For a recent review of this topic consult McLachlan (2000). Further development and validation of broadly applicable QSARs that reliably predict the transfer of chemicals to meat and milk is an urgent priority for improving human exposure estimates associated with this pathway (Bennett *et al*, 2003).

D.3 Aquatic toxicity

D.3.1 SAR

Verhaar *et al* (1992) developed a general rule-based classification system, that is based on chemical structure that can be applied to discrete organic chemicals. This classification system, which can be regarded as a SAR, was developed by evaluating the excess toxicity to fish (defined as the log ratio of the predicted baseline toxicity to the actual observed toxicity) for a variety of nonionic organic substances. Based on this analysis four substance classes were identified:

1. Inert chemicals exhibiting baseline toxicity;
2. Less inert chemicals which also appear to invoke a narcotic effect but exhibit toxicity that is a factor of 5 to 10 greater than baseline;
3. Reactive chemicals;
4. Specifically acting chemicals.

The latter two classes were found to exhibit toxicity in a range of 10 to 10000 greater than baseline. A recent external validation study of this classification system has been reported using the ECETOC (1993) aquatic toxicity database (Verhaar *et al*, 2000). Results from this exercise were promising, indicating that in most cases the proposed scheme predicted the correct range in acute fish toxicity, although some revisions to the initial rule-based scheme were identified. A limitation of this SAR is that metals, inorganics and ionisable organic chemicals are not presently included. The TGD (EC, 1996, 2003) uses the substance classes 1 and 2 specify 'mode of action' (MOA) specific QSARs. Similar approaches of the use of SAR to identify MOA specific QSARs can be found in the ASTER and OASIS systems, discussed further below.

D.3.2. QSAR

ECOSAR

ECOSAR uses molecular weight, structure and log K_{ow} to predict aquatic toxicity. The predictions are based on actual data for at least one member of a chemical class. The data (measured toxicity values) are correlated with molecular weight and log K_{ow} to derive a regression equation that may be used to predict aquatic toxicity of another chemical that belongs to the same chemical class. If a measured log K_{ow} is not available, ECOSAR utilises the SRC KOWWIN method to calculate this property. ECOSAR designates with an asterisk when estimated toxicity is above the solubility of a substance.

There are several special classes in ECOSAR that are not based on log K_{ow}. For example various surfactant SARs are based on the average carbon chain length and/or ethoxylate number. Anionic surfactant toxicity, for example, is predicted based entirely on chain length. Therefore, any anionic surfactant of a given chain length will be assigned the same toxicity value. The nuances of structure, which are often of critical importance, are not included in the analysis.

ECOSAR contains equations for >50 chemical classes (the full list can be found at www.epa.gov/opptintr/newchms/chemcat.htm), which can be categorised into four main areas:

1. Neutral organics that are nonreactive and nonionisable;
2. Organics that are reactive and ionisable and that exhibit 'excess toxicity' toxicity beyond narcosis associated with neutral organic toxicity);
3. Surface-active organic compounds (e.g. surfactants and polycationic polymers);
4. Inorganic compounds including organometallics.

To use ECOSAR for a particular chemical, it is necessary to select an appropriate SAR based on the following information:

- Chemical structure;
- chemical class;
- predicted log K_{ow};
- molecular weight;
- physical state;
- water solubility;
- number of carbons, ethoxylates or both;
- percent amine nitrogen or number of cationic charges, or both, per 1000 molecular weight.

Because the regression equations are chemical-specific and because they may vary by species (fish vs. daphnid vs. algae), a critical factor in reliable use of ECOSAR is identification of the appropriate chemical class for the chemical of concern.

If a measured log K_{ow} is not available, ECOSAR utilises the SRC KOWWIN method to calculate this property. ECOSAR designates with an asterisk when estimated toxicity is above the solubility of a substance.

A recent review of ECOSAR by Kaiser *et al* (1999) cautions the users against the indiscriminate use of the program and highlights the need for expert judgement on applicability of the predictions. Their overall conclusions:

1. Poor predictability (some (Q)SARs are developed with 2-4 points and their predictability is not verified);
2. No structural alerts of substructures that are not recognised;
3. Lack of transparency of training set;
4. Some equations are built on 1-3 data points and are therefore insufficient for quantitative predictions. The transparency of ECOSAR is mixed. The training data are not available but computational methods are very simple (all log K_{ow} based) and therefore easy to regenerate when data are made available.

ECOSAR – Procter and Gamble evaluation

Recently utility of ECOSAR to predict toxicity for surfactants, softeners, polymers and oligoamines has been reviewed by Procter and Gamble (Internal company report). ECOSAR could not make predictions for many polymers because the polymer module has not been developed or for softeners because they were predicted to be insoluble. Results for the remaining compounds are shown in the log-log plots below (Figures D.2 and D.3). Note the differences in the x-y axes. Overall, the predictive ability of ECOSAR was poor for compounds. ECOSAR predicted fish LC_{50} adequately for linear surfactants with chain lengths between 8 and 12 carbons; longer and shorter chain lengths and branched compounds were less well predicted. ECOSAR did not adequately predict algal toxicity of any of the chemicals in the Procter and Gamble data set.

Figure D.2: Fathead minnow versus ECOSAR predictions

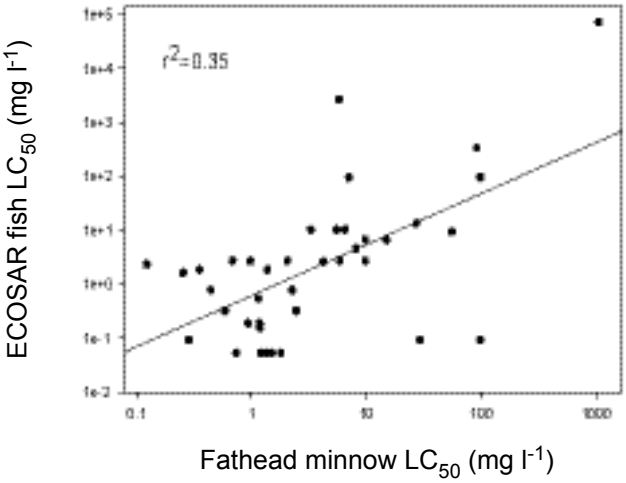
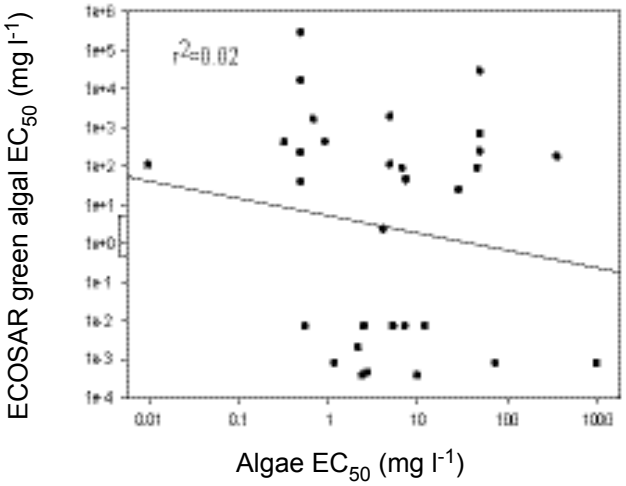


Figure D.3: Algal data versus ECOSAR predictions



ASTER

US EPA ORD developed expert system: Assessment Tools for the Evaluation of Risk (ASTER). This system selects QSARs based on the predicted mechanism of action of chemicals (Russom *et al*, 1997). The following modes of action were considered: narcosis I, narcosis II, narcosis III, oxidative phosphorylation uncoupling, acetylocholinesterase inhibition, respiratory inhibition, electrophile/proelectrophile reactivity mechanisms, central seizure mechanisms. To decide about a mechanism of action an expert system was constructed. A broad base of toxicodynamic information such as joint toxic action bioassays, fish acute toxicity syndromes (FATS), behavioural syndromes, LC_{50} ratios can been converted to 2 dimensional substructural rules that predict mechanism of action.

All QSARs in ASTER are log K_{ow} based. This expert system has been developed on a broader suite of information than the Verhaar system and empirical data. ASTER uses 617 fathead minnow data from the Duluth database, while Verhaar system is based on 117 guppy data. ASTER links two empirical databases: the AQUatic toxicity Information REtrieval (AQUIRE) database component of the US EPA's ECOTOX database http://www.epa.gov/ecotox/. Aster is no longer interactively available via Internet, but a user can submit a request in electronic form and predictions will be provided by US EPA.

TOPKAT

TOPKAT contains two models suitable for the prediction of acute toxicity. These are based on the fish (the Duluth fathead minnow) and invertebrates (*Daphnia magna* acute toxicity) database.

The fathead minnow LC_{50} module of the TOPKAT® package comprises of models for acyclics (hydrocarbons, alcohols, miscellaneous, alicyclics, multiple/fused benzenes, single benzenes) based on experimental data. The *Daphnia magna* module of the TOPKAT® package comprises of four models for phenols, single benzenes, other aromatics, aliphatics and (ii) a database of 252 experimental acute median effective concentration, EC_{50} values selected after critical review of open literature and the Acquire database.

MultiCASE

MultiCASE model has been developed for guppy toxicity (Klopman *et al*, 1999).

OASIS

OASIS is an expert system for prediction of acute aquatic fish toxicity of noncongeneric chemicals. It is based on the Duluth database of acute toxicities to fathead minnow for about 660 organic chemicals (i.e. 75% of the total number of chemicals from the database). The program has a build in expert system that chooses the most appropriate QSAR for a queried chemical and shows on a graph how the queried chemical is positioned vs chemicals used in the training set. The interaction of the non-covalent acting chemicals with the lipid-bilayer region of membranes is delineated by descriptors assessing the global electrophilic and hydrophobic character of molecules. The general form of the model used to evaluate non-covalent (narcosis) toxicities is:

$$\log(1/LC50) = b_0^N + b_1^N \log(BCF) + b_2^N GED \qquad [1]$$

where N is the type of narcosis, including nonpolar narcosis, amine narcosis and polar narcosis (phenols and anilines); BCF is bioconcentration factor (expressed as a non-linear function of the hydrophobicity); and GED is global electronic descriptor (e.g. lowest unoccupied molecular orbital and largest acceptor delocalisability of benzene ring atoms).

The domain of reactive chemicals is subdivided according to their putative mechanism of action conditioned by specific reactive groups. Quantum-chemical descriptors associated with those groups were combined with log(BCF) to model toxicity of these chemicals by making use of the following general formulae:

$$\log(1/LC50) = b_0^R + b_1^R \log(BCF) + \sum_i b_i^R RP_i \qquad [2]$$

where R corresponds to the specific class of covalently binding bioreactive chemicals and PR_i are descriptors of the group reactivity. The following reactivity parameters are used to model specific chemical classes: maximum donor delocalisability and acceptor delocalisability at α-C-atom, for α,β-unsaturated alcohols; charge at carbonyl oxygen, for aldehydes; bond order between carbon and halogens, for α,β- unsaturated halides (Mekenyan *et al*, 1996).

D.3.3 Use of (Q)SAR in statistical extrapolation method for environmental effect assessment

For environmental risk assessment of new or existing substances in Europe, PNECs are estimated from limited acute or chronic data though the use of conservative default application factors (AFs) (EC, 1996, 2003). The default AF chosen depends on the number of organism classes (e.g. vertebrates, invertebrates, plants) and endpoints (acute versus chronic) that are available in the ecotoxicity database for the substance. As a consequence, AFs implicitly take into consideration differences in species sensitivity and acute to chronic extrapolation. Alternatively, if sufficient ecotoxicity data are available, statistical extrapolation methods may be used that rely on species sensitivity distributions for deriving no effect concentrations (Aldenberg and Slob, 1993; Aldenberg and Jaworska, 2000; Posthuma *et al*, 2001). The advantages of using this approach in environmental risk assessment have recently been discussed by Grist *et al* (2002). However for most chemical substances insufficient ecotoxicity data are available for routine use of statistical extrapolation techniques. Moreover, owing to high costs associated with performing toxicity tests on multiple species, coupled with the general desire to reduce animal testing such methods will in the future continue in practice to be limited to a small subset of 'data rich' chemicals. Consequently, there is considerable incentive to use SAR or QSAR tools to improve the current paradigm for PNEC derivation based on default AFs. Recent research suggests a number of promising developments relevant to this regulatory management context.

In a recent analysis, de Zwart (2001) examined the relationship between the variance in the species sensitivity distribution (SSD) of a substance and the corresponding mode-of-action. It was hypothesised that substances with similar modes of actions would demonstrate SSDs with similar variance. Results suggest that SSD variance is indeed related to mode-of-action, with nonpolar narcotic chemicals yielding the most narrow range of interspecies sensitivities, while specific acting substances show greater species differences.

Roex *et al* (2000) examined the relationship between acute to chronic ratios (ACRs) in aquatic organisms and mode-of-action. These authors found that both the magnitude and variability of the ACR varied between classes (Table D.7). Substances designated as inert were found to exhibit the lowest and least variable ACRs while specific acting chemicals and metals demonstrated higher, more variable, ACRs with polar narcotics showing an intermediate behaviour. These authors concluded that given the consistent ACR observed for inert chemicals, acute toxicity tests could be used to provide reliable estimates of chronic effect endpoints. The above research indicates that mode-of-action based SARs may be a promising tool for refining default AFs used in PNEC derivation.

Table D.7: Mean and standard deviation of the acute to chronic ratio (ACR) by mode of action class

Class	**ACR**
Nonpolar narcotics (inert)	2.6 ± 1.6 (n = 11)
Polar narcotics (less inert)	9.8 ± 11.8 (n = 12)
Specific acting	17.3 ± 26.6 (n = 45)
Metals	15.3 ± 28.8 (n = 34)

Van Leeuwen *et al* (1992) and Verhaar *et al* (1994) have applied statistical extrapolation procedures to QSAR equations that were derived using available aquatic toxicity data on nonpolar narcotic chemicals for different test species and toxicity endpoints. The approach used can be outlined as follows. First, linear QSARs of the form:

$$\log NOEC = a * \log K_{ow} + b$$

are compiled, where NOEC is the chronic no effect concentration expressed on a molar basis, K_{ow} is the octanol-water partition coefficient and a and b are empirically derived constants obtained from regression analysis of toxicity data against K_{ow} for different test substances. For test species where only acute toxicity data are generally available, an ACR is applied to adjust the acute endpoint to a chronic NOEC. The no effect concentration (referred to as the HC_5) is then calculated from the statistics of the slope and intercept estimates obtained from the available population of QSAR equations:

$$Log\ HC_5 = \mu_a\ Log\ K_{ow} + \mu_b - K_z \sqrt{\sigma_a^2\ Log\ K_{ow} + \sigma_b^2 + \sigma_{ab}^2\ Log\ K_{ow}}$$

where:

HC_5 = aqueous concentration intended to protect 95% of species (mmol l)
μ_a = mean of QSAR slopes
μ_b = mean of QSAR intercepts
σ_b^2 = variance of QSAR intercepts
σ_{ab}^2 = covariance between QSAR slopes and intercepts
K_z = the 95% confidence extrapolation factor

Application of this QSAR-based approach for PNEC derivation was found to yield no effect concentrations that were in good agreement with US EPA final chronic values (FCV) for selected nonpolar narcotic chemicals (Verhaar *et al*, 1994). FCVs were derived from substance-specific aquatic toxicity data using a statistical extrapolation procedure developed by Stephan *et al* (1985) and are intended to serve as water quality criteria for aquatic ecosystem protection.

This technique has been subsequently extended to PNEC derivation for polar narcotic chemicals (Urrestarazu Ramos, 1998) as well as phthalate esters (Parkerton and Konkel, 2000). These concepts have also been further evaluated in different classes of narcotic chemicals by Di Toro *et al* (2000). This framework is based on the critical body residue (CBR) hypothesis advanced by McCarty and Mackay (1993) and is formalised in terms of a proposed target lipid model. In a recent critical review, Barron *et al* (2002) concluded that the QSAR-type approach developed by Di Toro *et al* (2000), that incorporates substance class differences in potency, organism lipid content and species sensitivity, offers the potential opportunity to advance the applicability of CBR theory in future regulatory use.

The advantages of using SAR/QSARs in support of PNEC derivation are clear. First, based on mechanistic principles, information that is available on many related substances can be collectively used in the effect assessment on a single substance. This maximises the information that can be gleaned from existing data while reducing animal use. In addition, such approaches would improve risk management decision-making, since the use of under or over conservative PNECs that are derived using default AFs would be avoided. Moreover, QSAR-based models for PNEC derivation are relatively straightforward to understand and apply from a practical user perspective. For a variety of chemical classes that act via a narcosis type mechanism, aquatic PNECs can in principle be reliably estimated from a knowledge of the substance's K_{ow}. Such transparency and ease of use should help to promote regulatory acceptance.

The potential utility of PNECs derived using QSAR based approaches was illustrated further by Mancini *et al* (2002). In this work, a comprehensive aquatic toxicity data set for the gasoline oxygenate methyl tert butyl ether (MTBE) was generated so that an ambient water quality criterion could be developed using US EPA methodology (Stephan *et al*, 1985). Acute toxicity tests were conducted for 8 families of freshwater and marine organisms (representing 25 individual test species) as well as chronic tests for three species, so that the minimum data set requirements for applying this methodology could be satisfied. Using these data, the final chronic value (FCV) for MTBE in fresh and marine water was calculated to be 51 and 18 mg/l, respectively.

Based on structural considerations MTBE is expected to be a classical nonpolar narcotic chemical (i.e. eliciting simple baseline toxicity). Inputting the log K_{ow} of this compound (= 1.56) obtained from the SPARC model (Karickhoff *et al*, 1991) into the linear equation given by Di Toro *et al* (2000) that relates log FCV to log K_{ow} for nonpolar narcotic chemicals yields an estimated FCV of 21 mg/l. It should be noted that Di Toro's target lipid model is based on toxicity data for both freshwater and marine organisms, and hence the agreement obtained between PNECs derived from the experimentally derived species sensitivity distributions for this substance and the narcosis-based QSAR prediction is excellent.

The above discussion highlights some of the recent developments in the field of SAR/QSAR that are relevant to environmental effect assessment. A critical evaluation of the relationship between inter-species variation in the species sensitivity distributions for both acute and chronic endpoints and mode-of-action would be a useful extension of the work by Roex *et al* (2000). The ECETOC (2003b) aquatic toxicity database would serve as an ideal source of information to support this analysis. Further validation, such as the example cited above, should be undertaken to determine if such SAR/QSAR tools can be exploited better in the future for improving the technical basis and reducing the resources required for environmental effect assessment.

APPENDIX E: SOFTWARE FOR HUMAN HEALTH ENDPOINTS

E.1 Software evaluated by TF

DEREK (Deductive Estimation of Risk from Existing Knowledge)

DEREK is a rule-based (knowledge-based) system for the qualitative prediction of several toxicological endpoints, developed and marketed by Sanderson and Earnshaw (1991) and which was updated by Ridings *et al* (1996).

DEREK identifies so-called toxicophores (segments of the molecule associated with a specific activity), highlighting them and providing information about the mechanism of activity as comments from the rule writer, literature and other available information. The likelihood of toxic activity for the submitted compound is provided. DEREK contains rules to identify segments for adverse effects, e.g. carcinogenicity, irritancy, lachrymation, mutagenicity, respiration and skin sensitisation. Some of these endpoints are not always covered sufficiently. DEREK has a very limited number of deactivating rules.

Development of the rules of the DEREK program is controlled by a group of commercial, educational and non-profit-making organisations.

The Microsoft Windows compatible version uses the MDL ISIS/Draw package, but MDL standard MOL files can also be imported. Entering the structure of the molecule, the program searches the rulebases and structural alerts are highlighted within the structure and the user is able to check the alert description for the rule thereof. For many structural alerts, the hazard evaluation is justified with relevant literature references. There are no single modules available.

Advantages of DEREK:

- The database is 'open', so new information can be added easily;
- the existing rules, as well as the database, can be changed or adjusted by the user;
- the program is not a 'black-box'. There are explanations why a certain rule was used or why not;
- batch mode processing enables large-scale throughput;
- all endpoints are covered in a single run;
- DEREK offers the possibility to include metabolic consideration (via link with METEOR).

Disadvantages of DEREK:

- DEREK is geared towards finding active substances and does not respect activating or detoxifying mechanisms in a balanced way;
- 3-dimensional QSARs, accounting receptor-based mechanisms, are not possible with DEREK;
- DEREK does not discriminate between non-toxic compounds and compounds for which no information on toxicity is available.

HAZARDEXPERT

HAZARDEXPERT is a rule-based system developed by Smithing and Darvas (1992) identifying toxic segments in a molecule and alerting the user. In addition, it provides species specific information across a range of trophic levels with different dosing regimens. The system works by searching the query structure for known toxicophores. These are held in the 'Toxic Fragments Knowledge Base', which is based on literature and on reports by the US EPA. The underlying chemicals database is accessible to the user. New substances can be added to the database by using a graphical interface or by incorporating the metabolites predicted by MetabolExpert. Similarly, the knowledge-bases on metabolic transformation, toxic fragments and references can be modified. For further information see Greene (2002), Dearden *et al* (1997).

No further evaluation of the program was performed by the ECETOC TF.

MultiCASE

The software is introduced in Section 3 and Appendix D.

TOPKAT (Toxicity Prediction by Komputer Assisted Technology)

The principles of TOPKAT are introduced in Appendix D.

OncoLogic

OncoLogic is a rule-based expert system for the prediction of the carcinogenic potential developed by the US EPA to systematise and make available the Agency's predictive expertise, which is used routinely in a regulatory setting when screening chemicals for potential carcinogenicity. The system consists of a structural arm, which searches for structural elements that may contribute to the carcinogenic potential and a functional arm, which evaluates the remainder of the molecule for deactivating structural elements.

The chemical universe is subdivided into 4 subsystems: fibers, polymers, metals and organic chemicals. Each is built with specific knowledge rules that are unique for the subsystem (Woo *et al*, 1995, 1997). The subsystems vary considerably in user interface, function and information content. To run the program, the query structure has to be assigned to one of the predefined chemical classes. Even for a chemical expert, this may often be arbitrary if the structure does not fit neatly into one of these classes. The program produces a detailed justification report which conveys the mechanism-based expert reasoning underlying the evaluation. The rulebase cannot be changed by the user. There is no batch mode available. Further information can also be obtained by Dearden *et al* (1997) and Greene (2002).

No further evaluation of this system was performed by the ECETOC TF.

BgVV DSS (Decision Support System)

The DSS was developed by the German Federal Institute for Health Protection and Veterinary Medicine (BgVV) (Zinke and Gerner, 2000; Gerner *et al*, 2000). It is a computerised rule-based model, for predicting the skin and/or eye irritation/corrosion potential of a chemical from its physico-chemical properties and molecular structure. The SAR rules were derived from physico-chemical and toxicological data for new chemical substances submitted to the regulatory authorities for the EU notification procedure. These rules are used to predict the potential of a chemical to act as an irritant/corrosive to skin and/or eye from reactive substructures within the molecule. The absence of an irritating potential is derived from data on aqueous and lipid solubility, partition coefficient log K_{ow}, surface activity, melting point and molecular weight.

The system is currently used by the BgVV to screen new chemicals notified within the EU.

Advantages:

- DSS is used by German competent authority (BgVV) and its training set consists of GLP-conforming test results, performed in accordance with official guidelines;
- the system gives quick results and on-screen messages provide information about why a rule is chosen or rejected;
- there is high flexibility in rule modification for someone with programming skills;
- takes into account skin permeation based on molecular weight;
- the software is available without costs.

Disadvantages:

- The input of the molecular formula is complicated and time-consuming, owing to a specific one-dimensional mathematical representation of the three dimensional chemical;
- since the program is built on proprietary data from notifications of new chemical substances, the training set is not available to the user.

TOXSYS/QSARIS

TOXSYS is primarily a searchable toxicological database of some 230,000 compounds divided into various toxicity categories including acute toxicity, reproductive toxicity, carcinogenicity, mutagenicity, eye irritation, human skin patch test, repeat dose toxicity, food additive toxic effects, hepatotoxic and nephrotoxic effects of drugs and endocrine disruptors. The major usefulness of TOXSYS is its ability to rapidly and easily search an extensive database for toxicological information (including literature reference, summary of toxicity data) based on chemical structure, chemical formula, MW, CAS number, substructure or compound similarity. The program provides an easy method for running similarity searches to find comparable compounds and for screening for probable toxic chemicals (i.e. SAR).

TOXSYS has a useful but limited QSAR module for predicting acute oral toxicity (i.e. mouse LD_{50} data) and mutagenicity (i.e. Ames Salmonella assay). The computational algorithm utilises a neural net approach for predicting LD_{50} values based on acute oral toxicity in the mouse. The oral LD_{50} and mutagenicity predictive models in TOXSYS are easy to run to screen large numbers of chemicals since they can be carried out in batch mode (compounds are processed as files in MDL mol or SMILES format). TOXSYS has the capability to export structure and toxicity data to the related QSARIS program in which the users can develop and validate their own QSAR predictive toxicity models.

Since TOXSYS is relatively new, this software program currently has found limited use in industry. The program runs under Windows and is very user friendly. TOXSYS costs US $7500 with a maintenance fee of US $7500 for biannual updates of the toxicity databases and release of new QSAR predictive models. QSARIS costs US $15,000 and requires some experience in modelling/data analysis. SciVision, a subsidiary of MDL Systems, supports/markets and updates both TOXSYS and QSARIS.

TOXSYS and QSARIS have not been extensively evaluated by the ECETOC TF.

E.2 Dermal penetration

Many studies have examined the relationship between skin permeability and physico-chemical properties using compilations of permeability coefficients K_p obtained from *in vivo* studies. Potts and Guy (1992) with a data set of more than 90 chemicals indicated that a multiple regression, involving log K_{ow} and molecular weight (MW), provided a very reasonable fit with log K_p (2).

$$\log K_p\ [\text{cm h}^{-1}] = 0.71 \times \log K_{ow} - 0.0061 \times MW - 2.72 \qquad (r^2 = 0.67) \qquad (2)$$

Further quantitative structure-permeability relationship approaches by different authors slightly refined this model confirming that human skin penetration is mainly governed by hydrophobicity and molecular size (Cronin *et al*, 1999).

In addition, attempts have been made by including further parameters to describe effects such as hydrogen bonding (El-Tayar *et al*, 1991; Roberts *et al*, 1996; Buchwald and Bodor, 2001) or the use of solvatochromic parameters (Abraham *et al*, 1995; Potts and Guy, 1995). Another prediction method using molecular fragments was reported by Pugh and Hadgraft (1994). More recently, a refined quantitative structure-permeability relationship model was described using a data set of 143 structurally diverse chemicals (Patel *et al*, 2002). Most of the quantitative structure-permeability relationship models tend to over estimate skin permeability for lipophilic compounds (log $K_{ow} > 4$). Furthermore, they do not include ionised compounds and that they are limited to aqueous vehicles.

Understanding the mechanism of skin penetration is crucial in risk assessment for dermal exposure. For instance, skin penetration may be regarded as the first step in a series of processes finally leading to skin sensitisation. DEREK, for example, integrates skin permeability (calculated using the Potts and Guy equation) for predicting skin sensitisation and photosensitivity.

E.3 Skin irritation and corrosion

Toxicological evaluation and assessment of skin reactivity usually considers irritation and corrosion. Skin irritation is characterised by moderate, reversible effects. Irritation is often accompanied by cytotoxicity and direct effects on dermal blood vessels and cell surface adhesion molecules. Skin corrosion has been distinguished from skin irritation in two ways. Corrosive skin reactions generally occur quickly after chemical exposures and are irreversible. It is also thought that the major direct processes leading to chemical corrosivity are more commonly physico-chemical in nature rather than the result of inflammatory biological processes. (Lewis and Botham, 1994.).

Irritant and corrosive effects are influenced by the molecular size (cut off: MW 1200), log K_{ow}, pKa, skin permeability, surface tension as well as water and lipid solubility (Gerner *et al*, 2000). Lipid solubility, in contrast to water solubility, is crucial for the potential of a chemical to cause skin lesions. Considerations on the expected local reactivity of a chemical can be based on log K_{ow}, molecular weight and on considerations on the nature of hetero atoms within the molecule. Surface tension is often correlated with a prediction of severe effects on skin and eye but not for prediction of moderate effects. The skin permeability of inorganic acids and bases and oxidising agents is dependent on their pKa-values and high polarity. They are often expected to be corrosive because of their destructive effect on the stratum corneum. Anionic and cationic surfactants have little skin penetrating properties and they can have corrosive effects. Cationic surfactants are more cytotoxic than anionic surfactants and corrosivity might be owing to solubilising the stratum corneum. Organic solvents remove cellular lipids from the skin, resulting in a loss of barrier properties leading mainly to irritant effects but rarely to corrosion.

For the development of (Q)SARs, mainly the above mentioned physico-chemical properties are correlated with the experimental data. However, development of good quality QSARs has been hindered by a lack of suitable *in vivo* data and a limited understanding of the complex mechanisms.

For many years the method to assess irritant effects of chemicals *in vivo* has been the Draize test for skin irritation. This test design is known to be overpredictive and not always suitable to reflect the human situation.

Current (Q)SARs are generally based on Draize test data and are therefore difficult to extrapolate to humans. There are no models available which use the more relevant human patch test data (Robinson *et al*, 2001).

The results of studies on the performance of the DEREK- and DSS-system (see Appendix F) show that the prediction capability of these systems with respect to skin irritancy is very limited.

E.4 Eye irritation

Eye irritancy is generally caused by comparable mechanisms as skin irritation. However, eye damage seems to be negatively correlated with lipid solubility and moderate irritation rather than serious damage needs some aqueous solubility. Descriptors often used for the prediction of eye irritation are hydrophobicity, dipole moments, molecular weight/size, log K_{ow} and pKa (Gerner *et al*, 2000).

- Chemicals with intermediate hydrophobicity, intermediate dipole moments and small molecular weight tend to be irritant to eyes;
- Chemicals with no or low dipole moments (not reactive) tend to have lower irritancy to eyes.

As for skin irritation, the available experimental data which are generally used for (Q)SAR development are from Draize tests, which are considered of limited relevance for the human situation. There are currently not sufficient data available from other probably more relevant tests (e.g. *in vivo*, *in vitro*, Low Volume Eye Test).

The prediction capabilities for the identification of irritation potential of substances of two available systems, DEREK and DSS, were evaluated by the Task Force (Appendix F). Neither performed to the desired level of accuracy, indicating that they are not generally applicable for the prediction of this endpoint.

E.5 Skin sensitisation

Allergic contact dermatitis evolves from two different phases. The first phase (induction phase) is initiated when a susceptible individual is exposed topically to the inducing chemical allergen. This stimulates a primary cutaneous immune response and results in allergic sensitisation. If the now-sensitised individual is exposed subsequently to the same allergen an accelerated and more aggressive secondary immune response will be provoked. Allergen-responsive T lymphocytes are activated at the site of contact and release cytokines and other inflammatory mediators that cause the inflammatory reaction recognised clinically as allergic contact dermatitis (Gerberick *et al*, 2000).

To elicit an immune response a chemical must gain access to the epidermis. Thus, a skin sensitising chemical requires the physico-chemical properties necessary for the passage through the stratum corneum, which under normal circumstances represents an effective barrier to many chemicals. In their native state chemical allergens are haptens and of insufficient size to provoke an immune response. For this to be achieved the chemicals must form stable conjugates with macromolecular proteins. For that reason skin sensitising agents are either protein-reactive or can be metabolised in the skin to a protein-reactive species (Basketter, 1998; Kimber *et al*, 2001). Ideally, a (Q)SAR should address all these parameters, which is difficult to achieve.

Until now, the guinea pig maximisation test (GPMT) (Magnusson and Kligman, 1969) and the Buehler occluded patch test (Buehler, 1965) have been the preferred methods in assessing skin sensitising potential of chemicals. An alternative to these methods is the more recently developed murine local lymph node assay (LLNA). This employs a different approach in which skin sensitising chemicals are identified on the basis of their ability to stimulate lymphocyte prophylactic responses during the induction phase (Kimber and Basketter, 1992; Kimber *et al*, 1994). This method correlates well with human data.

The existing commercial (Q)SAR systems for prediction of skin sensitisation are based on guinea pig data, no models for LLNA are available yet.

To assess the predictive power of commercial prediction systems in the identification of skin sensitising potential, different subsets of chemicals were evaluated with DEREK or TOPKAT. The results (Appendix G) show that the systems tend to be over-predictive for this endpoint which can be seen by the enhanced number of false positive predictions and the resulting low specificity (7 - 85%).

E.6 Acute mammalian toxicity (LD_{50})

The determination of acute oral toxicity provides information on the possible effects occurring within a short time after the administration of a single dose. For a biological endpoint such as death, a well-defined mode-of-action or common site of action are difficult to define because it involves various mechanisms which are not analysed in a standard acute toxicity study.

Although there are a considerable number of acute toxicity studies published in open literature it may be difficult to predict a whole animal phenomenon like death without specific knowledge of the mechanism of action. General (Q)SARs covering non-congeneric chemical compounds may thus not give reliable predictions. (Q)SARs for specific chemical classes may perform better, but more validation is necessary to establish the boundaries of such chemical classes.

The limited data on the performance of the acute toxicity models presented in Appendix H, do not allow for a general conclusion concerning the predictive power of such systems. However, the data indicate, that the quality of the prediction is dependent on the chemical class and thus the applicability of such systems is considered restricted.

E.7 Chronic mammalian toxicity

The determination of no observed (adverse) effect levels (NO(A)EL), lowest observed (adverse) effect levels (LO(A)EL) or maximum tolerated dose (MTD) in repeated dose toxicity studies is generally dependent on the toxicity on several organs, the dose and duration of dosing. Moreover, multiple mechanisms can lead to the same toxicological effect. Thus, numerous different mechanisms of action have to be taken into account when developing a generally applicable (Q)SAR for non-congeneric compounds. Moreover, the determination of a NO(A)EL/LO(A)EL or MTD in experimental studies is also influenced by the dose selection and is subject to biological and experimental variability and can therefore not be taken as an absolute number. To overcome these difficulties in modelling, careful selection of the data and analysis is necessary.

For well known mechanisms of action or well-defined congeneric groups of compounds the development of a comprehensive QSAR models may be feasible. General models for this endpoint are difficult to develop and should only be used with great care because of the complexity of the biological system, the lack of a well-defined mode-of-action or a common site of action and a lack of mechanistic data, coupled with the lack of data representing the chemical variability. As a consequence there are only few (Q)SAR models for chronic toxicity currently available on the market with limited reliability.

No thorough analysis of the predictive power of the current commercially available systems could be performed owing to lack of sufficient valid experimental data.

E.8 Reproduction toxicity

Reproduction toxicology examines the possible effects on the reproduction process, including effects on the reproduction organs and their hormonal regulation, conception, implantation and embryo-fetal development. Impairment of fertility represents disorders of male or female reproductive functions or capacity. Developmental toxicity is part of reproduction toxicology and studies the effects of chemicals during embryonal and fetal development, which can be death of pups, growth retardation, structural malformations, or functional disorders of organs.

The mechanism of action of reproductive or developmental toxic compounds is seldom known. Mechanisms of action for these endpoints are complex as several organs may be involved including their interaction. Specific growth factors and hormones may regulate this interaction.

There are no (Q)SAR models for the prediction of fertility. The focus in developmental toxicology is mainly on teratogenicity assessment and hence this is the focus area to derive predictive models.

Teratogenicity

In teratogenicity studies, the effects of chemicals on the development of organs during embryo and foetal development are studied. The outcome of these studies is dependent on a variety of factors which complicate the development of (Q)SARs such as:

- The large number of different mechanisms of action that are not exactly understood;
- the difficulty of predicting the distribution of a maternally administered dose and possible transplacental transfer of the compound;
- the problem of maternally mediated effects;
- the dependence of possible effects on the day of gestation when the substance is administered;

- the low number of compounds for which developmental toxicity is described in open literature resulting in an insufficient coverage of chemical substructures in the training sets and in insufficiently robust sets of descriptors.

These factors severely limit the applicability of the current commercial (Q)SAR systems. The study by Pearl *et al* (2001) shows that currently these systems are not able to reliably identify teratogenic compounds with sensitivities ranging from 12 to 53%. (Appendix I).

E.9 Mutagenicity, in vitro (Ames-Test)

Most (Q)SAR models were developed for the bacterial reverse mutagenicity assay (Ames Test). A considerable number of chemicals has been tested for bacterial mutagenicity. These data have been used for the development of prediction models.

Mutagenicity in the Ames Test is mainly determined by the electrophilicity of a compound, which determines the potential to interact with the genetic material. Owing to the relatively good understanding of the underlying mechanisms, mechanistically based modelling is facilitated. Additionally, this endpoint and test system is less complex and thus shows less biological and experimental variability than *in vivo* assays.

The performance of currently available prediction systems were analysed in different studies. The concordance in these studies ranged from 56 to 85%. For rule-based systems, the concordance could be increased by inclusion of class specific new rules into system. It was shown that the combination of different prediction systems often enhances the reliability of the predictions and their sensitivity. However, the number of compounds for which predictions were possible and the specificity was considerably decreased when different systems were used in combination.

Although the relatively good understanding and relatively low experimental variability, the commercial systems are still not generally applicable to all chemical classes. They still have deficiencies in predicting bacterial mutagenicity accurately because of insufficient coverage of the whole chemical universe and the lack of appropriate rules /descriptors. It was shown that the development of models for specific classes of chemicals could overcome these deficiencies (Appendix J).

E.10 Carcinogenicity

Commercial systems for predicting chemical carcinogenicity have been developed by examination of existing experimental rodent studies (NTP, FDA, CPDB databases), that meet minimum protocol requirements. These standard 2-year rodent carcinogenicity bioassays result in a subset of 50 or more possible tumour sites for each rodent submodel, i.e. male rat, female rat, male mouse, female mouse. However, information on tumour sites and rodent submodels is sparsely populated in the study reports. Some chemicals cause tumours only at a single site in one rodent submodel, others cause tumours at multiple sites in multiple rodent submodels.

Two factors that strongly influence the success of a modelling effort are the experimental reproducibility of rodent bioassay results and the categorical assignment of carcinogenic response within the training set. With regard to the issue of reproducibility, the concordance of NTP bioassay results with non-standardised bioassay results (i.e. results taken from literature studies), on a common set of >100 chemicals for which both sets of data were available, was found to be less than 60% (Gottmann *et al*, 2001). This points to a high degree of biological and experimental variability inherent in existing rodent carcinogenicity studies. With respect to categorical assignment (positive or negative) of summarised carcinogenic response, there is also variability in the interpretation of the results concerning the biological significance of single site/sex/species carcinogens and how to best combine rodent submodel results to be most predictive for human carcinogenic risk. Moreover, it has to be noted that the NTP bioassay test set is heavily biased towards 'difficult' chemicals already suspected of carcinogenicity leading to a lower prediction capability for chemicals more typical of the universe of mostly non-carcinogenic chemicals. To overcome these difficulties in modelling rodent carcinogenicity, careful selection of the rodent data and analysis is necessary. However, given the structural diversity and mechanistic complexity of the rodent carcinogenicity database, coupled with the lack of data representation for large portions of the chemical universe, it is questionable whether the current data and knowledge base relevant to this endpoint is sufficient to serve as a basis for a general prediction system. For some chemicals within well-defined classes or well known mechanisms, success is most probable; but for others that are poorly represented reliable modelling might not be possible (Richard and Benigni, 2001; Benigni and Passerini, 2002).

None of the commercially available programs could be used without expert judgement for non-congeneric compounds. For the prediction of the carcinogenic potential of a set of 44 compounds the (Q)SAR systems received an overall accuracy of 49-59% compared to 75% that was reached by an expert panel (Benigni, 1997) (Appendix K).

APPENDIX F: Prediction of Irritation / Corrosion (Skin / Eye)

DEREK

The DEREK rulebase consists of nine rules for the prediction of skin irritancy and a further 10 rules which are specific for the prediction of eye irritancy. The performance of this rulebase was studied by the Health and Safety Laboratory, UK, using a combined data set of about 300 chemicals. The irritancy data were retrieved from published EC classifications of new or existing chemicals, the ECETOC reference data banks for eye and skin irritation (ECETOC 1998b) and a large set of eye irritation data from a laboratory. About 75% of the chemicals were skin or eye irritants or corrosives. The other chemicals were non or only weak irritants. Only 23% of the skin irritants and only 30% of the eye irritants were correctly identified by DEREK (sensitivity: 23%). 19 of 45 non-irritating chemicals were predicted to be irritating (false positives). Only 20 – 40% of the chemicals classified as corrosive were predicted to be irritants by DEREK. The remaining chemicals, not being predicted as corrosive/irritant, were chemicals known to be reactive with water or biological molecules (e.g. anhydrides, peroxy acids) and surfactants. These results indicate that the prediction capability of DEREK to predict skin and eye irritancy and/or corrosivity is very limited. This may be owing to the various mechanisms of irritancy and corrosivity, which are not yet fully understood and resulting from this a lack of rules to predict irritancy/corrosivity.

DSS (Decision Support System)

The program was externally validated by the BgVV using 331 chemicals which are not part of the training set (Zinke *et al*, 2000).

Table F.1:

No. of chemicals	Experimental result	corrosive	Prediction corrosion probable	corrosion not probable	not corrosive	no prediction
24/331 (7.3%)	corrosive to skin	**16/24 (66.7%)**	5/24 (20.8%)	0/24 (-)	1/24 (42%)	2/24 (8.3)
258/331 (77.9%)	not corrosive to skin	1/258 (0.4%)	20/258 (7.8%)	33/258 (12.8%)	**163/258 (63.2%)**	41/258 (15.9%)
33/331 (10.0%)	not known	1/33 (3%)	4/33 (12.1%)	4/33 (12.1%)	8/33 (24.2%)	**16/33 (48.5%)**

The program was also tested by Degussa AG, submitting 25 different, GLP-conforming tested chemicals to the DSS-system.

Table F.2: Data for skin irritation/corrosion

Skin Irritation/corrosion	Correct prediction	Wrong	Recommendation) for testing (*in vitro*)	Open answer (no rule)
25 substances	6 (24%)	2 (8%)	7 (28%)	10 (40%)

Table F.3: Data for eye irritation/corrosion

Skin Irritation/corrosion	Correct prediction	Wrong	Recommendation) for testing (*in vitro*)	Open answer (no rule)
24 substances	3 (12.5%)	6 (25%)	1 (4%)	14 (58%)

For 40% of the chemicals, no prediction could be supplied by DSS for skin irritation. For eye irritation, no prediction could be made for 58% of the chemicals tested. Correct predictions of the skin irritation potency could only seen in 24% of all chemicals tested. The prediction competence for eye irritation resulted in 13% correct answers, only. The DSS is designed to be over sensitive providing a large number of 'false positive' predictions. The custom test of 25 substances shows that at present the program cannot be applied to gain reliable results for skin or eye irritancy/corrosivity.

APPENDIX G: PREDICTION OF SKIN SENSITISATION

For the endpoint skin sensitisation the following commercial (Q)SAR systems are available:

MultiCASE, TOPKAT and DEREK

The acceptance of *in silico* systems depends on validation studies with sufficient concordance of prediction results with experimental data. In order to assess the suitability of commercial (Q)SAR systems in the identification of skin sensitisation potential, different subsets of chemicals for which experimental data were available were evaluated with DEREK or TOPKAT.

Validation studies from literature

The BgVV validated the DEREK rulebase for identifying contact allergens against a regulatory database which contains data submitted under the procedure for notifying new chemicals within the European Union. The BgVV database includes 1039 chemicals that have reliable data for the assessment of a relevant skin sensitising potential (Zinke *et al*, 2002).

Table G.1: Assessment of skin sensitising potential

	Predicted as positive	**Predicted as negative**	**Σ**
Positive in exp.	150	253	403
Negative in exp.	85	541	636
Σ	235	794	1039

Concordance: 67 %
Sensitivity: 37 % (=> false negative prediction by DEREK)
Specificity: 85 %

The results of another example for a DEREK validation study are summarised below. 79 chemicals have undergone LLNA-testing and DEREK prediction (Seaman *et al*, 2001).

Concordance: 55 – 77 %
Sensitivity: 79 % (LLNA-pos. correct)
Specificity: 51 % (LLNA-neg. correct)

Compounds from IUCLID

The application of DEREK (v 5.01) and TOPKAT (v 5.01) for predicting skin sensitisation potential has also been examined using a set of 80 substances from the IUCLID database, for which guinea pig maximisation test results have been published. The results from the DEREK and TOPKAT predictions are presented in Tables G.2 and G.3 and summarised below:

Table G.2: DEREK results

Skin sensitisation	DEREK predicted as positive	DEREK predicted as negative	Σ
Positive in PGMT	20	12	32
Negative in PGMT	18	30	48
Σ	38	42	80

Concordance: (20+30)/80 = 62.5%
Sensitivity: 20/32 = 62.5%
Specificity: 30/48 = 62.5%

Table G.3: TOPKAT results with same data set (for comparison)

Skin sensitisation	TOPKAT predicted as positive	TOPKAT predicted negative	Σ
Positive in PGMT	23	9	32
Negative in PGMT	23	25	48
Σ	46	34	80

Concordance: (23+25)/80 = 60.0%
Sensitivity: 23/32 = 71.8%
Specificity: 25/48 = 52.1%

For the chemicals in this select group, predictions were in concordance with 60-63% of the sensitisers and non-sensitisers using DEREK (v 5.01) or TOPKAT (v 5.01). TOPKAT appears to predict slightly more false-positives compared to DEREK for this particular set of select chemicals.

Henkel compounds

A total of 82 chemicals (mostly aromatic amines) were validated with the DEREK for Windows 3.6.0 software. Previously, these chemicals had undergone experimental testing using guinea pig maximisation test and/or Buehler test (BT).

Overall, the predictions of DEREK were in concordance with about 42 % of the sensitisers and non-sensitisers when compared to the results of both test types or to the results of each test system. For the group of chemicals evaluated, the DEREK software was over-predictive for the endpoint skin sensitisation which was shown by many false-positive predictions. The current DEREK rules are not sufficiently developed for predictions within a group of chemicals with suspicious structural elements. Examination of the structures of false-positives and false-negatives should help to improve the DEREK rules with the aim of better predictivity. 85 % of the alerts were fired by two rules. Revision of these rules should improve the predictivity significantly (Delbanco, 2002).

Table G.4: DEREK results

	Predicted as positive	**Predicted as negative**	**Σ**
Positive in PGMT	27	1	28
Negative in PGMT	40	3	43
Σ	67	4	71
Positive in BT	4	0	4
Negative in BT	13	2	15
Σ	17	2	19

Concordance: 42 % (GPMT) / 40 % (BT)
Sensitivity: 96 % / 100 %
Specificity: 7 % / 13 % (=> false positive prediction by DEREK)

These chemicals were also investigated with TOPKAT (version 5.01) to compare the predictivity for the endpoint skin sensitisation. Chemicals being out of prediction space (OPS) are excluded. Overall, the predictions of TOPKAT were in concordance with 53 - 59 % of the sensitisers and non-sensitisers when compared to the results of both test types or to the results of each test system. For the group of chemicals evaluated, the TOPKAT software was over-predictive for the endpoint skin sensitisation which was shown by enhanced false-positive predictions.

Table G.5: TOPKAT results

	Predicted as positive	**Predicted as negative**	**Σ**
Positive in PGMT	20	3	23
Negative in PGMT	21	15	36
Σ	41	18	59
Positive in BT	4	0	4
Negative in BT	7	4	11
Σ	11	4	15

Concordance: 59 % (GPMT) / 53 % (BT)
Sensitivity: 87 % / 100 %
Specificity: 42 % / 36 % (=> false positive prediction by TOPKAT)

APPENDIX H: PREDICTION OF ACUTE AND CHRONIC ORAL TOXICITY

H.1. Applicability of the TOPKAT rat oral LD_{50} module for the prediction of acute toxicity

The rat oral LD_{50} module of TOPKAT comprises 19 quantitative-structure-toxicity relationship models. These models are derived from experimental LD_{50} values of approximately 4,000 chemicals from open literature (RTECS). As generally the most toxic value was used when multiple data existed, the model is meant to overestimate the toxicity. The results are reported in chemical weight/body weight units. A 95% confidence limit is also calculated.

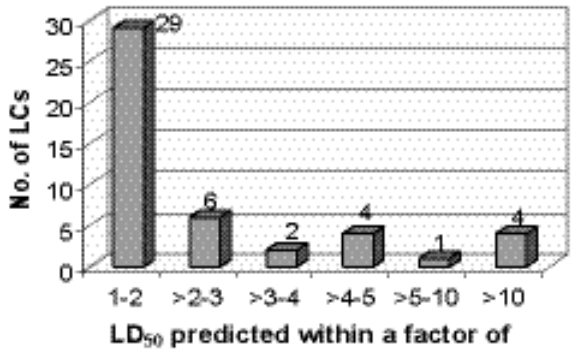
For the evaluation of the performance of the rat oral LD_{50} module, 45 proprietary pharmaceutical compounds and 144 mostly proprietary liquid crystal compounds were selected.

a) Liquid crystal compounds

The performance of TOPKAT for acute toxicity of 144 liquid crystal compounds is summarised in Table H.1. The predictions were of satisfactory reliability (33%) for only 47 of the liquid crystals. For the remaining 97 compounds, structural features were not represented in the training set to an adequate degree and the assessments did not meet all of the program's validation criteria and were excluded from the assessment.

Table H.1: TOPKAT predictions with acute toxicity module for liquid crystal compounds (LCs)

	LCs predicted within a factor of					
Factor	≤2	≤3	≤4	≤5	≤10	>10
No of LCs	29	35	38	42	43	4
[%]	62%	74%	81%	89%	91%	



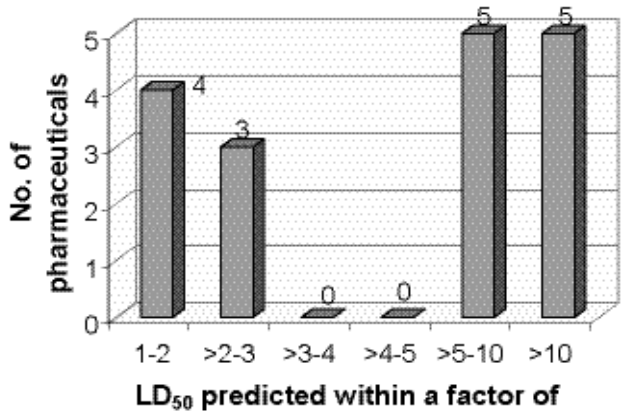
For 62% of those compounds for which a valid prediction was possible, the predicted LD_{50} value was correct within a factor of ≤2. 89% of the calculated LD_{50} values are within a factor of ≤5. For 4 LCs the predictions were >10 compared to the experimental value.

b) Pharmaceutical compounds

TOPKAT predictions for acute toxicity were calculated for 45 proprietary pharmaceutical structures (Table H.2). For 28 compounds, the validity of the prediction was insufficient (e.g. outside OPS). Thus the predictions were only considered reliable for 17 out of the 45 compounds (38%).

Table H.2: TOPKAT predictions with acute toxicity module for pharmaceutical compounds

	Pharmaceuticals predicted within a factor of			
Factor	2	3	10	>10
Number of pharmaceuticals	4	7	12	5
[%]	24%	41%	67%	29%



H.2. Applicability of the TOPKAT chronic LOAEL module for the prediction of chronic toxicity

To predict chronic effects, TOPKAT calculates a LOAEL. All data used were from oral rat chronic studies of at least 1 years duration (Mumtaz *et al*, 1995).

The computed chronic LOAEL value in the rat in weight/body weight units, along with 95% confidence limits. In order to avoid spurious results, the LOAEL models check whether the query structure is within the range of the octanol-water partition coefficients of the compounds from which the LOAEL model was developed, as well as the estimated LD_{50} (rat oral) for the query structure to check whether the LOAEL is more toxic than the LD_{50}. If either checks reveals a conflict, appropriate messages are issued by the system.

Chemical compounds from IUCLID-database

Eleven compounds from the IUCLID-database with experimental LOAEL values from oral >1year-studies were used to predict chronic LOAEL with TOPKAT.

7 out of the 11 compounds were in the training set of the module (2 of which were statistical outliers, not used for model development). Descriptors for one compound were outside the OPS and thus the prediction was considered unreliable.

For the remaining 3 compounds the following results were obtained:

Table H.3: Prediction of chronic toxicity with TOPKAT

Compound	CAS no.	Experimental LOAEL	Predicted LOAEL (95% c.i.)
Chlorocresol	59-50-7	500 mg/kg	92.2 mg/kg (17.9-475.5 mg/kg)
Cyanamide	420-04-2	7.5 mg/kg	16.5 mg/kg (3.5-78.2 mg/kg)
Pyridate	55512-33-9	115 mg/kg	43.2 mg/kg (9.9-188 mg/kg)

APPENDIX I: PREDICTION OF TERATOGENICITY

Pearl *et al* (2001) compared the predictivity of MultiCASE, TOPKAT and DEREK for a set of 105 compounds tested in *in vivo* rodent teratogenicity studies (34 positive rodent teratogens, 71 negative rodent teratogens; Table I.1). All three systems have an excessively high rate of false negative predictions and do not adequately identify the teratogenic compounds.

Table I.1: Teratogenicity predictions (Pearl et al, 2001)

System	Concordance	False positive	False negative
DEREK (3.4/3.6)	72%	0%	28%
MultiCASE (3.45, A49)	66%	7%	27%
TOPKAT(5.01)[1]	50%	35%	15%
DEREK and MultiCASE[2]	71%	7%	22%

[1] Includes predictions outside OPS

[2] DEREK negative predictions verified by MultiCASE

DEREK only identified 4 retinoic acids correctly as teratogens (sensitivity: 12%). Of the 18/34 compounds that TOPKAT correctly detected as teratogens only 7 were within the Optimum Prediction Space (sensitivity: 53% and 21%, respectively). The combination model DEREK and MultiCASE found 10/34 positive *in vivo* teratogens (sensitivity: 34%).

APPENDIX J: PREDICTION OF MUTAGENICITY

1. Prediction of drugs and non-drug chemicals with DEREK, TOPKAT and MultiCASE

Pearl *et al* (2001) compared the performance of DEREK (version 3.4 or 3.6), MultiCASE (version 3.45) and TOPKAT (version 5.01) for predicting mutagenicity. The compounds selected for the study were taken from open literature and US-government toxicity databases. The results are shown in Tables J.1 and J.2.

Table J.1: Ames mutagenicity validation with 123 drugs[1] (Pearl et al, 2001)

Model	Concordance	False Positive	False Negative	Indeterminate
DEREK	61%	31%	8%	0%
MultiCASE (A2H)	72%	15%	13%	3%
TOPKAT	67%	15%	18%	2%
DEREK and MultiCASE[2]	75%	19%	6%	3%
DEREK and MultiCASE and TOPKAT in agreement	86%	9%	5%	64%

[1] 49 Ames positive, 74 Ames negative

[2] DEREK positive predictions preceded by MultiCASE predictions

The DEREK, MultiCASE and TOPKAT consensus model provided the best concordance but had the lowest coverage and was applicable for approximately one-third of the compounds examined, only.

Table J.2: Ames mutagenicity validation with 516 non-drug compounds[1] (Pearl et al, 2001)

Model	Concordance	False Positive	False Negative	Indeterminate
DEREK	70%	24%	6%	0%
MultiCASE (A2H)	81%	12%	7%	2%
TOPKAT	56%	19%	25%	3%
DEREK and MultiCASE[2]	82%	16%	2%	2%
DEREK and MultiCASE and TOPKAT in agreement	83%	10%	7%	57%

[1] 285 Ames-positive, 231 Ames-negative

[2] DEREK positive predictions preceded by MultiCASE predictions

The drug validation set has approximately half as many Ames-positives as negatives while the non-drug validation set contains approximately equal number of positives and negatives. This bias for negative Ames mutagens in the drug validation artificially enhances the concordance of programs that are biased for negative predictions while reducing the concordance for programs biased for predicting positive mutagens (Pearl *et al*, 2001).

2. Prediction of pharmaceutical compounds using DEREK and TOPKAT

Cariello *et al* (2002) evaluated 414 compounds with DEREK (version 17.1, Java client) and TOPKAT (version 5.01) to predict bacterial mutagenicity. With TOPKAT, 111 out of 414 were excluded from the analysis because they were either out of OPS or produced probability estimates in the indeterminate region and the prediction of mutagenicity for these compounds is considered unreliable.

Table J.3: TOPKAT results

		Experimental data		
		Mutagen	**Non-mutagen**	**Total**
TOPKAT-prediction	Mutagen	21	49	70
	Non-mutagen	32	201	233
	Total	53	250	303

Sensitivity: 40%, Specificity: 80%, Concordance: 73%

DEREK could not process 5 of the 414 compounds since their molecular weight was too high. The overall performance of the program based on 409 compounds is shown in Figure J.4. Certain structural alerts were prevalent in discordant results for chemicals which were negative in the Ames assay but produced an alert. In particular, the rule for aromatic amine mutagenicity (rule 827) was triggered 43 times in the Ames-negative data. Other rules that were triggered inappropriately were α,β-unsaturated amide (rule 821, 10 instances) and aromatic amine with diaryl fusion (rule 848, 10 instances). The rule for aromatic amines was the most frequently triggered rule in both the discordant (43 instances) and concordant (9 instances) data sets. This illustrates one of the difficulties of simplistically applying a rule-based-approach (Cariello *et al*, 2002).

Table J.4: DEREK results

		Experimental data		
		Mutagen	**Non-mutagen**	**Total**
DEREK-prediction	Mutagen	38	101	139
	Non-mutagen	44	226	270
	Total	82	327	409

Sensitivity: 46%, Specificity: 69%, Concordance: 65%

Neither DEREK nor TOPKAT performed particularly well in the Ames-positive category. While the overall concordance of the TOPKAT program was higher than DEREK, TOPKAT fared more poorly in the Ames-positive category than DEREK. When both programs are used in conjunction, about one-quarter (26%) Ames-positive molecules are missed and the false positive rate is increased (Sensitivity: 16%, Specificity: 45%, Table J.5, Cariello *et al*, 2002).

Table J.5: Compounds where DEREK and TOPKAT predictions are in agreement (Cariello et al, 2002)

	# predicted positive	# predicted negative
Ames-positives: 82	13	21
Ames-negatives: 327	21	146

3. Prediction of bacterial mutagenicity of 500 proprietary Novartis pharmaceutical candidate compounds using DEREK

Table J.6: Prediction table using rules as defined in DEREK version 5.01

		DEREK prediction		
		Mutagen	**Non-mutagen**	**Total**
Ames tests	Mutagen	27	63	90
	Non-mutagen	52	358	410
	Total	79	421	500

Sensitivity (correctly predicted positives to total number of experimental positives): 27/90 = 30%
Specificity (correctly predicted negatives to total number of experimental negatives): 358/410 = 87%
Concordance (correctly predicted positives and negatives to total number of experimentally tested compounds): (27 + 358) / 500 = 77%

Table J.7: Prediction table using the same set of data but with incorporation of new (Novartis-specific) rules for mutagenicity into DEREK

		DEREK prediction		
		Mutagen	**Non-mutagen**	**Total**
Ames tests	Mutagen	74	16	90
	Non-mutagen	107	303	410
	Total	181	319	500

Sensitivity (correctly predicted positives to total number of experimental positives): 74/90 = 82%
Specificity (correctly predicted negatives to total number of experimental negatives): 303/410 = 73%
Concordance (correctly predicted positives and negatives to total number of experimentally tested compounds): (74 + 303) / 500 = 75%

The DEREK prediction of bacterial mutagenicity of Novartis structures was of unacceptably low sensitivity (30%) when using the original DEREK rules. There was a dramatic improvement in sensitivity to 82% with the incorporation of additional rules that seem to adequately reflect aspects of genotoxicity that are specific to Novartis structures and that are not adequately reflected in the open literature. At the same time the incorporation of these additional rules results in a loss of specificity as reflected by a decrease from 87% to 73%. With the incorporation of new rules, DEREK was assessed as useful for SAR analysis of Novartis structures for genotoxicity.

4. Prediction of bacterial mutagenicity of 169 new proprietary Novartis pharmaceutical candidate compounds using DEREK, version 5.01

Table J.8: Prediction table using DEREK including new (Novartis-specific) rules for mutagenicity

		DEREK prediction Mutagen	Non-mutagen	Total
Ames tests	Mutagen	16	8	24
	Non-mutagen	50	95	145
	Total	66	103	169

Sensitivity (correctly predicted positives to total number of experimental positives): 16/24 = 66.7%
Specificity (correctly predicted negatives to total number of experimental negatives): 95/145 = 65.5%
Concordance (correctly predicted positives and negatives to total number of experimentally tested compounds): (16 + 145) / 169 = 95.3%

Table J.9: Prediction table using MultiCASE mutagenicity modules

		DEREK prediction Mutagen	Non-mutagen	Total
Ames tests	Mutagen	12 $_{9+\ ;\ 3+\ ?}$	12 $_{7-\ ;\ 5-}$	24
	Non-mutagen	35 $_{21+\ ;\ 14+\ ?}$	110 $_{68-\ ;\ 42-}$	145
	Total	47	122	169

Sensitivity (correctly predicted positives to total number of experimental positives): 12/24 = 50%
Specificity (correctly predicted negatives to total number of experimental negatives): 110/145 = 75.9%
Concordance (correctly predicted positives and negatives to total number of experimentally tested compounds): (12 + 110) / 169 = 72.2%

Table J.10: Prediction table using DEREK and MultiCASE combined

		DEREK prediction Mutagen	Non-mutagen	Total
Ames tests	Mutagen	18	6	24
	Non-mutagen	80	65	145
	Total	98	71	169

Sensitivity (correctly predicted positives to total number of experimental positives): 18/24 = 75%
Specificity (correctly predicted negatives to total number of experimental negatives): 65/145 = 44.8%
Concordance (correctly predicted positives and negatives to total number of experimentally tested compounds): (18 + 65) / 169 = 49.1%

The evaluation of the prediction for new Novartis structures shows that the improved rule-based version of DEREK shows considerably higher sensitivity than the MultiCASE mutagenicity modules. This and the fact that a high percentage of Novartis structures had features that were not covered by MultiCASE illustrates the need for more comprehensive MultiCASE Salmonella mutagenicity modules that would have an increased coverage of pharmaceutical-like structures. The combination of both, DEREK and MultiCASE results in an increased sensitivity, but, as expected, results also in a considerable loss in specificity.

5. Prediction of a selection of 44 aromatic amines using DEREK and MultiCASE

44 simple aromatic amines have been studied for mutagenicity in *Salmonella typhimurium* TA98 and TA100 in the presence and absence of 10% aroclor-induced rat liver S9 mix. The following tables give the prediction results for this selection of compounds for DEREK, MultiCASE and its combination:

Table J.11: DEREK prediction

		DEREK prediction		
		Mutagen	**Non-mutagen**	**Total**
Ames tests	Mutagen	18	3	21
	Non-mutagen	11	12	23
	Total	29	15	44

Sensitivity (correctly predicted positives to total number of experimental positives): 18/21 = 85.7%
Specificity (correctly predicted negatives to total number of experimental negatives): 12/23 = 52.2%
Concordance (correctly predicted positives and negatives to total number of experimentally tested compounds): (18 + 12) / 44 = 68.2%

Table J.12: MultiCASE prediction

		MultiCASE prediction		
		Mutagen (+ and +?)	**Non-mutagen**	**Total**
Ames tests	Mutagen	17 $_{13+\ ;\ 4+?}$	4 $_{\text{one not covered}}$	21
	Non-mutagen	13 $_{8+\ ;\ 5+?}$	10	23
	Total	30	14	44

Sensitivity (correctly predicted positives to total number of experimental positives): 17/21 = 81%
Specificity (correctly predicted negatives to total number of experimental negatives): 10/23 = 43.5%
Concordance (correctly predicted positives and negatives to total number of experimentally tested compounds): (17 + 10) / 44 = 61.4%

Table J.13: Prediction of DEREK and MultiCASE combined

		DEREK + MultiCASE prediction		
		Mutagen	**Non-mutagen**	**Total**
Ames tests	Mutagen	21	0	21
	Non-mutagen	17	6	23
	Total	38	6	44

Sensitivity (correctly predicted positives to total number of experimental positives): 21/21 = 100%
Specificity (correctly predicted negatives to total number of experimental negatives): 6/23 = 26.1%
Concordance (correctly predicted positives and negatives to total number of experimentally tested compounds): (21 + 6) / 44 = 61.4%

For the aromatic amines under question, DEREK and MultiCASE show similar values with regard to sensitivity, specificity and concordance. The combination of both systems yields 100% sensitivity, whereas the specificity is reduced to less than 30%. One example may serve to illustrate that prediction should not normally be taken at face value, rather the data, on which the prediction was based should be critically analysed.

O-cresidine was part of the testing program of aromatic amines and was tested negative at Novartis using *Salmonella typhimurium* TA98 and TA100 in the presence of 10% aroclor induced rat liver S9 mix. However, it was predicted positive by MultiCASE because the data that were used to build the MultiCASE module were positive literature data in the presence of 30% S9 mix. At the same time, the literature reference indicates negative data for the use of 10% S9 mix.

6. Evaluation of 27 mutagenicity studies performed in Bayer Toxicology with DEREK

Table J.14: DEREK prediction

		DEREK prediction		
		Mutagen	**Non-mutagen**	**Total**
Ames tests	Mutagen	4	0	4
	Non-mutagen	9	14	23
	Total	13	14	27

Sensitivity: 100%, Specificity: 61%, Concordance: 67%

This set of compounds was also used for carcinogenicity evaluation (see Appendix I).

7. Evaluation of Ames studies from IUCLID with DEREK

For 54 compounds the experimental data on mutagenicity were extracted from IUCLID. This set of compounds was also used for carcinogenicity prediction (see Appendix I).

Table J.15: DEREK prediction

		DEREK prediction		
		Mutagen	**Non-mutagen**	**Total**
Ames tests	Mutagen	3	0	3
	Non-mutagen	8	43	51
	Total	11	43	54

Sensitivity: 100%, Specificity: 84%, Concordance: 85%

8. Evaluation of 247 chemical compounds from literature with TOPKAT

100 compounds which were not part of the TOPKAT training set were selected from the open literature and evaluated for mutagenicity prediction. For 14 compounds the predictions were considered unreliable because they were outside OPS. A further 3 compounds were excluded because the predictions were inconclusive (probability for mutagenicity between >0.3 and <0.7) (Mueller *et al*, 2000).

Table J.16: TOPKAT prediction

		TOPKAT prediction		
		Mutagen	**Non-mutagen**	**Total**
Ames tests	Mutagen	18	3	21
	Non-mutagen	17	45	62
	Total	35	48	83

Sensitivity: 86%, Specificity: 73%, Concordance: 76%

APPENDIX K: PREDICTION OF CARCINOGENICITY

The first exercise to predict carcinogenicity conducted by the NTP involved 44 chemicals. The predictions made by Multi-CASE (Rosenkranz and Klopman, 1990), TOPKAT (Enslein *et al*, 1990) and DEREK (Sanderson and Earnshaw, 1991) were published in advance of the bioassays being performed. The predictions were then compared to the results from the bioassays.

The overall results for the commercial prediction systems as published by Benigni (1997) (Table K.1) indicate that none of the commercially available programs performed to the desired level of accuracy for the set of unknown compounds. By way of comparison, toxicology experts (Ashby and Tennant), managed to correctly predict 75% of the compounds, significantly out-performing the computer systems.

Table K.1: Results of predicting carcinogenicity from the NTP 44 chemicals

System	Overall accuracy (%)
DEREK	59
TOPKAT	57
Multi-CASE	49
Expert panel	75

A subsequent, similar exercise was proposed involving a further 30 chemicals selected for testing by the National Toxicology Programme (NTP) (Bristol *et al*, 1996). Predictions from Multi-CASE (Zhang *et al*, 1996), DEREK (Marchant, 1996) and OncoLogic (Woo *et al*, 1997) were published in advance of the bioassays being performed.

The results for 26 of these chemicals, for which testing is complete, have been peer reviewed and the preliminary results of the performance of the commercial systems published (Benigni, 2000). The overall concordance figures (Table K.2), i.e. the percentage of correctly identified negative and positive carcinogens, show the OncoLogic system to be leading the field of commercial systems.

Table K.2: Preliminary results of commercial systems for the NTP 30 chemicals exercise

System	Overall accuracy (%)
OncoLogic	67
DEREK	38
MultiCASE	18

In the study by Prival (2001), the utility of the TOPKAT NTP carcinogenicity module was evaluated by determining the system's ability to predict the results of rodent carcinogenicity bioassays conducted by the NTP.

Table K.3: Predictivity of the TOPKAT-NTP-module (version 5.01)[a] (Prival, 2001)

	Oral	Oral + Inhalation	Oral + inhalation + dermal
Sensitivity	0.18	0.33	0.31
Specificity	0.80	0.78	0.80
Positive predictivity	0.45	0.64	0.63
Negative predictivity	0.51	0.49	0.51
Concordance	0.50	0.53	0.54

[a] predictions using <30% probability as predicting non-carcinogenicity and >70% probability as predicting carcinogenicity; excluding data found outside 'OPS'

TOPKAT was not effective in identifying potential rodent carcinogens and noncarcinogens in the data set analysed (Prival, 2001).

Durham *et al* (2001) compared the predictive strengths and weaknesses of the three commercial programs TOPKAT, MultiCASE and DEREK for the assessment of *in vivo* rodent carcinogenic properties of pharmaceutical compounds. The 142 drug compounds for this validation study were taken from public literature and US-governmental toxicity databases (71 positive, 71 negative).

Table K.4: Predictivity of DEREK, MultiCASE and TOPKAT (Durham et al, 2001)

Model	Concordance	False Positive	False Negative	Indeterminate	Sensitivity	Specificity
DEREK[1]	57%	30%	13%	0%	75%	40%
MultiCASE[2]	74%	17%	8%	1%	85%	65%
TOPKAT[3]	60%	31%	4%	5%	87%	32%

[1] Version 3.6 or 3.4

[2] Version 3.45, AF-5-8

[3] Version 5.01, NTP-Carcinogenicity Module (requires 1 positive model for carcinogen)

Additionally, the MultiCASE carcinogenicity correlation for 141 drugs (71 positive, 70 negative) with the number of sex/species models was analysed (Table K.5).

Table K.5: MultiCASE Carcinogenicity Correlation vs. # of positive models

# Positive Models	# of Compounds	Concordance	False positive	False negative
0	40	85%	-	15%
1	15	33%	67%	-
2	32	56%	44%	-
3	17	71%	29%	-
4	37	86%	14%	-

A paper addressing the performance of the enhanced MultiCASE-FDA-module for carcinogenicity (AF5-8) was published by Matthews and Contrera (1998) (see Table K.6). The 126 compounds were selected from contemporary pharmaceuticals submitted to the authority, 27 compounds studied in the NTP programme and 54 compounds from literature.

Table K.6: Predictivity of FDA-OTR/MultiCASE proprietary database module (Matthews and Contrera, 1998)

		Experimental data		
	Carcinogenic	**Non-carcinogenic**		**Total**
MultiCASE prediction	Carcinogenic	34	1 falsely positive	35
	Non-carcinogenic	24 falsely negative	41	65
	Total	58	42	100

Compounds with unreliable prediction: 26/126

Predictability: 100/126 = 79%

Sensitivity (carcinogens correctly identified): 34/58 = 59%

Specificity (non-carcinogens correctly identified): 41/42 = 98%

Concordance (correctly predicted positives and negatives to total number of experimentally tested compounds): (34 + 41) / 100 = 75%

Table K.7: Evaluation of 30 chronic rat studies performed by Bayer Toxicology

		DEREK		
		Carcinogenic	**Non-carcinogenic**	**Total**
Carcinogenicity	Carcinogenic	3	2	5
	Non-carcinogenic	17	6	23
	Total	20	8	28

Sensitivity: 60%, Specificity: 26%, Concordance: 32%

All studies were performed according to GLP between 1981 and 1989. The 30 studies were performed on 28 different compounds (21 agrochemicals, 6 pharmaceutical chemicals, 1 organic chemical). The information, if MTD was reached in the study and the mechanism for carcinogenicity was primary or secondary, were important for the evaluation. Parallel to the carcinogenicity the mutagenicity (Ames test) was also evaluated (see Appendix J).

It is obvious that the concordance for the prediction of the carcinogenicity was insufficient (see Tables K.6 and K.7).

Table K.8: Evaluation of carcinogenicity studies from IUCLID

		DEREK prediction		
		Carcinogenic	**Non-carcinogenic**	**Total**
Carcinogenicity	Carcinogenic	15	7	22
	Non-carcinogenic	15	24	39
	Total	30	31	61

Sensitivity: 68%, Specificity: 62%, Concordance: 64%

The experimental data on carcinogenicity for 61 compounds were extracted from IUCLID. All studies were performed according to GLP. It was very difficult and in some cases not possible to get the right information from IUCLID, especially for the assessment of the endpoint carcinogenicity.

Parallel to the carcinogenicity, the mutagenicity (Ames-test) of 54 substances was also evaluated (see Appendix J).

The prediction for this organic chemical based data set is better than for drugs or other active substances (see Tables K.8 and K.9).

Table K.9: Evaluation of carcinogenicity studies from IUCLID

substance	mutagenicity															carcinogenicity			target-organ (dose)	MTD reached
	DEREK	in vitro									in vivo					DEREK	primary	secondary		
		Sac	SMT	Pol A	HGPRT	MLA	EMT	UDS	DNA	Cyt	MNT	DNA	SMT	UDS	DLT					
1	-		-			-	-			-	-	+			-	-	-		(0-2000ppm)	+
2	-		2-	2-			-	-		+	-			-	-	+	-	+	(0-2500ppm)	(+)
3	+		2+	-							-				-	+	-		(0-50ppm)	
4	+	+	+			+		-	-		-				-	+	-		(0-25ppm)	+
5	-	-	-	-			-	-		-	-				-	+	-		- (5-45ppm)	+
6	-	-	-								-	+		+	-	+	-		(0-1500ppm)	+
7	+		+/3-			- (+)										+	-		(0-250ppm)	+
8	-		-	-	-			-		-	-		-	-	-	-	-		(0-2500ppm)	
9	+		-	-					-	-	-				-	+	-		(0-32ppm)	
10	-		-								-				-	-	-	(?)	(0-4000 ppm) (0-1000 ppm)	+
11	+	no Ames-Test available									-					+	-		(0-100 ppm) (0-2500 ppm)	+
12	-		-		-		-	-	-	-	-				-	+	-		(0-4000 ppm) (0-1000 ppm)	m + / f -
13	-		-	-		-	-		-		-				-	+	-		(0-1000 ppm)	(+)
14	+		-				-	-		-/+	us				-	+	-		(0-25ppm) 0-100ppm)	(-)
15	+		-	-	-						-					+	-		(0-25ppm) (0-100 ppm)	(-)
16	+		-						+		-	-			-	+	-/-		(0-2500ppm)	-
17	+		-		-			-		+	-					+	-	(?)	(0-500ppm) (0-6000ppm)	+
18	-		-		-			-		(+)	(+)				2-	+	-		(0-1000 ppm) (0-4500 ppm)	+
19	+		-						-		-					+	-		(0-25ppm)	m - / f +
20	+		+				+		2+	+	2-				-	+	-	+ thyroids	(0-4500 ppm)	+
21	-	-	2-			2-	-	-	-	+	-				-	-	+		thyroids (0-3600 ppm)	+
22	+	-	-				-	-		-	-				-	+	-		(0-500ppm)	-
23	+		-				-		-	+	-					+	-		(0-500ppm) (0-1800ppm)	-
24	-	-	-		-					-	-	2-		-		+	-	+ bladder + liver	(0-2000ppm)	m - / f (-)
25	-	-	-	-		-	-	-		-	-				-	-	-	+ thyroids	(0-1800 ppm)	+
26	+	-/+	2+	-		-	-	+	+		-				-	+	- -		(0-1)ppm) (0-32ppm)	+
27	-		+/2-		-			-			-					-	-		(0-10000ppm)	+
28	-		-						-	-/+	-				-	+	- -		(0-2000ppm)	+

SMT = Salmonella-microsome-test
Sac = Saccharomyces cerevisiae
DLT = Dominant-lethal-test (mouse) (rat only BAY q 7621)*
Cyt = Cytogenetic test
MLA = Mouse Lymphoma cell reverse mutation system
EMT = E.coli microsome test
MNT = Micronucleus-test
UDS = hepatocyte unscheduled DNA synthesis assay

us = unsuitable test

APPENDIX L: Databases

L.1 IUCLID

A potential source of data for developing or validating (Q)SARs is the **I**nternational **U**niform **C**hemica**L** **I**nformation **D**atabase, IUCLID. IUCLID is the database management software that was developed by the European Commission Services. The IUCLID format was initially designed as an electronic repository to store, organise, exchange and use the information collected under the EC Council Regulation EEC/793/93 on the Evaluation and Control of the Risks of Existing Substances. The data submitted to the Commission are used for assessment of the risks to human health and the environment, evaluation and control of existing substances (EC, 1993).

Among the data collected are physico-chemical, toxicological and ecotoxicological properties, as well as information related to environmental fate.

The latest update of the non-confidential version of IUCLID (version 4.0 database, February 2000) (EC, 2000a) includes approximately 2,600 chemicals.

The following section evaluates some of the data collected in the IUCLID database and whether these could be used to enlarge data sets of (Q)SAR models, or as a test set for external validation of (Q)SARs.

Limiting factors are the quality and the inherent variability of the data. Sources of variability include:

- Use of different methods (test system, protocol) for one and the same endpoint;
- heterogeneity of data sets;
- variability in chemical composition (possible presence of impurities) for apparently similar substances.

The query module provided with the software includes several standard queries using the SQL language. However, only a few of them are concerned with toxicological data. Additional simple queries were therefore designed using several selected parameters to screen toxicological endpoints for data of sufficient quality for use in a training or test set for (Q)SAR models.

L.1.1 Chemical classes

Most of the (Q)SAR models are developed for organic chemicals and do not apply to other specific classes, such as metals, inorganics, salts or mixtures. Among the chemicals data sets reported in IUCLID, only those responding to the specific chemical classes covered by each (Q)SAR models should be selected.

The following classes are represented in IUCLID (Allanou *et al*, 1999):

- Discrete organics;
- inorganics;
- mixtures of discretes;
- organo-metallics;
- coal and petroleum-based;
- UVCB (unknown or variable composition, complex reaction products or biological materials).

Organics, which are the most frequently used chemicals in the design of (Q)SAR models, represent approximately 50% of the chemical data sets in IUCLID.

The present evaluation did not discriminate for specific classes in the queries (the relevant data-table was not accessible in the query module). This criterion will have to be included in a further evaluation, e.g. specific to each (Q)SAR model, or by adding a criterion in the queries.

L.1.2 Quality criteria

The following criteria were used to screen the IUCLID data:

- Were the data assessed for reliability?
- were the data obtained in studies performed in accordance to Good Laboratory Practices?
- methodology: a source of variability in the data sets of QSAR models may arise from the differences in the methods used for generating the data. For the purpose of the present exercise, only records with reference to OECD guidelines and/or EC standard methods were selected;
- balance of data: the distribution of the values of the descriptors.

Reliability

There are four levels of reliability in the more recent version of IUCLID, which correspond to criteria described by Klimisch *et al* (1997):

1. Valid without restriction;
2. Valid with restriction;
3. Invalid;
4. Not assignable.

Levels 1 and 2 are the more appropriate to select the most reliable data.
In the IUCLID 4.0 database, the query results indicate that a total of 1466 records currently have a reliability criteria assigned (see Table L.1):

- 837 with reliability information (1 to 4) (regardless of information on GLP)
- 515 with reliability 1 or 2 (regardless of information on GLP)
- 114 with reliability 1 or 2 *and* GLP

These results indicate that the reliability criterion is poorly represented in the IUCLID database. This is explained by the recent inclusion of the parameter in the database. Therefore, this criterion cannot be used at this time as a selective parameter to retrieve a sufficient number of reliable data. This figure might improve in the future, with a more systematic assignment of validity criteria associated with the review of the chemicals targeted by the HPV programme, the update of data sets and the development of the proposed EC REACH programme.

GLP studies

A number of records in the IUCLID database were indicated as compliant with GLP. For the toxicological endpoints, usually between 32 and 53 % of data records were GLP studies, except for carcinogenicity data for which only approximately 20% were GLP studies, possibly because this type of studies were relatively old, or literature data.

In the subsequent searches of this assessment exercise, only GLP studies were selected (no query were made with GLP= 'No' or GLP = 'no data').

Experimental test methods

For some endpoints, several methods may be applicable. A glossary of standard methods is available in the IUCLID software to help the users during the data entry. The query for IUCLID records relative to acute toxicity, repeated dose toxicity and carcinogenicity showed that the majority of records for GLP studies are generally referring to OECD standardised guidelines.

In contrast, very few IUCLID records refer to US EPA methods for toxicological endpoints, although they are available in the glossary.

Table L.1: Number of IUCLID records with a reliability criterion assigned

		Acute Oral	**Acute Dermal**	**Acute Inhalation**	**Skin Irritation**	**Eye Irritation**	**Sensitisation**	**Genetic *in vitro***	**Genetic *in vivo***	**Repeat Dose Toxicity**	**Carcinogenicity**	**Reprotoxicity**	**Developmental Toxicity**	**Total hits**
GLP = Yes	Reliability 1	7	5	3	11	10	6	16	10	14	5	3	9	
	Reliability 2	1	0	2	2	0	1	4	2	2	1	0	0	
	Total	8	5	5	13	10	7	20	12	16	6	3	9	114
GLP = Yes, No and No Data	Reliability 1	20	7	6	20	20	8	36	17	24	10	7	13	
	Reliability 2	49	21	38	44	39	21	34	15	29	13	7	17	
	Total hits	69	28	44	64	59	29	70	32	53	23	14	30	515
GLP =	Reliability 1 to 4	107	45	69	101	84	57	107	54	94	42	21	56	837

L.1.3 Human health endpoints

Acute toxicity

In the acute oral toxicity section, 326 records out of 408 (almost 80%) referred either to OECD Guideline No. 401 or to EC Method B.1 which is described in Directive 84/449/EEC. Logically, the more recent methodologies (OECD Guidelines No. 420, 423 and 425 and the corresponding EC methods) were not used at the time the Phase I and Phase II HEDSET were compiled and no GLP studies for high volume existing substances refered to these methods.

Similarly, nearly 79% of the acute dermal toxicity data and 72% of the acute inhalation toxicity data refered to OECD guideline or EC methods (Table L.2).

In each section, the remaining records used the glossary term 'other' and it is clear that they should not be selected for building up a (Q)SAR data set, as it is assumed that the information available on the method is either absent or limited.

Table L.2: Number of IUCLID Records Using Standardised Test Procedures for Acute Toxicity

Acute Toxicity Endpoint	Method	Number of hits	Comments
Oral	OECD 401	273	No records were found with a reference to OECD 420, 423 or 425.
	B.1 84/449/EEC	53	No records were found with a reference to EC method B1 92/69/EEC, B1bis 92 /69/EEC or B.1tris 96/54/EEC
	US EPA	0	3 methods in the glossary
	Other	82	
Dermal	OECD 402	176	
	B.3 92/69/EEC	0	
	US EPA	0	3 methods in the glossary
	other	47	
Inhalation	OECD 403	119	
	B.2 84/449/EEC	7	No records found with a reference to EC method B2 92/69/EEC or B2 93/21/EEC
	US EPA	0	3 methods in the glossary
	Other	48	

Distribution of descriptors

Previous evaluations have shown that composition of the data set in terms of descriptors distribution can influence the predictability of the (Q)SAR models (ECETOC, 1998a).

The quality of the IUCLID database records was checked for distribution of descriptors within European hazard classification categories for the endpoints, skin and eye irritation and sensitisation.

In IUCLID, it is assumed that a 'classification' is generally based on validated results and therefore only information present in this field were retained for the queries, as opposed to the 'Results' field. As for the previous sections, queries were limited to GLP records that included a reference to standardised methods (OECD guidelines or EC methods). The other categories of the IUCLID glossary, such as 'other' and 'other (calculated)' were not taken into account in the following IUCLID queries, mainly because they are not amenable to an automated selection process. The *in vitro* category was also not selected because details on methodology were often lacking. The EC Method B.40 (corrosion assay) was validated only recently for classification of chemicals, which explains why no corresponding records were found in the database (EC, 2000b).

In the following queries, the GLP studies were screened for the number of records in each hazard classification category (selected from the glossary) and method (selected from the glossary). Tables L3, L4 and L5 show the distribution of GLP studies in each hazard classification category. The query results shows that the IUCLID database contains a larger number of substances 'not classified' as compared with substances that require a classification.

Distribution of IUCLID records among hazard classification categories for skin irritation, eye irritation and sensitisation

The query results were limited to standardised OECD guidelines and corresponding EC methods. For skin and eye irritation, test results obtained with the Draize assay were also included when classification was assigned in the database (Table L.3 and L.4). With regard to sensitisation, records referring to OECD 406 and EC B.6 refer to only Magnusson and Kligman or Buelher assays (Table L.5).

It should be mentioned that figures obtained with the queries described here may underestimate the actual data present in IUCLID, as some records were entered without using the terms available in the glossary.

Table L.3: Number of IUCLID records in each hazard classification category and using standardised test procedures for skin irritation

Method	Hazard classification			
	Not irritating	Irritating (R38)	Causes burns (R34)	Causes severe burns (R35)
OECD 404	257	113	28	4
EC B.4 (84/449/EEC)	36	15	4	0
EC B.4 (92/69/EEC)	0	0	0	0
Draize	116	81	2	0
Total hits	409	209	32	4

Table L.4: Number of IUCLID records in each hazard classification category and using standardised test procedures for eye irritation

Method	Hazard classification		
	Not irritating	Irritating (R36)	May cause severe damage to eyes (R41)
OECD 405	204	49	36
EC B.5 (84/449/EEC)	25	5	3
EC B.5 (92/69/EEC)	0	0	0
Draize	166	12	2
Total hits	395	66	41

Table L.5: Number of IUCLID records in each hazard classification category and using standardised test procedures for skin sensitisation

Method	Hazard classification	
	Not sensitising	Sensitising (R43)
OECD 406	105	42
EC B.6 (84/449/EEC)	23	9
EC B.6 (92/69/EEC)	0	0
Total hits	128	51

Distribution of IUCLID records among result categories for genetic toxicity in vitro

The IUCLID database was searched for distribution of negative and positive results in the Ames test or reverse mutation assay. Results described as 'ambiguous' were left out of the search results. The positive (42 hits) and negative (533 hits) results were distributed in the following categories of test-types: Ames test, bacterial reverse mutation assay, *Salmonella typhimurium* reverse mutation assay. Methods were generally OECD Guideline No. 471 or 472, EC B.10 or B.14 (84/449/EEC). Again, the IUCLID database contains more negative results than positive results.

Distribution of IUCLID records among result categories for specific effects

For the remaining endpoints, the IUCLID database is more difficult to evaluate with simple queries because repeated dose toxicity, carcinogenicity, toxicity to the reproduction and developmental toxicity are complex endpoints.

Of the 247 records for repeated dose toxicity studies performed in rats using the oral route (oral route includes the following categories of the IUCLID glossary: 'oral feed', 'gavage', 'drinking water' and 'oral unspecified'), only 20 GLP studies were of a 2-year duration.

With regard to the carcinogenicity endpoint, the query retrieved 79 records for GLP studies in rats via the oral route and of at least a 2-year duration. However, each record should be checked individually to assess details of the methodology, results and quality prior to use the substance in a (Q)SAR model either as part of the training set or the test set.

The ITIC database (see section L2.1) should contribute for endpoints such as carcinogenicity and mutagenicity to a more comprehensive database of reliable data that could be used to develop improved prediction models.

Conclusions for human health endpoints

The evaluation of the IUCLID database (version 4.0 updated February 2000) covering phase I and phase II existing chemicals showed that currently:

- Only a limited number of IUCLID records are available with reliability criteria; this important feature is therefore not sufficiently represented in the database to select data for (Q)SAR development;
- GLP studies and standardised methods (OECD guidelines and EC methods) are well-represented in IUCLID and can be used as criteria for selection of good quality records;

- the usually larger number of 'negative' versus 'positive' results may introduce bias in a (Q)SAR model when selecting data from IUCLID;
- IUCLID can be a data source for (Q)SAR development for relatively simple endpoints (local effects, Ames test). However, more complex endpoints require further data validation.

L.1.4 Environmental endpoints

Table L.6 presents results of a simple search of the IUCLID database for availability of data for a specific endpoint and a consistent source, i.e. one protocol. For certain endpoints there would seem to be sufficient good quality data for (Q)SAR development or validation purposes, although no further assessment has been carried out to assess the chemical domain across which data are available. It is probable that this picture will improve over the next few years as the HPV programme generates data, which should be entered into IUCLID.

Table L.6: Availability of selected environmental test endpoints in IUCLID

Test endpoint	Total hits	Number of single substances with experimental data of potentially reliable quality
Ready Biodegradation (OECD 301B, 301F and Sturm)	293	55
Bioconcentration Factor (OECD 305C)	444	60
Acute Algae Toxicity		
OECD 201	293	ca. 50%
EC C	50	
Acute Daphnia Toxicity		
OECD 202	336	ca. 50%
EC C2	187	
Acute Fish Toxicity		
OECD 203	478	ca. 50%
EC C1	60	
Chronic Fish Toxicity (OECD 204 and other)	92	33

Conclusions

The evaluation of the IUCLID database version 4.0 updated in February 2000 (covering phase I and phase II existing chemicals) showed that currently:

- It potentially contains useful environmental data for (Q)SAR-related uses.

L.1.5 General conclusions

There are a number of concerns about the use of IUCLID that must be addressed to improve its applicability for (Q)SAR purposes:

1. It is well known that the data included is of a very variable quality. As HPV substances have gone through the ESR process the IUCLID records for those chemicals have been investigated, corrected and in many cases quality criteria assigned. This assessment of data quality needs to be consistently applied across the entire database, a daunting task, but one that again should be addressed as part of the HPV programme or in the early stages of the implementation of the proposed EU REACH process.
2. There is a practical need for a more user-friendly interface that would enable IUCLID to be directly queried to extract data in a given format amenable for (Q)SAR research. The searching capabilities provided to the routine user are currently quite limited and data can only be exported in an ASCII format making IUCLID awkward for use in (Q)SAR activities. Further software improvements in IUCLID for searching and exporting information in a convenient format is therefore needed.
3. It is strongly recommended that if IUCLID is to assume a key role in (Q)SAR development and validation, a clear process be identified by which new data that is entered may be flagged with a date stamp. In this way, any new data for a given test endpoint can be used for the purposes of validation (if the new chemical falls within the domain of a previously existing (Q)SAR) or model improvement (if outside the domain of existing (Q)SARs). In this way, each IUCLID update could serve as a new source of data that provides an opportunity for either confirming or expanding existing (Q)SAR models. Such a strategy will ensure that (Q)SAR knowledge can be maximised from the wealth of new, good quality data that will be generated as part of the HPV programme or the proposed EU REACH system.

L.2 Other databases for human health and environmental endpoints

L.2.1 ITIC database

This database project is driven by the ILSI Health and Environmental Science Institute (HESI) Structure Activity Relationship (SAR) Database Subcommittee. The mission of the SAR database subcommittee was to utilise the vast collection of toxicology data that has been developed by the international government, industry and academic community to establish a comprehensive database of toxicity test results, which will be useful for predictive toxicology. During a 2-year project (July 2000-2002), an international collaboration of ILSI HESI and LHASA, at the University of Leeds (UK), supported by global industry and regulators, developed a pilot database (ITIC) of toxicity testing data.

The pilot database was populated with toxicology data for four endpoints – carcinogenicity, mutagenicity, skin sensitisation and hepatic toxicity compiled from publicly available sources. The pilot web-based database was derived from IUCLID and was modified to meet data searching needs. It can be used for structure, substructure and biological data searches, easy data input and output in standard formats as well as report generation.

As a continuation, the development of a full-scale comprehensive database is planned. This project is open for global participation of industry, regulatory bodies and academic institutions. The comprehensive database will be populated with data from additional toxicological human health and environmental endpoints, ideally associated with quality identifier. In contrast to other database projects, participants are encouraged to share data not yet freely available. The database will be made publicly available, probably supported by a subscription fee, for development and improvement of predictive toxicological models and as a screening tool for early chemical hazard identification.

L.2.2 Distributed Structure-Searchable Toxicity (DSSTox) public database network

Most of the database efforts try to establish a global repository of toxicity information that fulfils the needs for access to public data linked with chemical structure searching. In contrast, the Distributed Structure-Searchable Toxicity (DSSTox) Public Database Network does not attempt to build a central repository database, which can be difficult to construct and to maintain, but is being proposed as a community-supported, web-based effort. The proposed DSSTox network has the following elements: (1) Adopt and encourage the use of a common standard data format (SDF) for public toxicity databases that will include chemical structures and that can be easily imported into available chemical relational database (CRD) applications. (2) Implement a distributed source approach that will enable a decentralised free public access to toxicity data files and that will link toxicity data sources with the potential user and (3) Engage public, commercial, academic and industry groups in contributing and expanding the DSSTox database network.

DSSTox SDF files will be created for a wide variety of available public toxicological databases, which will then be easily convertible to data tables or importable to any commercial or private CRD application. The SDF file will provide summary toxicity information as opposed to detailed descriptive text. SDF files will be downloaded to the users' PC and a user-customised toxicity CRD can be created, that could be fully searched, assessed, reformulated or merged with proprietary or other data. A central DSSTox web site will serve as the hub of the DSSTox project, providing general information on SDF files, data sources etc. DSSTox SDF files are in various stages of development for a selection of public toxicity databases spanning human health and environmental endpoints. DSSTox SDF files for the Carcinogenic Potency Databases are nearly complete. Other SDF files currently under development include the US EPA/IARC-GAP database, the US EPA-genotox database, NTP Salmonella database, the US EPA-ECOTOX database etc. The DSSTox proposal is a simple way to make existing data more useful and universally accessible to toxicologists and modellers. The DSSTox files will be made freely and publicly available. The proposal however, does not tackle issues related to toxicity data representation, reproducibility, relevance and quality.

L.2.3 Repeated dose toxicity database – Fraunhofer Institute for Toxicology

This database project is being carried out by the Fraunhofer Institute for Toxicology (Hannover, Germany) as a Cefic LRI project to identify SARs for substances with low NOELs. The objective of this project is to generate a central database with data on repeated dose toxicity. The aim is to use the database to derive relationships between certain functional groups and toxic effects and to help to classify chemicals into combinations of chemical functional groups or categories and toxic endpoints, i.e. target organs. Review documents, reviewed by expert groups, (e.g. MAK, EHC, BUA documentation, EU risk assessments) were used to obtain high quality data for the database. At present, the database consists of 100 chemicals and 243 studies with a total of about 650 specific effects. Chemicals entered into the database are limited to simple structures (no inorganics, metal compounds, mixtures) with LOELs between 0.05 and 6000 mg/kg bw/day. Future work will include continuous data recording for a wider group of chemicals as well as evaluation of the data. It has not yet been decided how this database should be distributed.

L.2.4 Commercial human health toxicity databases

Three commercial programs, Toxicity Finder provided by MDL Information Systems, Inc., LeadScope's ToxScope database™ and the SciVision ToxSys database offer chemical database searching through published human health toxicity data. The LeadScope technology also allows for substructure searching and combines search outputs with strong data visualisation tools. The new Leadscope virtual decision-making software provides access to 150,000 chemical structures and expert statistical analysis of complex data. The Leadscope ToxScope databases™ encompass acute toxicity, hepatotoxicity, mutagenicity and carcinogenicity data from RTCES, NTP and some other sources and they allow for data comparisons, exploration of chemical space or management of in-house data.

The MDL Toxicity Finder is an oracle based structure searchable database, which covers over 150,000 toxic chemicals from *in vivo* and *in vitro* studies of acute toxicity, mutagenicity, skin and eye irritation, tumorigenicity and carcinogenicity, reproductive effects and multiple dose effects.

ToxSys comprises about 230,000 compounds divided into various categories similar to MDL and Leadscope. In addition it also covers human skin patch test data, repeat dose toxicity, food additive toxic effects, nephrotoxicity and endocrine disruption. ToxSys is primarily a structure searchable database but it also has some limited QSAR modules.

Additionally, two prediction applications, TOPKAT and MultiCASE provide a user with limited access to toxicity databases for a variety of endpoints. The toxicity databases have been compiled from primary sources for the purpose of model development and are only available with purchase of the commercial program, not just the database. In their current forms, TOPKAT and MultiCASE place limits on a user's open access to their internal toxicity data and they have limited searching capabilities.

L.2.5 C-QSAR

This system stems from the mind of the 'father of QSAR' Corwin Hansch. The QSAR database has been assembled at MedChem (Pomona College) over a period of more than 20 years, with input from academic leaders in specific fields. It consists of a dual database of QSAR equations relating biological and physico-chemical activities to structural parameters. BIO currently contains 5600+ equations and PHYS over 7500. Training sets for all equations are accessible and full references are provided.

The purpose of this dynamic, integrated electronic system is more than a compilation of QSAR from all areas of chemistry and (chemical) biology. The most important reason is that of establishing the validity of a QSAR. The developers believe that only via *lateral correlation* of a given QSAR with many others can one develop confidence in the validity of any approach to the correlation of structure with biological activity. Users find the database valuable for validation of new equations as they are being developed; i.e. to see if the emerging structure-activity relationship bears a resemblance to others with known mechanisms.

Open version of C-QSAR exists on VAX/VMS platform. Limited functionality is provided via queries on the Internet (http://www.biobyte.com/ or http://clogp.pomona.edu/chem/qsar-db/sets/ghindex.html). The VAX/VMS version allows multiple access and contains the complete set of dependent and independent parameter values and structural information for each compound (SMILES). The VAX/VMS version is about 30 000 USD and among others includes ClogP and MASTERFILE, a database with over 36,000 structures, 43,000 names, 53,000 logP values in various solvent systems (including over 11,000 high-confidence octanol/water values, called LOGPSTAR), 13,000 pKa values, 18,000 CAS numbers and much more. BioByte Corporation currently supports it.

L.2.6 Other databases for environmental toxicity endpoints

Databases covering ecotoxicological data were reviewed by ECETOC (ECETOC, 1998a). Table L.7 summarises the available databases incorporating environmental information.

Table L.7: Databases incorporating environmental information

Database	Data	Comments
ENVIROLINE	Very wide range including air, water and land pollution and environmental impact of chemicals	
CA Search	Very wide database	Usually referred to as Chemical Abstracts
BIOSIS Previews	Life Sciences database	
EMBASE	Medical and pharmaceutical database	
ECDIN	Data on 65,000 substances	
Environmental Fate	Transport and fate of chemicals	Links 4 sub files : Datalog, Databases Biolog, Chemfate and Biodeg
Hazardous Substances Database	Scientifically reviewed and edited data,includes biodegradation, ecotoxicity and log K_{ow}	
LOG P Database	Log K_{ow}	Contains over 30,000 data on 14,000 compounds.
AQUIRE	US EPA database for aquatic tox data, periodic updates	AQUatic Information REtrieval system - with simple quality rating system
ISHOW	US EPA database of phys chem data developed by Duluth Lab for QSAR studies simple quality rating system	Information System for Hazardous Organics in Water -
ENVIROFATE	US EPA Office of Toxic Substances + SRC database of summary environmental fate and effects data	With simple quality rating system
BIODEG	US EPA Office of Toxic Substances + SRC database of summary environmental fate and effects data	With simple quality rating system
BIOLOG	Bibliographic database - containing references to papers covering biodegradation and bacterial inhibition studies.	No quality rating
DATALOG	Containing references to papers covering phys chem/occ health/ monitoring studies	Bibliographic database only - no quality rating
CHRIS	US Coastguard database containing full datasheets for materials transported by sea for emergency response	Chemical Hazard Response Information System - no quality rating
OHMTADS	US EPA database containing safety datasheets physico-chemical/ecotox/environmental fate	Oil and Hazardous Materials/Technical Assistance Data System - no quality rating
TSCATS	Bibliographic references to data submitted to US EPA under section 8d of TSCA	Toxic Substances Control Act Test Submissions database - Bibliographic - no quality rating.
RTECS	Data and bibliographic references to sources	Registry of Toxic Effects of Chemical Substances - National Institute for Occupational Safety and Health NIOSH
SIGEDB	Numerical database of phys chem/ environmental fate data	In German
HEILBRON : Now Chapman and Hall Chemical Databases (CHCD)	Numerical/bibliographic data from Chapman and Hall dictionaries	No quality ratings
Arizona DB	Aqueous solubility	10,000 data points - includes recommended values
CHEMINFO	Acute and chronic data	Limited to 600 chemicals
IUCLID	Any available data	On all HPVCs submitted by Industry
ECETOC – EAT3	Aquatic toxicity (acute and chronic)	High quality but limited

GLOSSARY

Coverage
Ratio of reliable predictions positives and negatives to the total number of experimentally tested compounds.

Predictive values
Positive: Ratio of correctly predicted positives to total number of predicted positives.
Negative: Ratio of correctly predicted negatives to total number of predicted negatives.

Predictivity
Ratio of correctly predicted positives and negatives to the total number of predicted compounds.

QSAR (quantitative structure-activity relationship)
A mathematical model that relates a quantitative measure of chemical structure (e.g. a physico-chemical property) to a physical property or to a biological effect (e.g. a toxicological endpoint).

SAR (structure-activity relationship)
A (qualitative) association between a chemical substructure and the potential of a chemical containing the substructure to exhibit a certain biological effect.

Sensitivity
Ratio of correctly predicted positives to total number of experimental positives

Specificity
Ratio of correctly predicted negatives to total number of experimental negatives

ABBREVIATIONS

ADME	Adsorption, Distribution, Metabolism, Excretion
ATSDR	Agency for Toxic Substances and Disease Registry
CSTEE	Scientific Committee on Toxicology, Ecotoxicology and the Environment
DSS	Decision Support System
DSStox	Distributed Structure Searchable Toxicity
ECB	European Chemicals Bureau
ECVAM	European Centre for Validation of Alternative Methods
FDA	Food and Drug Administration
HPV	High production volume
ICCA	International Council of Chemical Associations.
ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
ITC	Interaging Testing Committee
ITIC	International Toxicity Information Centre
IUCLID	International Uniform Chemical Information Database
OECD	Organisation for Economic Cooperation and Development
OPS	Optimum Prediction Space
OSPAR	Oslo Paris Commission
PC	Physico-chemical
PMN	Pre-manufacturing Notification
QDSS	QSAR Decision Support System
(Q)SAR	(Quantitative) Structure Activity Relationship
SAR	Structure Activity Relationship
SIDS	Screening Information Data Set
SMILES	Simplified molecular input line entry system
TGD	Technical Guidance Document
TSCA	Toxic Substances Control Act
US EPA	United States Environmental Protection Agency

BIBLIOGRAPHY

Abraham MH, Chadha HS, Mitchell RC. 1995. The factors that influence skin penetration of solutes. *J Pharm Pharmacol* 47:6-16.

Abraham MH, Le J. 1999. The correlation and prediction of the solubility of compounds in water using an amended solvation energy relationship. *J Pharm Sci* 88:868-880.

Aldenberg T, Jaworska JS. 2000. Uncertainty of the hazardous concentration and fraction affected for normal species sensitivity distributions. *Ecotoxicol Environ Saf* 46:1-18.

Aldenberg T, Slob W. 1993. Confidence limits for hazardous concentrations based on logistically distributed NOEC toxicity data. *Ecotoxicol Environ Saf* 25:48-63.

Alexander M. 1981. Biodegradation of chemicals of environmental concern. *Science* 211:132-138.

Allanou R, Hansen BG, van der Bilt Y. 1999. Public availability of data on EU high production volume chemicals. Report EUR 18996. European Commission, Joint Research Centre, Institute for Health and Consumer Protection, European Chemicals Bureau, Ispra, Italy.

Ashby J, Tennant RW (1994) Prediction of rodent carcinogenicity for 44 chemicals: results. Mutagenesis 9:15.

Atkinson R. 1985. Kinetics and mechanisms of gas-phase reactions with the hydroxyl radical with organic compounds under atmospheric condition. *Chem Rev* 85:61-201.

Atkinson R. 1987. A structure–activity relationship for the estimation of rate constants for the gas-phase reactions of OH radicals with organic compounds. *Int J Chem Kinet* 19:799–828.

Atkinson R. 1988. Estimation of gas phase hydroxyl radical rate constants for organic chemicals. *Environ Toxicol Chem* 7:435-442.

Atkinson R. 1989. Kinetics and mechanisms of the gas-phase reactions of the hydroxyl radical with organic compounds. *J Phys Chem Ref Data Monograph No 1* American Institute of Physics and American Chemical Society, New York, NY, USA.

Atkinson R. 2000. Atmospheric oxidation. In Boethling RS, Mackay D, eds, *Handbook of Property Estimation Methods for Chemicals*. Lewis Publishers 14:335-354.

Atkinson R, Carter WPL. 1984. Kinetics and mechanisms of the gas-phase reactions of ozone with organic compounds under atmospheric conditions. *Chem. Rev*. 84: 437-470.

Balk F, Ford RA. 1999. Environmental risk assessment for polycyclic musks AHTN and HHCB in the EU I. fate and exposure assessment. *Toxicol Lett* 111:57-79.

Barron MG, Hansen JA, Lipton J. 2002. Association between contaminant tissue residues and effects in aquatic organisms. *Rev Environ Contam Toxicol* 173:1-37.

Basketter DA. 1998. Chemistry of contact allergens and irritants. *Am J Contact Derm* 9:119-124.

Benigni R. 1997. The first US National Toxicology Programme exercise on the prediction of rodent carcinogenicity: definitive results. *Mutat Res* 387:35-45.

Benigni R. 2000. Structure-activity relationships in mutagenesis and carcinogenesis. In Balls M, van Zeller AM, Halder ME, eds, *Progress in the Reduction, Refinement and Replacement of Animal Experimentation*. Elsevier, Amsterdam, The Netherlands, pp 469-478.

Benigni R, Passerini L. 2002. Carcinogenicity of the aromatic amines: from structure-activity relationships to mechanisms of action and risk assessment. *Mutat Res* 511:191-206.

Bennett DH, Margni DM, Mckone TE, Jolliet O. 2003. Intake fraction for multimedia pollutants, A tool for lifecycle analysis and comparative risk assessment. *Risk Analysis*, *In Press.*

Bintein S, Devilliers J, Karchers W. 1993. Non-linear dependence of fish bioconcentration on n-octanol water partition coefficient. *SAR QSAR Environ Res* 1:29-39.

Boethling RS, Sabljic A. 1989. Screening-level model for aerobic biodegradability based on a survey of expert knowledge. *Environ Sci Technol* 23:672-679.

Boethling RS, Howard PH, Meylan W, Stiteler W, Beaumann J, Tirado N. 1994. Group contribution method for predicting probability and rate of aerobic biodegradation. *Environ Sci Technol* 28:459-465.

Boethling RS, Mackay D. 2000. eds, *Handbook of Property Estimation Methods for Chemicals*. Environmental and Health Sciences. Lewis Publishers, New York, NY, USA.

Brennan RA, Nirmalakhandan N, Speece RE. 1998. Comparison of predictive methods for Henry's law coefficients of organic chemicals. *Water Res* 32:1901-1911. *[JH please change date in Section C.8. Thanks]*

Bristol DW, Wachsman JT, Greenwell A. 1996. The NIEHS predictive-toxicology evaluation project: chemcarcinogenicity bioassays. *Environ Health Perspect* 104:1001-1010.

Buchwald P, Bodor N. 2001. A simple, predictive, structure-based skin permeability model. *J Pharm Pharmacol* 53:1087-1098.

Buehler EV. 1965. Delayed contact hypersensitivity in the guinea pig. *Arch Dermatol* 91:171-177.

Burkhard LA. 2000. Estimating dissolved organic carbon partition coefficients for nonionic organic chemicals. *Environ Sci Technol* 34:4663-4668.

Bysshe SE. 1990. Bioconcentration factor in aquatic organisms. In: Lyman WJ, Reehl WF and Rosenblatt DH, eds, *Handbook of Chemical Property Estimation Methods – Environmental Behaviour of Organic Compounds*. American Chemical Society, Washington DC, USA.

Cariello NF, Wilson JD, Britt BH, Wedd DJ, Burlinson B and Gombar V. 2002. Comparison of the computer programs DEREK and TOPKAT to predict bacterial mutagenicity. *Mutagenesis* 17:321-329.

Carr KH, Coyle GT, Kimerle RA. 1997. Bioconcentration of [^{14}C]butyl benzyl phthalate in bluegill sunfish (*Lepomis macrochirus*). *Environ Toxicol Chem* 16:2200–2203.

Cefic. 2002. Proceedings of workshop on regulatory acceptance of (Q)SARs for human health and environmental endpoints, Setubal, March 4-6, 2002.

Chen XQ, Cho SJ, Li Y, Venkatesh S. 2002. Prediction of aqueous solubility of organic compounds using a quantitative structure - property relationship. *J Pharm Sci* 91:8.

Connell DW. 1987. Bioaccumulation behaviour of persistent organic chemicals with aquatic organisms. *Reviews of Environ Contam Toxicol* 101:311-320.

Connell DW, Hawker DW. 1988. Use of polynomial expressions to describe the bioconcentration of hydrophobic chemicals by fish. *Ecotoxicol Environ Saf* 16:242-257.

Cronin M, Dearden JC, Moss GP, Murray-Dickson G. 1999. Investigation of the mechanism of flux across human skin *in vitro* by quantitative structure-permeability relationships. *Eur J Pharm Sci* 7:325-330.

Cronin M, Jaworska J, Walker J, Comber M, Watts C. 2003a. Use of QSARs in international decision-making frameworks to predict ecological effects and environmental fate of chemical substances. *Environ Health Perspect*. 111(10);1376-1390

Cronin M, Jaworska J, Walker J, Comber M, Watts C. 2003b. Use of QSARs in international decision-making frameworks to predict human health effects of chemical substances. *Environ Health Perspect*. 111(10): 1391-1418

de Maagd PGJ. 1997. Polycyclic aromatic hydrocarbons: fate and effects in the aquatic environment. Ph.D. Dissertation, University of Utrecht, The Netherlands.

de Wolf W, Seinen W, Hermens JLM. 1993. Biotransformation and toxicokinetics of trichloroanilines in fish in relation to their hydrophobicity. *Arch Environ Contam Toxicol* 25:110-117.

de Zwart D. 2001. Observed regularities in species sensitivity distributions for aquatic species. In Posthuma L, Suter GW, Traas TP, eds, *Species sensitivity distributions in ecotoxicology*. Lewis publisher CRC Press, Boca Raton, Fl, USA.

Dearden JC, Barratt MD, Benigni R, Bristol DW, Combes RD, Cronin MTD, Judson PN, Payne MP, Richard AM, Tichy M, Worth AP, Yourick JJ. 1997. The development and validation of expert systems for predicting toxicity – The report and recommendations of an ECVAM/ECB workshop (ECVAM workshop 24) ATLA 25:223-252.

Dearden JC, Netzeva TI, Bibby R. 2002. Comparison of a number of commercial software programs for the prediction of aqueous solubility. *Journal of Pharmacy and Pharmacology* 54, Supplement, page S-66 (one page only).

Delbanco EH. 2002. Use of the prediction software DEREK in the hazard assessment of raw materials. *Naunyn Schmiedeberg's Arch Pharmacol* Suppl 365:R 639

Devillers J, Domine D, Karcher W. 1995a. Estimating n-octanol/water partition coefficients from autocorrelation method. *SAR and QSAR Environ Res* 3:301-306.

Devillers J, Bintein S, Domine D. 1995b. Comparison of BCF models based on log P. Conducted for: Ministere de l'Environment, Bureau des Substances et Preparations Chimiques, CTIS Report No 95038, Lyon France.

Di Toro DM, McGrath JA, Hansen DJ. 2000. Technical basis for narcotic chemicals and PAH criteria: I. Water and tissue. *Environ Toxicol Chem* 19:1951-1970.

Durham SK, Pearl GM. 2001. Computational methods to predict drug safety liabilities. *Curr Opin Drug Discov Devel* 4:110-114.

EC. 1993. Council Regulation (EC) 793/93 of 23 March 1993 on the evaluation and control of the risks of existing substances. *Off J of the EC* L 84/1.

EC. 1996. Technical Guidance Document on risk assessment for new and existing substances. *Off J of the EC*, Luxembourg.

EC. 2000a. IUCLID CD ROM - Year 2000 Edition, Public Data on High Volume Chemicals, European Commission, Joint Research Centre, Institute for Health and Consumer Protection, European Chemicals Bureau, Ispra, Italy.

EC. 2000b. Council Directive 2000/33/EC of 25 April 2000 adopting to technical progress for the 27th time Council directive 67/548/EEC on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances. *Off J of the EC*, L136/90.

EC. 2003. 2nd Edition of the TGD in support of Commission Directive 93/67/EEC on risk assessment for new notified substances and commission regulation (EC) No 1488/94 on risk assessment for existing substances (1996) ISBN 92-827-8012-0. *Off J of the EC.*

ECETOC. 1993. Aquatic toxicity data evaluation task force. Technical Report No. 56. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels, Belgium.

ECETOC. 1995. The role of bioaccumulation in environmental risk assessment: the aquatic environment and related food webs. Technical Report No.67. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels, Belgium.

ECETOC. 1998a. QSARs in the assessment of the environmental fate and effects of chemicals. Technical Report No. 74. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels, Belgium.

ECETOC. 1998b. Eye irritation reference chemical data bank (second edition). Technical Report No 48(2). European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels, Belgium.

ECETOC. 2001. Genomics, transcript profiling, proteomics and metabonomics (GTPM): an introduction. Document No 42. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels, Belgium.

ECETOC. 2003a. Persistence of chemicals in the environment. Technical Report No. 90. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels, Belgium.

ECETOC. 2003b. Aquatic hazard assessment II. Technical Report No. 91. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels, Belgium.

El Tayar N, Tsai RS, Tasta B, Carrupt PA, Hansch C, Leo A. 1991. Percutaneous penetration of drugs: a quantitative structure-permeability relationship study. *J Pharm Sci* 80:744-749.

Enslein K. 1984. Estimation of toxicological endpoints by structure-activity relationships. *Pharmacol Rev* 36:131-135.

Enslein K, Blake BW, Borgstedt HH. 1990. Prediction of probability of carcinogenicity for a set of ongoing NTP bioassays. *Mutagenesis* 5:305-306.

Enslein K, Gombar VK, Balke BW. 1994. Use of SAR in computer-assisted prediciton of carcinogenicty and mutagencity of chemcials by the TOPKAT program. *Mutat Res* 305:47-61.

Eriksson L, Jaworska J, Cronin M, Worth A, Gramatica P, McDowell R. 2003. Existing methods for reliability and uncertainty assessment and applicability evaluations of regression based and classification QSARs. *Environ Health Perspect 111:1361-1375*.

Gawlik BM, Sotiriou N, Feicht EA, Schulte-Hostede S, Kettrup A. 1997. Alternatives for the determination of the soil adsorption coefficient, K-oc, of non-ionicorganic compounds - A review. *Chemosphere* 34:2525-2551.

Gerberick GF, Ryan CA, Kimber I, Dearman RJ, Lea LJ, Basketter DA. 2000. Local lymph node assay: validation assessment for regulary purposes. *Am J Contact Derm* 11:3-18.

Gerner I, Zinke S, Graetschel G, Schlede E. 2000. Development of a decision support system for the introduction of alternative methods into local irritancy/corrosivity testing strategies. Creation of fundamental rules for a decision support system. *ATLA* 28: 665-698.

Gobas FAPC, Morrison HA. 2000. Bioconcentration and biomagnification in the aquatic environment. In Boethling RS, Mackay D, eds, *Handbook of property estimation methods for chemicals* - Vol 9. Lewis Publishers, Boca Raton, Fl, USA, pp189-231.

Gold P, Ogle G. 1969. Estimating thermophysical properties of liquids. Part 4 – Boiling freezing and triple point temperatures. *Chem Eng* 113:119-122.

Gottmann E, Kramer S, Pfahringer B, Helma C. 2001. Data quality in predictive toxicology: reproducibility of rodent carcinogenicity experiments. *Environmental Health Perspectives* 109:509-514.

Greene N. 2002. Computer systems for the prediction of toxicity: an update. *Advanced Drug Delivery Reviews* 54:417-431.

Grist EPM, Leung KMY, Wheelen JR, Crane M. 2002. Better bootstrap estimation of hazard concentration thresholds for aquatic assemblages. *Environ Toxicol Chem* 21:1515-1524.

Harris JC. 1990. Rate of hydrolysis. In Lyman WJ, Reehl WF, Rosenblatt, DH, eds, *Handbook of Chemical Property Estimation Methods: Environmental Behaviour of Organic Compounds.* American Chemical Society, Washington DC, USA.

Hilal SH, Karickhoff SW. 1995. A rigorous test for SPARCs chemical reactivity models: Estimation of more than 4300 ionisation PK_as. *Quant Struct Act Relat* 14:348-355.

Hilal SH, Carreira A, Karickhoff SW. 1994. Estimation of chemical reactivity and parameters and physical properties of organic molecules using SPARC. In Politzer P, Murray JS, eds, *Quantitative treatments of solute/solvent interactions* - Vol 1. Elsevier, Amsterdam, pp. 291-348.

Hilal SH, Karickhoff SW, Carreira LA. 1999. Estimation of microscopic, zwitterionic ionization constants, isoelectric point and molecular speciation of organic compounds. *Talanta* 50, 827-840.

Howard PH, Meylan WM. 1997. Prediction of physical properties, transport and degradation for environmental fate and exposure assessments In: Quantitative Structure-Activity Relationships in Environmental Sciences-VII, eds, Chen F, Schuurmann G, SETAC Press 185-205, Pensacola Fl, USA.

Howard PH, Hueber AE, Boethling RS. 1987. Biodegradation data evaluation for structure/biodegradability relations. *Environ Toxicol Chem* 6:1-10.

Howard PH, Boethling RS, Stiteler WM, Meylan WM, Hueber AE, Beauman JA, Larosche ME. 1992. Predictive model for aerobic biodegradability developed from a file of evaluated biodegradation data. *Environ Toxicol Chem* 11:593-603.

Hüls. 1997. Diethylhexylphthalate: study on the determination of the bioaccumulation in carp, Final Report BA 006, Marl, Germany.

Hulzebos EM, Schielen PCJI, Wijkhuizen-Maslankiewicz L. 1999. (Q)SARs for human toxicological endpoints: a literature search. *RIVM Report* 601516.001, The Netherlands.

Isnard P, Lambert S. 1988. Estimating bioconcentration factors from octanol water partition coefficient and aqueous solubility. *Chemosphere* 17:21-34.

Jaworska JS, Dimitrov S, Nikolova N, Mekenyan O. 2002. Probabilistic assessment of biodegradability based on metabolic pathways: CATABOL system. *SAR QSAR Environ Res* 13:307-323.

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

Joback KG. 1982. A unified approach to physical property estimation using multivariate statistical techniques. Steven Institute of Technology, submitted to the Department of Chemical Engineering for M.S. degree at the Massachusetts Inst. of Technol, June 1984.

Jouravleva D. 2002. ACD/PhysChem product Overview, http://www.acdlabs.co.uk/download/publ/2002/aaps2002/physchem.pdf.

Kaiser KLE, Dearden J, Klein W, Schultz TW. 1999. Note of caution to the users of ECOSAR. *Wat Qual Res J Can* 34:179-182.

Karickhoff SW, McDaniel VK, Melton C, Vellino AN, Nute DE, Carreira LA. 1991. Predicting chemical reactivity by computer. *Environ Toxicol Chem* 10:1405-1416.

Kelcka GM. 1985. Biodegradation. In: *Environmental exposure from chemicals*, eds, Neely WB, Blau GE, CRC Boca Raton, Fl, USA, 109-155.

Kimber I, Basketter DA. 1992. The murine local lymph node assay: commentary on

collaborative studies and new directions. *Food Chem Toxicol* 30:165-169.

Kimber I, Basketter DA, Berthold K, Butler M, Garrigue J-L, Lea L, Newsome C, Roggeband R, Steiling W, Stropp G, Watermann S, Wiemann C. 2001. Skin sensitization testing in potency and risk assessment. *Toxicol Sci* 59:198-208.

Kimber I, Dearman RJ, Scholes EW, Basketter DA. 1994. The local lymph node assay: developments and applications. *Toxicology* 93:13-31.

King EF, Painter HA. 1983. Ring-test programme 1981-82. Assessment of biodegradability of chemicals in water by manometric respirometry. Commission to the EC, DG XI, Report No EUR 8631 EN.

Kitano M, Takatsuki M. 1988. Evaluation of the 1988 OECD ring-test of ready biodegradability, 12th September 1988, Chem Inspect and Test Inst, Japan.

Klimisch HJ, Andreae M, Tillmann U. 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. *Regulatory Toxicol Pharmacol* 25:1-5.

Klopman G. 1984. Computer automated structure evaluation of organic molecules. *J Am Chem Soc* 106:7315-7324.

Klopman G. 1992. MULTI-CASE 1. A hierarchical computer automated structure evaluation program. *Quant Struct Act Rel* 11:176-184.

Klopman G, Tu M. 1997. Structure-biodegradability study and computer-automated prediction of aerobic biodegradation of chemicals. *Environ Toxicol Chem* 16:1829-1835.

Klopman G, Tu M, Talafous J. 1997. META 3: A genetic algorithm for metabolic transform priorities optimization. *J Chem Inf Comput Sci* 37:329-334.

Klopman G, Saiakhov R, Rosenkranz HS, Hermens JLM. 1999. Multiple Computer Automated Structure Evaluation program study of aquatic toxicity. 1. Guppy. *Environ Toxicol Chem* 18:2497-2505.

Kristensen P, Tyle H. 1990. The assessment of bioaccumulation. In Nagel R, Loskill R (eds), Bioaccumulation in aquatic systems:Contribution to the assessment, International workshop, Berlin, Germany. VCH, Berlin 189-222.

Kwok ESC, Atkinson R. 1995. Estimation of hydroxyl radical reaction rate constants for gas-phase organic compounds using a structure–reactivity relationship: an update. *Atmos Environ* 29:1685–1695.

Leo A. 1993. Calculating logPoct from structures. *Chem Rev* 93:1281-1306.

Leo A. 2000. Octanol/water partition coefficient. In Boethling RS, Mackay D, eds, *Handbook of property estimation methods for chemicals* – Vol 5. Lewis Publishers, Boca Raton, Fl, USA, pp89-114.

Letinksi DJ, Connelly MJ, Peterson DR, Parkerton TF. 2002. Slow-stir water solubility measurements of selected alcohols and diesters. *Chemosphere* 48:257-265.

Lewis RW, Botham PA. 1994. Measurement of transcutaneous electrical resistance to assess the skin corrosivity potential of chemicals. In Rougier A, Goldberg AM, Maibach HI, eds, In vitro *skin toxicology*. Mary Ann Liebert, New York, NY, USA, pp 161-169.

Lyman WJ. 2000. Boiling Point. In Boethling RS, Mackay D, eds, *Handbook of Property Estimation Methods for Chemicals* – Vol 2. Lewis Publishers, Boca Raton, Fl, USA pp 29-51.

Lyman WJ, Reehl WF, Rosenblatt, DH. 1982. *Handbook of Chemical Property Estimation Methods: Environmental Behaviour of Organic Compounds*. McGraw Hill, New York, USA.

Lyman WJ, Reehl WF, Rosenblatt, DH. 1990. *Handbook of Chemical Property Estimation Methods: Environmental Behaviour of Organic Compounds*. American Chemical Society, Washington DC, USA.

Mackay D. 1982. Correlation of bioconcentration factors. *Environ Sci Technol* 16:274-278.

Mackay D. 2000. Solubility in water. In Boethling RS, Mackay D, eds, *Handbook of Property Estimation Methods for Chemicals* – Vol 7. Lewis Publishers, Boca Raton, Fl, USA pp 125-139.

Mackay D, Shiu WY, Ma KC. 2000. Henry's Law constant. In Boethling RS, Mackay D, eds, *Handbook of Property Estimation Methods for Chemicals* – Vol 4. Lewis Publishers, Boca Raton, Fl, USA pp 69-87.

Magnusson B, Kligman AM. 1969. The identification of contact allergens by animal assay. The guinea pig maximization test. *J Invest Dermatol* 52:268.

Mancini ER, Steen A, Rausina GA, Wong DCL, Arnold WR, Gostomski FE, Davies T, Hockett JR, Stubblefield WA, Drottan KR, Springer TA, Enrico P. 2002. MTBE ambient water quality criteria development: a public private partnership. *Environ. Sci Technol* 36:125-129.

Mannhold R, Dross K. 1996. Calculation procedures for molecular lipophilicity: a comparative study. *Quant Struct-Activ Relat* 15:403-409.

Marchant CA. 1996. Prediction of rodent carcinogenicity using the DEREK system for 30 chemicals currently being tested by the National Toxicology Program. *Environ Health Perspect* 104:1065-1074.

Matthews EF, Contrera JF. 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.

Mayer FL. 1976. Residue dynamics of di-2-ethylhexyl phthalate in fathead minnow (*Pimephales promelas*). *J Fish Res Bd Can* 33:2610-2613.

McCarty LS, Mackay D. 1993. Enhancing ecotoxicological modelling and assessment. *Environ Sci Technol* 27:1719-1728.

McLachlan MS. 2000. Vegetation/air partition coefficient. In Boethling RS, Mackay D, eds, *Handbook of Property Estimation Methods for Chemicals* - Vol 6. Lewis Publishers, Boca Raton, Fl, USA, pp 115-123.

Meissner HP. 1949. Critical constants from parachor and molar refraction. *Chem Eng Prog* 45:149-153.

Mekenyan O, Karabunarliev St, Ivanov J, Dimitrov D. 1996. The OASIS computer system and its new tools for QSAR studies. In Dubois J-E, Gershon N, eds, *Industrial information and design issues*, Springer, Berlin, pp 113-124.

Meylan WM, Howard PH. 1991. Bond contribution method for estimating Henry's Law constants. *Environ Toxicol. Chem* 10:1283-1293.

Meylan WM, Howard PH. 1993a. User's guide for HYDRO. Syracuse Research Corporation. Environ. Sci. Center, Syracuse, NY, USA.

Meylan WM, Howard PH. 1993b. Computer estimations of the atmospheric gas-phase reaction rate of organic compounds with hydroxyl radicals and ozone, *Chemosphere* 26, 2293 - 2299.

Meylan WM, Howard PH, 1994. User's Guide for AOPWIN. Atmospheric Oxidation Program for Microsoft Windows 3.1® Syracuse Research Corporation. Environ Sci Center, Syracuse, NY, USA.

Meylan WM, Howard PH. 1995. Atom/fragment contribution method for estimating octanol-water partition coefficients. *J Pharm Sci* 84:83-92.

Meylan WM, Howard PH, Boethling RS. 1992. Molecular topology / fragment contribution method for predicting soil sorption coefficients. *Environ Sci Technol* 26:1560-1567.

Meylan WM, Howard PH, Boethling RS. 1996. Improved method for estimating water solubility from octanol/water partition coefficient. *Environ Toxicol Chem* 15:100-106.

Meylan WM, Howard PH, Boethling RS, Aronson D, Printup H, Gouchie S. 1999. Improved method for estimating bioconcentration / bioaccumulation factor from octanol/water partition coefficient. *Environ Toxicol Chem* 18:664-672.

Mill T. 2000. Photoreactions in surface waters. In Boethling RS, Mackay D, eds, *Handbook of Property Estimation Methods for Chemicals* – Vol 15. Lewis Publishers, Boca Raton, Fl, USA, pp355-381.

MITI. 1992. Bioaccumulation and biodegradation data on existing chemicals based on the CSCL Japan, Ministry of International Trade and Industry Japan, Chemicals Inspection and Testing Institute of Japan.

Moore DRJ. 2000. *A comparison of model performance for six QSAR packages that predict acute toxicity to fish*. Report prepared for Environment Canada, Chemicals Evaluation Division by The Cadmus Group Inc. Ottawa, CA.

Mueller F, Simon-Hettich B, Kramer P-J. 2000. An evaluation of the predictability of the Ames-test mutagenicity using the TOPKAT™ Program. *Naunyn-Schmiedeberg's Archives of Pharmacology*, Supplement to volume 361:R173 (Abstract).

Mumtaz MM, Knauf LA, Reisman DJ, Peirano WB, DeRosa CT, Gombar VK, Enslein K, Carter JR, Blake BW, Huque KI, Ramanujam VMS. 1995. Assessment of effect levels of chemicals from quantitative structure-activity relationship (QSAR) models. I. Chronic lowest-observed-adverse-effect level (LOAEL). *Toxicol Lett* 79:131-143.

Nendza M. 1998. Structure-activity relationships in environmental sciences. Chapman and Hall, London, UK.

OECD. 1993. Guideline for testing of chemicals - ready biodegradability. Organisation for Economic Cooperation and Development, Paris, France.

OECD. 1994. US EPA/EC Joint project on the evaluation of (Quantitative) Structure Activity Relationship, OECD Report No OECD/GD/(94) 28, Paris, France.

Parham M, Hall I, Kier L. 2000. Accurate prediction of logP using e-state indices with neural net network Am Chem Soc annual meeting. Presentation stored at http://www.logp.com

Parkerton TF. 2001. Bioaccumulation. In Jaworska J, Parkerton T, Thompson E, eds, Summary of the Cefic's ad-hoc QSAR group's Preliminary Evaluation of the Danish

EPA's self-advisory QSAR classification system, 20 Nov 2001.

Parkerton TF, Konkel WJ. 2000. Application of quantitative structure activity relationships for assessing the aquatic toxicity of phthalate esters. *Ecotoxicol Environ Saf* 45:61-78.

Patel H, ten Berge W, Cronin MTD. 2002. Quantitative structure-activity relationships (QSARs) for the prediction of skin permeation of exogenous chemicals. *Chemosphere* 48:603-613.

Pearl GM, Livingston-Carr S, Durham SK. 2001. Integration of computational analysis as a sentinel tool in toxicological assessments. *Current Topics in Medicinal Chemistry* 1:247-255.

Peijnenburg WJGM, de Beer KGM, de Haan MWA, den Hollander HA, Stegeman MHL, Verboom H. 1992. Development of a structure-reactivity relationship for the photohydrolysis of substituted aromatic halides. *Environ Sci Technol* 11:2116-2121.

Petrauskas AA, Kolovanov E. 2000. ACD/Log P Method Description. *Perspectives in Drug Discovery and Design* 19:99-116.

Posthuma L, Suter GW II, Traas TP. 2001. Species Sensitivity Distributions in Ecotoxicology. In Boethling RS, Mackay D, eds, *Handbook of Property Estimation Methods for Chemicals* - Vol 8. Lewis, Boca Raton, Fl, USA, pp 133-154.

Potts RO, Guy RH. 1992. Predicting skin permeability. *Pharm Res* 9:663-669.

Potts RO, Guy RH. 1995. A predictive algorithm for skin permeability: the effects of molecular size and hydrogen bond activity. *Pharm Res* 12:1628-1633.

PREDICT. 2002. Chemical thermodynamics and transport properties prediction software complete manual Dragon Technology. free dowload http://www.mwsoftware.com /dragon/manual.html.

Prival MJ. 2001. Evaluation of the TOPKAT system for predicting the carcinogenicity of chemicals. *Environ Molecular Mutagenesis* 37:55-69.

Pugh WI, Hadgraft J. 1994. *Ab initio* prediction of human skin permeability coefficients. *Int J Pharm* 103:163-178.

Richard AM, Benigni R. 2001. AI and SAR approaches for predicting chemical carcinogenicity: survey and status report. *SAR QSAR Environ Res 13:1-19.*

Ridings JE, Barratt MD, Cary R, Earnshaw CG, Eggington CE, Ellis MK, Judson PN, Langowski JJ, Marchant CA, Payne MP, Watson WP, Yih TD. 1996. Computer prediction

of possible toxic action from chemical structure: an update on the DEREK system. *Toxicology* 106:267-279.

Rimkus GG, Butte W, Geyer HJ. 1997. Critical considerations on the analysis and bioaccumulation of musk xylene and other synthetic nitro musks in fish. *Chemosphere* 35:1497-1507.

Roberts DW. 2002. Application of octanol/water partition coefficients in surfactant science: A Quantitative Structure - Property Relationship for micellization of anionic surfactants. *Langmuir* 18:345-352.

Roberts MS, Pugh WJ, Hadgraft J. 1996. Epidermal permeability: penetrant structure relationships. 2. The effect of H-bonding groups in penetrants on their diffusion through the stratum corneum. *Int J Pharm* 132:23-32.

Robinson MK, McFadden JP, Basketter DA. 2001. Validity and ethics of the human 4-h patch test as an alternative method to assess acute skin irritation potential. *Contact Dermatitis* 45:1-12.

Roex EWM, van Gestel CAM, van Wenzel AP, van Straalen NM. 2000. Ratios between acute aquatic toxicity and effects in population growth rates in relation to toxicant mode of action. *Environ Toxicol Chem* 19:685-693.

Rorije E, Peijnenburg WJGM, Klopman G. 1998. Structural requirements for anaerobic biodegradation of organic chemicals: a fragment model analysis. *Environ Toxicol Chem* 17:1943-1950.

Rosenkranz HS, Klopman G. 1990. Prediction of the carcinogenicity in rodents of chemicals currently being tested by the US National Toxicology Program: structure activity correlations. *Mutagenesis* 5:425-432.

Russom CL, Bradbury SP, Broderius SJ, Hammermeister DE, Drummond RA. 1997. Predicting modes of toxic action from chemical structure: acute toxicity in fathead minnow (*Pimephales promelas*). *Environ Toxicol Chem* 16:948-967.

Sabljic A, Güsten H. 1995. In Hermens JLM, ed, *QSARs for soil sorption*, in *Overview of structure-activity relationships for environmental endpoints*. Report prepared within the framework of the project "QSAR for prediction of fate and effects of chemicals in the environment", an international project of the Environmental Technologies RTD Programme (DG XII/D-1) of the European Commission under contract number EV5V-CT92-0211.

Sage ML, Sage GW. 2000. Vapor pressure. In Boethling RS, Mackay D, eds, *Handbook*

of Property Estimation Methods for Chemicals – Vol 3. Lewis Publishers, Boca Raton, Fl, USA, pp53-65.

Sanderson DM, Earnshaw CG. 1991. Computer prediction of possible toxic action from chemical structure. The DEREK system. *Hum Exp Toxicol* 10:261-273.

Schüürmann G, Kühne R, Kleint F, Ebert R-U, Rothenbacher C, Herth P 1997. A software system for automatic chemical property estimation from molecular structure. In Chen F, Schüürmann G, eds, *Quantitative Structure-Activity Relationships in Environmental Sciences*, Vol VII. SETAC Press, Pensacola, FL, USA, pp. 93-114.

Seaman CW, Guerriero FJ, Sprague GL. 2001. The use of DEREK (a structure/toxicity prediction program) in the identification of skin sensitizers. *Toxicologist* 60:1452 (abstract).

Seth R, Mackay D, Muncke J. 1999. Estimating the organic carbon partition coefficient and its variability for hydrophobic chemicals. *Environ Sci Technol* 33:2390-2394.

Smithing MP, Darvas F. 1992. HazardExpert: an expert system for predicting chemical toxicity. In: Finlay JW, Robinson SF, Armstron DJ, eds, *Food safety assessment*. American Chemical Society, Washington, DC, USA, pp191-200.

Staples CA, Weeks J, Hall JF, Naylor CG. 1998. Evaluation of aquatic toxicity and bioaccumulation of C_8 and C_9-alkylphenol ethoxylates. *Environ Toxicol Chem* 17:2470-2480.

Stegeman MHL, Peijnenburg WJGM, Verboom H. 1993. A quantitative structure-activity relationship for the direct photohydrolysis of meta-substituted halobenzene derivatives in water. *Chemosphere* 5:837-849.

Stein SE, Brown RL. 1994. Estimation of normal boiling points from group contributions. *J Chem Inf Comput Sci* 34:581-587.

Stephan CE, Mount DI, Hansen DJ, Gentile JH, Chapman GA, Brungs WA. 1985. Guidelines for deriving numerical water quality criteria for the protection of aquatic organisms and their use. *National Technical Information Service*, Springfield, VA, USA.

Tesconi M, Yalkowsky SH. 2000. Melting point. In Boethling RS, Mackay D, eds, *Handbook of Property Estimation Methods for Chemicals* – Vol 1. Lewis Publishers, Boca Raton, Fl, USA, pp3-27.

Tetko IV, Tanchuk VY, Kasheva T, Villa AEP. 2001. Internet software for the calculation of the lipophilicity and aqueous solubility of chemical compounds. *J Chem Inf Comput Sci* 41:246-252.

Tolls J, Haller M, De Graaf I, Thijssen MATC, Sijm DHTM. 1997. Bioconcentration of

LAS: experimental determination and extrapolation to environmental mixtures. *Environ Sci Technol* 31:426-431.

Tolls J, Haller M, Labee E, Verweij M, Sijm DTHM. 2000. Experimental determination of bioconcentration of the nonionic surfactant alcohol ethoxylate. *Environ Toxicol Chem* 19:646-653.

Tolls J, van Dijk J, Bioconcentration of n-dodecane and its highly branched isomer 2,2,4,6,6-pentamethylheptane in fathead minnows *Chemosphere* 2002, 47, 1049-1057.

Tsuda T, Takino A, Muraki K, Harada H, Kojima M. 2001. Evaluation of 4-nonylphenols and 4-tert-octylphenol contamination of fish in rivers by laboratory accumulation and excretion experiments. *Water Res* 35:1786-1792.

Urrestarazu Ramos E. 1998. Aquatic toxicity of polar narcotic pollutants: mechanisms, modelling and environmental effects assessment, *PhD thesis, Research Institute of Toxicology*, Utrecht University, Utrecht, The Netherlands, 197 pp.

US EPA (OPPT). 1997. "Improved Method for Estimating Bioconcentration Factor (BCF) from Octanol-Water Partition Coefficient", SRC TR-97-006 (2nd Update), July 22, 1997; prepared for: Robert S. Boethling, EPA-OPPT, Washington, DC; Contract No. 68-D5-0012; prepared by: William M. Meylan PH, Aronson D, Printup H, Gouchie S. Syracuse Research Corp., Environmental Science Center, 6225 Running Ridge Road, North Syracuse, NY 13212.

US EPA. 1999. ASTER user guide, assessment tools for the evaluation of risk. Version 1.0. Duluth, MN, USA.

US patent. 2001. US patent 6,036,349 Optimum Prediction Space for mathematical models V. Gombar

Van Egmond R, Hambling S, Marshall S. 1999. Bioconcentration, biotransformation and chronic toxicity of sodium laurate to zebrafish (*Danio rerio*). *Environ Toxicol Chem* 18:466-473.

Van Haelst AG, Loonen H, van der Wielen FWM, Govers HAJ. 1996. Comparison of bioconcentration factors of tetrachlorobenzyltoluenes in the guppy (*Poecilia reticulata*) and zebra mussel (*Dreissena polymorpha*). *Chemosphere* 32:1117-1122.

van Leeuwen CJ, van der Zandt PTJ, Aldenberg T, Verhaar HJM, Hermens JLM. 1992. Application of QSARs, extrapolation and equilibrium partitioning in aquatic effects assessment. I. Pollutants that act by narcosis. *Environ Toxicol Chem* 11:267-282.

Veith GD, Kosian P. 1983. Estimating bioconcentration potential from octanol/water

partition coefficients. In Mackay D, Peterson S, Eisenreich SJ, Simons MS, eds, *Physical behaviour of PCBs in the Great Lakes.* Ann Arbor Science Publishers, Ann Arbor, USA, pp269-282.

Veith GD, DeFoe DL, Bergstedt BV. 1979. Measuring and estimating the bioconcentration factor of chemicals in fish. *J Fish Res Board Can* 36:1040-1048.

Veith GD, Macek KJ, Petrocelli R, Carrol J, 1980. An evaluation of using partition coefficients and water solubility to estimate bioconcentration factors for organic chemicals in fish. In Eaton JG, Parrish PR, Hendricks AC, eds, *Aquatic toxicology*, ASTM STP 707, Philadelphia, USA, pp116-129.

Verhaar HJM, van Leeuwen CJ, Hermens JLM. 1992. Classifying environmental pollutants. 1. Structure-activity relationships for prediction of aquatic toxicity. *Chemosphere* 25:471-491.

Verhaar HJ, van Leeuwen CJ, Bol J, Hermens JLM. 1994. Application of QSARs in risk management of exisitng chemicals. *SARs QSARs Env Res* 2:39-58.

Verhaar HJM, Solbe J, Speksnijden J, van Leeuwen CJ, Hermens JLM. 2000. Classifying environmental pollutants: part 3, external validation of the classification system. *Chemosphere* 40:875-883.

Wolfe NL, Jeffers PM. 2000. Hydrolysis. In Boethling RS, Mackay D, eds, *Handbook of Property Estimation Methods for Chemicals* –Vol 13. Lewis Publishers, Boca Raton, Fl, USA, pp311-333.

Woo Y, Lai DY, Argus MF, Arcos JC. 1995. Development of structure-activity relationship rules for predicting carcinogenic potential of chemicals. *Toxicol Lett* 79:219-228.

Woo Y, Lai DL, Arcos J, Argus M. 1997. Mechanism-based structure-activity relationship (SAR) analysis of carcinogenic potential of 30 NTP test chemicals. *Environ Carcino Ecotoxi Revs* C15:139-160.

Worth A, Cronin MTD, van Leeuwen CJ. 2003. A framework for promoting the acceptance and regulatory use of (Q)SARs. Submitted to *Modelling Environ Fate Toxicity*.

Yalkowski SH, Banerjee S. 1992. Aqueous solubility: methods of estimation for organic chemicals. Marcel Dekker Inc., NY, USA.

Zhang YP, Sussman N, Macina OT, Rosenkranz HS, Klopman G. 1996. Prediction of the carcinogenicity of a second group of organic chemicals undergoing carcinogenicity testing. *Environ Health Perspect* 104:1059-1060.

Zinke S, Gerner I. 2000. Local irritation/corrosion testing strategies: extending a decision support system by applying self-learning classifiers. *ATLA* 28:651-663.

Zinke S, Gerner I, Graetschel G, Schlede E. 2000. Local irritation/corrosion testing

strategies: development of a decision support system for the introduction of alternative methods. *ATLA* 28:29-40.

Zinke S, Gerner I, Schlede E. 2002. Evaluation of a rule base for identifying contact allergens by using a regulatory database: comparison of data on chemicals notified in the European Union with "structural alerts" used in the DEREK system. *ATLA* 30:285-298.

MEMBERS OF THE TASK FORCE

J. Jaworska (Chairman)	Procter and Gamble B – Brussels
A. Buard	Rhodia Services F – Lyon
E. Delbanco	Henkel D – Düsseldorf
W. de Wolf	DuPont de Nemours B – Mechelen
F. Dreher	L'Oréal F – Aulnay-sous-Bois Cedex
M. Geurts	Akzo Nobel Chemicals NL – Arnhem
M. Hetheridge	AstraZeneca UK – Brixham
E. Jacob	BASF D – Ludwigshafen
P. Kern	Procter and Gamble B – Brussels
L. Low	ExxonMobil USA – Annandale
L. Mueller	Novartis Pharma AG CH – Basel
T. Parkerton	ExxonMobil B – Machelen
U. Schmidt	Bayer AG D – Wuppertal
B. Simon-Hettich	Merck KgaA D – Darmstadt
V. Soballa	Degussa D – Trostberg
M. Holt	ECETOC B – Brussels

MEMBERS OF THE SCIENTIFIC COMMITTEE

(Peer Review Committee)

Dr. B. Hildebrand (Chairman)	Consultant D – Weinheim
E. Bomhard Industrial Toxicologist	Bayer D – Wuppertal
C. Braun Occupational Toxicologist	Akzo Nobel NL – Arnhem
P. Calow Professor of Zoology	University of Sheffield UK – Sheffield
N. Carmichael Head, Toxicology	Bayer CropScience F – Sophia Antipolis
C. d'Hondt Head, Environmental Safety Department	Syngenta CH – Basel
P. Douben Head, SEAC Environmental Protection Department	Unilever UK – Sharnbrook
T. Feijtel[a] Manager, Professional and Regulatory Services	Procter and Gamble B – Brussels
H. Greim Director, Institute of Toxicology and Environmental Hygiene	Technical University Munich D – Munich
C. Money Industrial Hygiene Adviser – Europe	ExxonMobil B – Brussels
G. Randall Director, Environmental Laboratory	AstraZeneca UK – Brixham
A. Sarrif Director, Health and Environmental Sciences	DuPont B – Brussels
G. Swaen Head, Occupational Epidemiology Unit	Maastricht University NL – Maastricht
B. van Ravenzwaay Director, Experimental Toxicology and Ecology	BASF D – Ludwigshafen
H-J. Wiegand[a] Head, Product Safety Department	Degussa D – Marl

[a] Steward responsible for peer review

ECETOC PUBLISHED REPORTS

Monographs

No.	Title
No. 1	Good Laboratory Practice
No. 2	A Contribution to Strategy for Identification and Control of Occupational Carcinogens
No. 3	Risk Assessment of Occupational Chemical Carcinogens
No. 4	Hepatocarcinogenesis in Laboratory Rodents: Relevance for Man
No. 5	Identification and Assessment of the Effects of Chemicals on Reproduction and Development (Reproductive Toxicology)
No. 6	Acute Toxicity Tests, LD_{50} (LC50) Determinations and Alternatives
No. 7	Recommendations for the Harmonisation of International Guidelines for Toxicity Studies
No. 8	Structure-Activity Relationships in Toxicology and Ecotoxicology: An Assessment (Summary)
No. 9	Assessment of Mutagenicity of Industrial and Plant Protection Chemicals
No. 10	Identification of Immunotoxic Effects of Chemicals and Assessment of their Relevance to Man
No. 11	Eye Irritation Testing
No. 12	Alternative Approaches for the Assessment of Reproductive Toxicity (with emphasis on embryotoxicity/teratogenicity)
No. 13	DNA and Protein Adducts: Evaluation of their Use in Exposure Monitoring and Risk Assessment
No. 14	Skin Sensitisation Testing
No. 15	Skin Irritation
No. 16	Early Indicators of Non-Genotoxic Carcinogenesis
No. 17	Hepatic Peroxisome Proliferation
No. 18	Evaluation of the Neurotoxic Potential of Chemicals
No. 19	Respiratory Allergy
No. 20	Percutaneous Absorption
No. 21	Immunotoxicity: Hazard Identification and Risk Characterisation
No. 22	Evaluation of Chemicals for Oculotoxicity
No. 23	Receptor Mediated Mechanisms in Chemical Carcinogenesis
No. 24	Risk Assessment for Carcinogens
No. 25	Practical Concepts for Dose Selection in Chronic Toxicity and Carcinogenicity Studies in Rodents
No. 26	Aquatic Toxicity Testing of Sparingly Soluble Volatile and Unstable Substances
No. 27	Aneuploidy
No. 28	Threshold-Mediated Mutagens - Mutation Research Special Issue
No. 29	Skin Sensitisation Testing for the Purpose of Hazard Identification and Risk Assessment
No. 30	Genetic Susceptibility to Environmental Toxicants
No. 31	Guidance on Evaluation of Reproductive Toxicity Data
No. 32	Use of Human Data in Hazard Classification for Irritation and Sensitisation
No. 33	Application of Physiological - Toxicokinetic Modelling to Health Hazard Assessment of Chemcial Substances

Technical Reports

No.	Title
No. 1	Assessment of Data on the Effects of Formaldehyde on Humans
No. 2	The Mutagenic and Carcinogenic Potential of Formaldehyde
No. 3	Assessment of Test Methods for Photodegradation of Chemicals in the Environment
No. 4	The Toxicology of Ethylene Glycol Monoalkyl Ethers and its Relevance to Man
No. 5	Toxicity of Ethylene Oxide and its Relevance to Man
No. 6	Formaldehyde Toxicology: An Up-Dating of ECETOC Technical Reports 1 and 2
No. 7	Experimental Assessment of the Phototransformation of Chemicals in the Atmosphere
No. 8	Biodegradation Testing: An Assessment of the Present Status
No. 9	Assessment of Reverse-Phase Chromatographic Methods for Determining Partition Coefficients
No. 10	Considerations Regarding the Extrapolation of Biological Data in Deriving Occupational Exposure Limits
No. 11	Ethylene Oxide Toxicology and its Relevance to Man: An Up-Dating of ECETOC Technical Report No. 5
No. 12	The Phototransformation of Chemicals in Water: Results of a Ring-Test
No. 13	The EEC 6th Amendment: A Guide to Risk Evaluation for Effects on the Environment
No. 14	The EEC 6th Amendment: A Guide to Risk Evaluation for Effects on Human Health
No. 15	The Use of Physical-Chemical Properties in the 6th Amendment and their Required Precision, Accuracy and Limiting Values
No. 16	A Review of Recent Literature on the Toxicology of Benzene
No. 17	The Toxicology of Glycol Ethers and its Relevance to Man: An Up-Dating of ECETOC Technical Report No. 4
No. 18	Harmonisation of Ready Biodegradability Tests
No. 19	An Assessment of Occurrence and Effects of Dialkyl-o-Phthalates in the Environment
No. 20	Biodegradation Tests for Poorly-Soluble Compounds
No. 21	Guide to the Classification of Carcinogens, Mutagens, and Teratogens under the 6th Amendment
No. 22	Classification of Dangerous Substances and Pesticides in the EEC Directives. A Proposed Revision of Criteria for Inhalational Toxicity
No. 23	Evaluation of the Toxicity of Substances to be Assessed for Biodegradability
No. 24	The EEC 6th Amendment: Prolonged Fish Toxicity Tests
No. 25	Evaluation of Fish Tainting
No. 26	The Assessment of Carcinogenic Hazard for Human Beings exposed to Methylene Chloride
No. 27	Nitrate and Drinking Water
No. 28	Evaluation of Anaerobic Biodegradation
No. 29	Concentrations of Industrial Organic Chemicals Measured in the Environment: The Influence of Physico-Chemical Properties, Tonnage and Use Patterns
No. 30	Existing Chemicals: Literature Reviews and Evaluations (Fifth Edition) (No longer available)
No. 31	The Mutagenicity and Carcinogenicity of Vinyl Chloride: A Historical Review and Assessment
No. 32	Methylene Chloride (Dichloromethane): Human Risk Assessment Using Experimental Animal Data
No. 33	Nickel and Nickel Compounds: Review of Toxicology and Epidemiology with Special Reference to Carcinogenesis
No. 34	Methylene Chloride (Dichloromethane): An Overview of Experimental Work Investigating Species Differences in Carcinogenicity and their Relevance to Man
No. 35	Fate, Behaviour and Toxicity of Organic Chemicals Associated with Sediments
No. 36	Biomonitoring of Industrial Effluents
No. 37	Tetrachlorethylene: Assessment of Human Carcinogenic Hazard
No. 38	A Guide to the Classification of Preparations Containing Carcinogens, Mutagens and Teratogens

No. 39 Hazard Assessment of Floating Chemicals After an Accidental Spill at Sea
No. 40 Hazard Assessment of Chemical Contaminants in Soil
No. 41 Human Exposure to N-Nitrosamines, their Effects and a Risk Assessment for N-Nitrosodiethanolamine in Personal Care Products
No. 42 Critical Evaluation of Methods for the Determination of N-Nitrosamines in Personal Care and Household Products
No. 43 Emergency Exposure Indices for Industrial Chemicals
No. 44 Biodegradation Kinetics
No. 45 Nickel, Cobalt and Chromium in Consumoducts: Allergic Contact Dermatitis
No. 46 EC 7th Amendment: Role of Mammalian Toxicokinetic and Metabolic Studies in the Toxicological Assessment of Industrial Chemicals
No. 47 EC 7th Amendment "Toxic to Reproduction": Guidance on Classification
No. 48 Eye Irritation: Reference Chemicals Data Bank (Second Edition)
No. 49 Exposure of Man to Dioxins: A Perspective on Industrial Waste Incineration
No. 50 Estimating Environmental Concentrations of Chemicals using Fate and Exposure Models
No. 51 Environmental Hazard Assessment of Substances
No. 52 Styrene Toxicology Investigation on the Potential for Carcinogenicity
No. 53 DHTDMAC: Aquatic and Terrestrial Hazard Assessment (CAS No. 61789-80-8)
No. 54 Assessment of the Biodegradation of Chemicals in the Marine Environment
No. 55 Pulmonary Toxicity of Polyalkylene Glycols
No. 56 Aquatic Toxicity Data Evaluation
No. 57 Polypropylene Production and Colorectal Cancer
No. 58 Assessment of Non-Occupational Exposure to Chemicals
No. 59 Testing for Worker Protection
No. 60 Trichloroethylene: Assessment of Human Carcinogenic Hazard
No. 61 Environmental Exposure Assessment
No. 62 Ammonia Emissions to Air in Western Europe
No. 63 Reproductive and General Toxicology of some Inorganic Borates and Risk Assessment for Human Beings
No. 64 The Toxicology of Glycol Ethers and its Relevance to Man
No. 65 Formaldehyde and Human Cancer Risks
No. 66 Skin Irritation and Corrosion: Reference Chemicals Data Bank
No. 67 The Role of Bioaccumulation in Environmental Risk Assessment: The Aquatic Environment and Related Food Webs
No. 68 Assessment Factors in Human Health Risk Assessment
No. 69 Toxicology of Man-Made Organic Fibres
No. 70 Chronic Neurotoxicity of Solvents
No. 71 Inventory of Critical Reviews on Chemicals (Only available to ECETOC members)
No. 72 Methyl tert-Butyl Ether (MTBE) Health Risk Characterisation
No. 73 The Value of Aquatic Model Ecosystem Studies in Ecotoxicology
No. 74 QSARs in the Assessment of the Environmental Fate and Effects of Chemicals
No. 75 Organophosphorus Pesticides and Long-term Effects on the Nervous System
No. 76 Monitoring and Modelling of Industrial Organic Chemicals, with Particular Reference to Aquatic Risk Assessment
No. 77 Skin and Respiratory Sensitisers: Reference Chemicals Data Bank
No. 78 Skin Sensitisation Testing: Methodological Considerations
No. 79 Exposure Factors Sourcebook for European Populations (with Focus on UK Data)
No. 80 Aquatic Toxicity of Mixtures
No. 81 Human Acute Intoxication from Monochloroacetic Acid: Proposals for Therapy
No. 82 Risk Assessment in Marine Environments
No. 83 The Use of T25 Estimates and Alternative Methods in the Regulatory Risk Assessment of Non-threshold Carcinogens in the European Union
No. 84 Scientific Principles for Soil Hazard Assessment of Substances

No. 85 Recognition of, and Differentiation between, Adverse and Non-adverse Effects in Toxicology Studies
No. 86 Derivation of Assessment Factors for Human Health Risk Assessment
No. 87 Contact Sensitisation: Classification According to Potency
No. 88 Environmental Risk Assessment of Difficult Substances

Joint Assessment of Commodity Chemicals (JACC) Reports

No. Title

No. 1 Melamine
No. 2 1,4-Dioxane
No. 3 Methyl Ethyl Ketone
No. 4 Methylene Chloride
No. 5 Vinylidene Chloride
No. 6 Xylenes
No. 7 Ethylbenzene
No. 8 Methyl Isobutyl Ketone
No. 9 Chlorodifluoromethane
No. 10 Isophorone
No. 11 1,2-Dichloro-1,1-Difluoroethane (HFA-132b)
No. 12 1-Chloro-1,2,2,2-Tetrafluoroethane (HFA-124)
No. 13 1,1-Dichloro-2,2,2-Trifluoroethane (HFA-123)
No. 14 1-Chloro-2,2,2-Trifluoromethane (HFA-133a)
No. 15 1-Fluoro 1,1-Dichloroethane (HFA-141B)
No. 16 Dichlorofluoromethane (HCFC-21)
No. 17 1-Chloro-1,1-Difluoroethane (HFA-142b)
No. 18 Vinyl Acetate
No. 19 Dicyclopentadiene (CAS: 77-73-6)
No. 20 Tris-/Bis-/Mono-(2 ethylhexyl) Phosphate
No. 21 Tris-(2-Butoxyethyl)-Phosphate (CAS:78-51-3)
No. 22 Hydrogen Peroxide (CAS: 7722-84-1)
No. 23 Polycarboxylate Polymers as Used in Detergents
No. 24 Pentafluoroethane (HFC-125) (CAS: 354-33-6)
No. 25 1-Chloro-1,2,2,2-tetrafluoroethane (HCFC 124) (CAS No. 2837-89-0)
No. 26 Linear Polydimethylsiloxanes (CAS No. 63148-62-9)
No. 27 n-Butyl Acrylate (CAS No. 141-32-2)
No. 28 Ethyl Acrylate (CAS No. 140-88-5)
No. 29 1,1-Dichloro-1-Fluoroethane (HCFC-141b) (CAS No. 1717-00-6)
No. 30 Methyl Methacrylate (CAS No. 80-62-6)
No. 31 1,1,1,2-Tetrafluoroethane (HFC-134a) (CAS No. 811-97-2)
No. 32 Difluoromethane (HFC-32) (CAS No. 75-10-5)
No. 33 1,1-Dichloro-2,2,2-Trifluoroethane (HCFC-123) (CAS No. 306-83-2)
No. 34 Acrylic Acid (CAS No. 79-10-7)
No. 35 Methacrylic Acid (CAS No. 79-41-4)
No. 36 n-Butyl Methacrylate; Isobutyl Methacrylate (CAS No. 97-88-1) (CAS No. 97-86-9)
No. 37 Methyl Acrylate (CAS No. 96-33-3)
No. 38 Monochloroacetic Acid (CAS No. 79-11-8) and its Sodium Salt (CAS No. 3926-62-3)
No. 39 Tetrachloroethylene (CAS No. 127-18-4)
No. 40 Peracetic Acid (CAS No. 79-21-0) and its Equilibrium Solutions

Special Reports

No.	Title
No. 8	HAZCHEM; A Mathematical Model for Use in Risk Assessment of Substances
No. 9	Styrene Criteria Document
No. 10	Hydrogen Peroxide OEL Criteria Document (CAS No. 7722-84-1)
No. 11	Ecotoxicology of some Inorganic Borates
No. 12	1,3-Butadiene OEL Criteria Document (Second Edition) (CAS No. 106-99-0)
No. 13	Occupational Exposure Limits for Hydrocarbon Solvents
No. 14	n-Butyl Methacrylate and Isobutyl Methacrylate OEL Criteria Document
No. 15	Examination of a Proposed Skin Notation Strategy
No. 16	GREAT-ER User Manual

Documents

No.	Title
No. 32	Environmental Oestrogens: Male Reproduction and Reproductive Development
No. 33	Environmental Oestrogens: A Compendium of Test Methods
No. 34	The Challenge Posed by Endocrine-disrupting Chemicals
No. 35	Exposure Assessment in the Context of the EU Technical Guidance Documents on Risk Assessment of Substances
No. 36	Comments on OECD Draft Detailed Review Paper: Appraisal of Test Methods for Sex-Hormone Disrupting Chemicals
No. 37	EC Classification of Eye Irritancy
No. 38	Wildlife and Endocrine Disrupters: Requirements for Hazard Identification
No. 39	Screening and Testing Methods for Ecotoxicological Effects of Potential Endocrine Disrupters: Response to the EDSTAC Recommendations and a Proposed Alternative Approach
No. 40	Comments on Recommendation from Scientific Committee on Occupational Exposure Limits for 1,3-Butadiene
No. 41	Persistent Organic Pollutants (POPs) Response to UNEP/INC/CEG-I Annex 1
No. 42	Genomics, Transcript Profiling, Proteomics and Metabonomics (GTPM). An Introduction
No. 43	Contact Sensitisation: Classification According to Potency, A Commentary