Organisation de Coopération et de Développement Economiques
Organisation for Economic Co-operation and Development

15-Feb-2007

English - Or. English

ENVIRONMENT DIRECTORATE
JOINT MEETING OF THE CHEMICALS COMMITTEE AND
THE WORKING PARTY ON CHEMICALS, PESTICIDES AND BIOTECHNOLOGY

Cancels & replaces the same document of 14 August 2006

REPORT ON THE REGULATORY USES AND APPLICATIONS IN OECD MEMBER COUNTRIES OF (QUANTITATIVE) STRUCTURE-ACTIVITY RELATIONSHIP [(Q)SAR] MODELS IN THE ASSESSMENT OF NEW AND EXISTING CHEMICALS

JT03221927

Document complet disponible sur OLIS dans son format d'origine
Complete document available on OLIS in its original format

ENV/JM/MONO(2006)25
Unclassified

English - Or. English

OECD Environment Health and Safety Publications

Series on Testing and Assessment

No. 58

REPORT ON THE REGULATORY USES AND APPLICATIONS IN OECD MEMBER COUNTRIES OF (QUANTITATIVE) STRUCTURE-ACTIVITY RELATIONSHIP [(Q)SAR] MODELS IN THE ASSESSMENT OF NEW AND EXISTING CHEMICALS

Environment Directorate

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

Paris, 2006

Also published in the Series on Testing and Assessment:

No. 1, *Guidance Document for the Development of OECD Guidelines for Testing of Chemicals (1993; reformatted 1995, revised 2006)*

No. 2, *Detailed Review Paper on Biodegradability Testing (1995)*

No. 3, *Guidance Document for Aquatic Effects Assessment (1995)*

No. 4, *Report of the OECD Workshop on Environmental Hazard/Risk Assessment (1995)*

No. 5, *Report of the SETAC/OECD Workshop on Avian Toxicity Testing (1996)*

No. 6, *Report of the Final Ring-test of the Daphnia magna Reproduction Test (1997)*

No. 7, *Guidance Document on Direct Phototransformation of Chemicals in Water (1997)*

No. 8, *Report of the OECD Workshop on Sharing Information about New Industrial Chemicals Assessment (1997)*

No. 9, *Guidance Document for the Conduct of Studies of Occupational Exposure to Pesticides during Agricultural Application (1997)*

No. 10, *Report of the OECD Workshop on Statistical Analysis of Aquatic Toxicity Data (1998)*

No. 11, *Detailed Review Paper on Aquatic Testing Methods for Pesticides and industrial Chemicals (1998)*

No. 12, *Detailed Review Document on Classification Systems for Germ Cell Mutagenicity in OECD Member Countries (1998)*

No. 13, *Detailed Review Document on Classification Systems for Sensitising Substances in OECD Member Countries 1998)*

No. 14, *Detailed Review Document on Classification Systems for Eye Irritation/Corrosion in OECD Member Countries (1998)*

No. 15, *Detailed Review Document on Classification Systems for Reproductive Toxicity in OECD Member Countries (1998)*

No. 16, *Detailed Review Document on Classification Systems for Skin Irritation/Corrosion in OECD Member Countries (1998)*

No. 17, *Environmental Exposure Assessment Strategies for Existing Industrial Chemicals in OECD Member Countries (1999)*

No. 18, *Report of the OECD Workshop on Improving the Use of Monitoring Data in the Exposure Assessment of Industrial Chemicals (2000)*

No. 19, *Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals used in Safety Evaluation (1999)*

No. 20, *Revised Draft Guidance Document for Neurotoxicity Testing (2004)*

No. 21, *Detailed Review Paper: Appraisal of Test Methods for Sex Hormone Disrupting Chemicals (2000)*

No. 22, *Guidance Document for the Performance of Out-door Monolith Lysimeter Studies (2000)*

No. 23, *Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures (2000)*

No. 24, *Guidance Document on Acute Oral Toxicity Testing (2001)*

No. 25, *Detailed Review Document on Hazard Classification Systems for Specifics Target Organ Systemic Toxicity Repeated Exposure in OECD Member Countries (2001)*

No. 26, *Revised Analysis of Responses Received from Member Countries to the Questionnaire on Regulatory Acute Toxicity Data Needs (2001)*

No 27, *Guidance Document on the Use of the Harmonised System for the Classification of Chemicals Which are Hazardous for the Aquatic Environment (2001)*

No 28, *Guidance Document for the Conduct of Skin Absorption Studies (2004)*

No 29, *Guidance Document on Transformation/Dissolution of Metals and Metal Compounds in Aqueous Media (2001)*

No 30, *Detailed Review Document on Hazard Classification Systems for Mixtures (2001)*

No 31, *Detailed Review Paper on Non-Genotoxic Carcinogens Detection: The Performance of In-Vitro Cell Transformation Assays (draft)*

No. 32, *Guidance Notes for Analysis and Evaluation of Repeat-Dose Toxicity Studies (2000)*

No. 33, *Harmonised Integrated Classification System for Human Health and Environmental Hazards of Chemical Substances and Mixtures (2001)*

No. 34, *Guidance Document on the Development, Validation and Regulatory Acceptance of New and Updated Internationally Acceptable Test Methods in Hazard Assessment (2005)*

No. 35, *Guidance notes for analysis and evaluation of chronic toxicity and carcinogenicity studies (2002)*

No. 36, *Report of the OECD/UNEP Workshop on the use of Multimedia Models for estimating overall Environmental Persistence and long range Transport in the context of PBTS/POPS Assessment (2002)*

No. 37, *Detailed Review Document on Classification Systems for Substances Which Pose an Aspiration Hazard (2002)*

No. 38, *Detailed Background Review of the Uterotrophic Assay Summary of the Available Literature in Support of the Project of the OECD Task Force on Endocrine Disrupters Testing and Assessment (EDTA) to Standardise and Validate the Uterotrophic Assay (2003)*

No. 39, *Guidance Document on Acute Inhalation Toxicity Testing (in preparation)*

No. 40, *Detailed Review Document on Classification in OECD Member Countries of Substances and Mixtures Which Cause Respiratory Tract Irritation and Corrosion (2003)*

No. 41, *Detailed Review Document on Classification in OECD Member Countries of Substances and Mixtures which in Contact with Water Release Toxic Gases (2003)*

No. 42, *Guidance Document on Reporting Summary Information on Environmental, Occupational and Consumer Exposure (2003)*

No. 43, *Draft Guidance Document on Reproductive Toxicity Testing and Assessment (in preparation)*

No. 44, *Description of Selected Key Generic Terms Used in Chemical Hazard/Risk Assessment (2003)*

No. 45, *Guidance Document on the Use of Multimedia Models for Estimating Overall Environmental Persistence and Long-range Transport (2004)*

No. 46, *Detailed Review Paper on Amphibian Metamorphosis Assay for the Detection of Thyroid Active Substances (2004)*

No. 47, *Detailed Review Paper on Fish Screening Assays for the Detection of Endocrine Active Substances (2004)*

No. 48, *New Chemical Assessment Comparisons and Implications for Work Sharing (2004)*

No. 49, *Report from the Expert Group on (Quantitative) Structure-Activity Relationships [(Q)SARs] on the Principles for the Validation of (Q)SARs (2004)*

No. 50, *Report of the OECD/IPCS Workshop on Toxicogenomics (2005)*

No. 51, *Approaches to Exposure Assessment in OECD Member Countries: Report from the Policy Dialogue on Exposure Assessment in June 2005 (2006)*

No. 52, *Comparison of emission estimation methods used in Pollutant Release and Transfer Registers (PRTRs) and Emission Scenario Documents (ESDs): Case study of pulp and paper and textile sectors (2006)*

No. 53, *Guidance Document on Simulated Freshwater Lentic Field Tests (Outdoor Microcosms and Mesocosms) (2006)*

No. 54, *Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application (2006)*

No. 55, *Detailed Review Paper on Aquatic Arthropods in Life Cycle Toxicity Tests with an Emphasis on Developmental, Reproductive and Endocrine Disruptive Effects (2006)*

No. 56, *Guidance Document on the Breakdown of Organic Matter in Litter Bags (2006)*

No. 57, *Detailed Review Paper on Thyroid Hormone Disruption Assays (2006)*

No. 58, *Report on the Regulatory Uses and Applications in OECD Member Countries of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models in the Assessment of New and Existing Chemicals (2006)*

© OECD 2006
Applications for permission to reproduce or translate all or part of this material should be made to: Head of Publications Service, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France

About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 30 industrialised countries in North America, Europe and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in ten different series: **Testing and Assessment**; **Good Laboratory Practice and Compliance Monitoring**; **Pesticides and Biocides**; **Risk Management**; **Harmonisation of Regulatory Oversight in Biotechnology**; **Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents; and the Safety of Manufactured Nanomaterials.** More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (http://www.oecd.org/ehs/).

This publication was produced within the framework of the Inter-Organisation Programme for the Sound Management of Chemicals (IOMC).

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The participating organisations are FAO, ILO, OECD, UNEP, UNIDO, UNITAR and WHO. The World Bank and UNDP are observers. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

This publication is available electronically, at no charge.

For this and many other Environment, Health and Safety publications, consult the OECD's World Wide Web site (www.oecd.org/ehs/)

or contact:

OECD Environment Directorate, Environment, Health and Safety Division

2 rue André-Pascal
75775 Paris Cedex 16
France

Fax: (33-1) 44 30 61 80

E-mail: ehscont@oecd.org

FOREWORD

This case studies report is part of the OECD effort to develop guidance for regulatory applications of (Quantitative) Structure-Activity Relationship [(Q)SAR] models, and emphasizes the use of programme-specific case studies to highlight the importance of legal and practical constraints and information requirements of individual regulatory programmes within OECD member countries in applying (Q)SAR approaches. This report provides a snapshot of the experiences of OECD member countries with respect to the use of (Q)SAR models in chemical assessment. The document provides both current regulatory uses in OECD member countries as well as prospective regulatory applications -- especially those within the European Union (EU) member states as a result of proposed legislation on chemicals (i.e., Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH)). As such, the case studies report should be regarded as a living document which will be updated periodically as requested by member countries to describe the expanding role of (Q)SAR in regulatory settings.

This report includes a brief description of the activity of OECD Member countries with the aim of enhancing the regulatory acceptance of (Q)SAR models and expanding the opportunities for future applications of the models. Several case studies on (Q)SAR model applications by OECD member countries have been summarized in case studies I-XII of this report, and they illustrate both their present uses and potential utility, as well as possible difficulties, associated with incorporating non-test methods such as (Q)SAR models into the various regulatory frameworks. The reports will also demonstrate the difficulties associated with defining universal (Q)SAR principles for acceptability that could fit the demands of all regulatory frameworks. Also described is the recent initiation of the EU Experience Project led by the Netherlands in order to facilitate the sharing of (Q)SAR models in response to legislative initiatives in the EU, and summaries of additional (Q)SAR activities within other OECD bodies.

This report was based on information provided to the Secretariat by the OECD member countries and is not an OECD-wide endorsement of specific (Q)SAR models or approaches. The first draft was produced by the U.S. EPA and circulated to the members of the OECD Ad hoc Group on (Q)SARs for their input in September 2005. Comments were discussed at the meeting of the OECD Steering Group on (Q)SARs in December 2005, and a revised draft was circulated to the Ad hoc Group for final review and endorsement in March 2006 and endorsed in June 2006.

This document is published on the responsibility of the Joint Meeting of the Chemicals Group and Management Committee of the Special Programme on the Control of Chemicals of the OECD.

This document has been produced with the financial assistance of the European Union. The views expressed herein can in no way be taken to reflect the official opinion of the European Union.

TABLE OF CONTENTS

INTRODUCTION

1. An OECD workshop was held in 1989 on notification schemes for new chemicals. One of its recommendations focused upon the need to evaluate the predictive power of the (Quantitative) Structure-Activity Relationship [(Q)SAR] models used by the U.S. EPA/OPPT by comparing the results of the OPPT's (Q)SAR assessment with those obtained from the base-set testing of new chemicals (physical/chemical, environmental fate, toxicological properties/health effects, and ecotoxicity parameters) required by the European Union (EU, then the European Commission/EC).

2. Soon thereafter, the relative accuracy of these two diverse systems became the subject of a joint EPA/EU evaluation called the "Structure Activity Relationship/Minimum Premarketing Dataset" (SAR/MPD) study. This comparison was undertaken from 1991-1993 and it contrasted the OPPT (Q)SAR predictions with the new chemical base-set test data for 175 chemicals that had been submitted to, and selected by, the EU. The results of the SAR/MPD study were subsequently jointly published in 1994 by the EPA and the EU (OECD) [U.S EPA, 1994a; OECD, 1994] and they confirmed that some EPA/OPPT (Q)SAR methods appeared to be both reliable and useful.

3. Also in the early to mid-1990's, the OECD undertook several additional efforts to assess the usefulness of a variety of (Q)SAR methodologies. In 1990, they held the OECD Workshop on Quantitative Structure Activity Relationships (QSARs) in Aquatic Effects Assessment (OECD, 1992a) and included the subject of (Q)SARs in the 1990 OECD Workshop on the Extrapolation of Laboratory Aquatic Toxicity Data in the Real Environment (OECD, 1992b). The outcomes of those workshops were then used in the development and publication of the OECD's Guidance Document for Aquatic Effects Assessment (OECD, 1995), which contains several sections that discuss the relative merits, and limitations, of using (Q)SARs in the hazard and risk assessment of chemicals found, or expected in, the aquatic environment. In the early 1990's, the OECD also published two other (Q)SAR documents that were based upon member country-led projects. The first was on the Application of Structure Activity Relationships to the Estimation of Properties Important in Exposure Assessment (OECD, 1993a), and the second was on Structure Activity Relationships for Biodegradation (OECD, 1993b)

4. At the 5th International Workshop on QSAR in Environmental Sciences (Duluth, Minnesota; July 1992), the initiative was taken to establish an expert group on QSARs. The work originated from activities in the US EPA and the European Commission related to priority setting of existing chemicals (van der Zandt and van Leeuwen, 1992; Hansen et al., 1999). During this work it appeared that most of the existing chemicals did not have basic public information that would allow a proper ranking (van Leeuwen et al., 1996; Allanou et al., 1999). Similar observations regarding data gaps of existing industrial chemicals were made in the USA (National Research Council, 1984). From 1993 onwards an international group led by the OECD involving the USA, Japan, EU Member States and the European Commission (European Chemicals Bureau, Ispra, Italy) made an inventory of existing SARs and QSARs. These estimation tools were applied in order to provide basic information of existing chemicals with the aim to rank them for further risk assessment. Furthermore, the US EPA provided aquatic toxicity estimates and comparative aquatic toxicity distributions with the EU HPVC's (Clements et al., 1995), Japan provided a complete summary of all their experimental biodegradation data to the OECD, and the Netherlands provided the aquatic toxicity estimates for all narcotic high production volume chemicals (Bol et al., 1993a and 1993b).

5. The OECD's Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology (Joint Meeting) was informed on several occasions of this work (Hansen and van Leeuwen, 1995; Karcher et al., 1995), but work on this activity in the Commission was subsequently discontinued. The relevance of the work was rediscovered a few years later when the chemical industry

took the initiative to put the need for this activity on the international agenda again (Jaworska et al., 2003) and the Commission launched its proposal for REACH (Pedersen et al., 2003). Nontesting alternative assessment methods, i.e. SARs and QSARs, are seen as an essential tool for an effective and efficient assessment of chemicals (Worth et al., 2004a, b; Bradbury et al. 2004)

6. Concerns in Europe about the lack of basic information on chemicals, the slow progress of the assessment of existing chemicals and the relatively high attention to new industrial chemicals necessitated a fundamental review of the EU policies on chemicals. This review was completed in November 1998. With input from stakeholders, the Commission subsequently issued a White Paper on a Strategy for a future Chemicals Policy (CEC, 2001). The proposals in that EU White Paper were subsequently further developed and extensively discussed with major stakeholders, and resulted in their October 2003 proposal for REACH (Registration, Evaluation, and Authorization of Chemicals), a new EU regulatory framework for chemicals management (CEC, 2003).

7. Following circulation of the original White Paper that highlighted a need to assess health and environmental effects in existing chemical inventories, a workshop on "Regulatory Use of (Q)SARs for Human Health and Environmental Endpoints," was held in Setubal, Portugal in March 2002. This workshop was organized by the European Chemical Industry Council (CEFIC) and the International Council of Chemical Associations (ICCA) and gathered more than 60 human health and environmental experts from industry, academia, and regulatory agencies from around the world. Participants noted that while (Q)SARs have been promoted as alternative assessment methods to experimental work for several years, there has been resistance to apply them in a regulatory context, because much of the existing legislation requires the development of experimental test data. In the interim however, (Q)SAR science has been advancing steadily by providing better ways of leveraging statistical/mathematical tools and predicting mechanistic information. Consequently, (Q)SARs are now being increasingly viewed as one of the most cost effective ways to estimate ecological and health effects of chemicals. To increase confidence in (Q)SAR predictions and minimize their misuse, a number of criteria for establishing the state of development and validation of (Q)SARs were agreed upon by the workshop participants (CEFIC, 2002). The recommendations from the workshop were then submitted to the OECD, and were the subject of review as a special activity at the OECD's 34th Joint Meeting of Chemical Committee and the Working Party on Chemicals, Pesticides, and Biotechnology in November 2002.

8. The 34th Joint Meeting held a Special Session on (Quantitative) Structure-Activity Relationships [(Q)SARs], to review and discuss papers and other information submitted from the OECD Member countries and other organisations [ENV/JM/RD(2002)21]. The record of the (Q)SAR Special Session pointed out the need for transparency in (Q)SAR programs and clear procedures for applicability evaluation and validation for (Q)SARs. Based upon this need, the OECD Member countries agreed to initiate an activity to develop an internationally accepted set of criteria, as well as procedures for the evaluation of existing and promising (Q)SAR models. An OECD Expert Group on (Q)SARs was subsequently established to develop a Work Plan, which was initiated at their first meeting in early 2003.

9. The two-year Work Plan eventually proposed by the Expert Group on (Q)SARs was endorsed at the 35th Joint Meeting in mid-2003 [ENV/JM(2003)18] and consisted of three major Work Items including:

- Work Item 1: Apply the specific validation principles agreed upon at the ICCA Workshop on Regulatory Acceptance of (Q)SARs, and the general validation principles for new and updated test methods, to selected (Q)SARS in use.

- Work Item 2: Develop guidance documents for the development, validation and use of (Q)SARs in regulatory assessments.

- Work Item 3: Identify practical approaches to enable (Q)SARs to be readily available and accessible, including the development of a database of accepted (Q)SARs.

10. The acceptability criteria derived at the Industry sponsored ICCA/CEFIC workshop were considered and modified by the OECD (Q)SAR Expert Group at their second meeting in September 2004, and the OECD's edited criteria were brought forth as the OECD principles for (Q)SAR validation for regulatory purposes. In addition, case studies for the use of (Q)SAR models in regulatory applications were reviewed and accepted by that Expert Group, and initial discussions on the development of a (Q)SAR Application Toolbox were also held.

11. The 37th Joint Meeting in November, 2004, approved the OECD principles for the validation, for regulatory purposes, of (Q)SAR models [See Annex] and considered the report from the Expert Group on (Q)SAR, and the U.S. case study on the regulatory applications of (Q)SARs. This completed Work Item 1 to develop a set of validation principles, and attention has now turned to the development of guidance on the validation of (Q)SAR models (Work Item 2) and the opportunities for making (Q)SAR models readily accessible by developing: a.) an overview of case studies on the use of (Q)SAR models in the OECD Member countries [this document]; and b.) a (Q)SAR application toolbox for Member countries (Work Item 3).

12. The report from the (Q)SAR Expert Group, the U.S. case study, and the OECD principles for (Q)SAR validation for regulatory purposes, were the major tangible outcomes from the first year of the two-year OECD Work Programme on (Q)SARs. The focus of the second year (2005) is on the following:

- Development of an OECD Guidance Document on the principles for (Q)SAR validation for regulatory purposes: A drafting team, led by the European Commission, was established to develop detailed and non-prescriptive guidance, which will explain and illustrate the application of the principles to different types of (Q)SAR models.
- Case studies on the regulatory application of (Q)SAR models: The Expert Group reviewed the US case study on the regulatory application of (Q)SARs [ENV/JM/TG(2004)27REV2]. The general conclusions from this case study will be useful in other jurisdictions, taking into account information from the work on battery and weight-of-evidence approaches and from other members of the Expert Group. It will help form the basis for the overview of the case studies on use of (Q)SAR models in OECD Member countries.
- Information exchange and work-sharing in the validation of individual (Q)SAR models: The Expert Group has exchanged information about national/regional initiatives to validate individual (Q)SAR models, notably by the European Commission and Japan. The Expert Group will continue information exchange, and consider how best to mutually benefit from these initiatives and share the burden.
- Guidance for reporting (Q)SAR estimations: Some members of the Expert Group will investigate the possibility and utility of developing guidance on reporting the estimations from (Q)SAR models. If this work progresses, it may be combined into the project on the Guidance Document for validation.

13. The 37th Joint Meeting recognized that, with the shift in focus from development of the validation principles for (Q)SAR models to the regulatory use and application of (Q)SAR models, the general oversight of these OECD (Q)SAR activities would move to the Task Force on Existing Chemicals in consultation with the other OECD groups with interest in the issue such as Test Guidelines, New Chemicals, Classification and Labelling, Pesticides and Biocides. The 37th Joint Meeting also agreed that the "(Q)SAR Expert Group" was changed to "Ad hoc Group on (Q)SARs" and the membership of the Ad hoc Group was re-established to include not only (Q)SAR experts but also those who use (Q)SARs for

regulatory purposes. Following receipt of the nominations from the Member Countries, the 38th Joint Meeting in June 2005 agreed to establish a smaller Steering Group, consisting of those members of the Ad hoc Group who are most closely involved in the planning and routine management of the (Q)SAR project. The development of projects in the OECD (Q)SAR project will be the responsibility of an Ad hoc Group on (Q)SARs.

14. In order to try and address the actual needs of the OECD Member countries and to facilitate the acceptance of (Q)SAR approaches in the regulation and assessment of chemicals, the following activities were agreed to by the 38th Joint Meeting (June 2005), as reflected in the revised (Q)SAR activities after the 38th Joint Meeting [ENV/JM(2005)18/REV1; 13 July 2005; pp. 21-22]

15. The Work Plan for (Q)SARs includes a work item to make acceptable (Q)SAR models available to member countries for regulatory use. This activity will start with the identification of (Q)SAR models that fulfill the needs of member countries in the New and Existing Chemicals Programmes, and that demonstrate consistency with the OECD principles of validation. This work can start this year, aiming at producing the first list of (Q)SAR models with available information about their use and performance in 2006. An increase in the range and number of (Q)SAR models that are available and validated, and guidance on how such models can be used, will be of considerable assistance to OECD member countries as part of the overall approaches needed for testing and assessment.

16. The Expert Group reviewed a thought-starter paper for making (Q)SAR models available in a system to assist regulatory use and decision-making, provide guidance to assist in the selection of the best (Q)SAR models, and provide information on their validation status. The issue of an Application Tool Box will be an item for discussion by the Ad hoc Group on (Q)SARs, as the scope and purpose of this Tool Box have not yet been agreed in to by the OECD. Nevertheless, there are a number of elements of an Application Tool Box that Member countries could consider.

17. The review of approaches for making (Q)SAR models readily assessable to member countries suggests that a central library of (Q)SAR models could be created, along with the current status of the validation for each (Q)SAR model with respect to the principles. One approach for providing this information to stakeholders is the use of web-based technology. The (Q)SAR Expert Group discussed this technology and suggested that emphasis be placed most heavily on the delivery of the (Q)SAR tools, acknowledging that different levels of support are required for different types of decisions and such prioritizing should be done by the regulating authority and not by the toolbox.

USE OF (Q)SAR MODELS IN REGULATORY FRAMEWORKS:

18. The scientific foundation of (Q)SAR models lies in physical organic chemistry, where aspects of chemical structure are used to estimate chemical behaviour and activity directly from chemical structure. The use of (Q)SAR methods to predict the results of test methods has been a part of chemical engineering for more than 50 years, and over the last 25 years has emerged as a useful tool in the regulatory assessments of chemicals in some OECD Member countries (Auer et al., 1990; Auer et al., 1994). The use of (Q)SARs for regulatory purposes among OECD Member countries are, generally, more widespread in those programs where regulatory constraints and assessment schemes limit the amount of data available from experimental test methods. For example, due to constraints on testing new industrial chemicals under the Toxic Substances Control Act (TSCA), the United States Environmental Protection Agency has developed one of the more extensive programs using (Q)SAR models to estimate important data for untested industrial chemicals (Clements et al., 1988; Lipnick et al., 1985). These models are now also being used to support assessment of pesticide inert ingredients in the United States (U.S. EPA, 2002g). Regulatory programs which require comprehensive testing, and therefore have more consistent sources of test data, have generally not had to rely on or investigate the use of alternative assessment methods such as (Q)SAR models. However, an increasing need to enhance efficiencies in these risk assessment processes suggests potential applications in the future (Bradbury et al., 2004; Matthews et al., 2004).

19. In the assessment of existing chemicals, the paucity of test data has been determined to be a global problem. And thus, as noted above, OECD Member countries have been cooperating in the evaluation of such non-testing (Q)SAR models for OECD's High Production Volume (HPV) Program chemicals and other chemicals for more than a decade (OECD, 1992a&b, 1993a&b, 1994, 1995, 2000, 2001, 2004a-e). And recently, the REACH legislative initiative in the EU provides for a possible increased reliance on (Q)SAR models for the assessment of existing chemicals (Pedersen et al., 2003; Bradbury et al., 2004). Member countries have responded by expanding the OECD activities to facilitate the development of guidance for the validation of (Q)SAR models as well as to assist in making the best available (Q)SAR models readily accessible to OECD Member countries.

20. A central issue in the use of alternative assessment methods in the absence of adequate test data is being able to enhance and promote the acceptability of (Q)SAR models for regulatory purposes. Through OECD principles for the validation, for regulatory purposes, of (Q)SAR models, it has been agreed upon that a (Q)SAR should have : 1) a defined endpoint; 2) an unambiguous algorithm; 3) a defined domain of applicability; 4) appropriate measures of goodness-of-fit, robustness and predictivity; and 5) a mechanistic interpretation, if possible [See Annex]. Validation of (Q)SAR models needs to take into account the fact that the reliable predictions can only be expected within the applicability domain that is defined by the content of the training set and the understanding of underlying mechanisms of the measurements. In this regard, the design of validation studies for (Q)SAR models may be different from those of more global laboratory test methods in that a (Q)SAR model may be limited to only a predictable subset of chemicals for which the (Q)SAR model can make a suitable estimate.

21. Additionally, it must be noted that the variability of different laboratories in reproducing measurements (ring testing) using the same test guideline creates a boundary condition for the internal performance and predictivity of (Q)SAR models. If the variation in measurement were large, it would be unrealistic to expect (Q)SAR models based on those measurements to have higher predictivity. Acceptance of a (Q)SAR model as an alternative assessment method may be enhanced if the validation

studies place the performance results in the context of the performance of the test methods from which the model was derived.

22. In general, (Q)SARs are reductionist models that are inevitably associated with limitations in terms of the types of chemicals structures for which they may be appropriate. A major focus of the validation of (Q)SAR models has been to understand the domain, or classes of chemicals for which the models meet appropriate statistical requirements for validity. It is, therefore, important that the regulatory bodies first clearly identify their regulatory domains of concern (chemical list for which model predictions and decisions are needed) before deciding which particular model sets will be appropriate to assist in chemical management activities.

23. OECD Member countries often operate under different regulatory constraints and may have different regulatory domains (chemical inventories) for the same assessment processes. For example, the EU's inventory of existing chemicals is not the same as the U.S. inventory of existing chemicals. Even within the regulatory bodies of a Member country, regulatory domains can vary. For example, the U.S. EPA Office of Pesticides Programs makes regulatory decisions for chemicals that are designed to have biological activity (fungicides, biocides, etc.), whereas the U.S. EPA Office of Pollution Prevention and Toxics evaluates and regulates industrial chemicals, often exhibit different molecular properties/structures (polymers, surfactants, etc.). This leads to differences in the regulatory domains of concern and, in turn, to different (Q)SAR models that are appropriate for each authority. Without first defining the specific regulatory domain for which decisions need to be made, one will not be able to identify or develop appropriate (Q)SARs to assist in regulatory decision making. If the (Q)SAR applicability domain lies within the one regulatory domain and not another, it is reasonable to find regulatory acceptance under one authority and not under a different authority. Where possible, the (Q)SAR model evolution is made efficient when chemicals are strategically chosen to create a high-quality (Q)SAR training set within the regulatory domain of concern

24. At the center of selecting reliable (Q)SAR models is the definition of the domain of applicability (Netzeva et al., 2005; Tunkel et al., 2005). There are numerous methods that define the application domain using such parameters as structural fragments, descriptor ranges, and confidence interval parameters in the respective assessments. However, different methods may lead to different conclusions regarding the domain of applicability for a model, and the use of specific models for specific regulatory applications is still a matter of experience.

25. The regulatory acceptance of (Q)SAR models has generally developed separately in OECD Member countries, using individual (Q)SAR models to meet specific regulatory needs rather than developing from a harmonised international process. This is demonstrated in case studies I-VIII provided. Since most regulatory frameworks have different assessment needs and constraints, it may be very difficult to rigorously define universal (Q)SAR principles for (Q)SAR validation that will appropriately fit the demands of each regulatory framework. In an effort to demonstrate how different regulatory bodies currently use (Q)SAR to assist in decision making processes, and to better understand how these countries address issues outlined in the OECD principles for the validation of (Q)SAR models, the attached case studies have been provided by a number of OECD Member countries to describe the current status of (Q)SAR models in chemicals regulation at the national and/or regional level.

CURRENT REGULATORY APPLICATIONS OF (Q)SAR WITHIN OECD MEMBER COUNTRIES

Case Study I: Canada

New Substances

26. Under the New Substances Provisions of the Canadian Environmental Protection Act (1999), all new substances must undergo an assessment to determine if they are "toxic". The New Substances Branch of Environment Canada is responsible for conducting the environmental risk assessment on all new industrial chemicals and polymers, while Health Canada conducts an evaluation of the risk to human health. The notification scheme is tiered, with increasing information requirements as the volume of import or manufacture increases.

27. The risk assessments are conducted through an examination of effects and an assessment of environmental fate and exposure. The program requires the submission of physical/chemical and effects data, however, when adequately validated by the notifier and acceptable to the government evaluators, data generated using (Q)SARs is accepted as an alternative to test data and can be used as a basis for regulatory decision making.

28. The New Substances Program recognizes that some (Q)SARs are suitable for certain chemical classes, but fail to provide reliable information for others. This is often due to a substance falling outside the applicability domain of a model, with features of the molecule not represented in the training set. In addition, models are not presently available to address some endpoints needed to assess the impact of new substances e.g. benthic toxicity, chronic aquatic toxicity, mutagenicity. Consequently, modeled data are only used when it is believed to be a valid representation of a substance's physical/chemical or toxicological characteristics.

29. The New Substances Program assesses a wide variety of substances, many which are considered "model difficult". Examples of these classes include polymers, UVCB's, organometalics, inorganics, surfactants, ionizable substances and fluorinated substances. While statistics are not currently available on new substance notifications, it is noted that inorganic, UVCB and polymer substances, considered "model difficult", represent 48% of the Domestic Substances List (DSL), the Canadian inventory of existing substances. It is believed that a significant portion of substances notified to the New Substances program are not amenable to extensive modeling.

30. In practice, experimental, closest analogue and (Q)SAR data are all used to prepare a weight of evidence approach in arriving at an assessment conclusion. In circumstances where modeled data are inappropriate, these data will not be considered in the assessment.

31. While notifiers may in some cases use (Q)SARs to address mandatory information requirements, regulators also employ models for the following purposes:

- bridging of information gaps by estimating data elements that are not required under the New Substances Notification Regulations, but which may be helpful in completing the assessment

- assisting in the evaluation of requests to waive prescribed data elements

- cross-verification of data and estimates as part of the process to validate the submitted information and aiding in establishing the physical/chemical/toxicity profile of the notified substance

32. Under the New Substances Notification Regulations, it is incumbent upon the notifier to address information requirements enabling the program to conduct an assessment of the risk to the environment and human health. Modelled data can play a role in the assessments, as data submitted by the notifier to address information requirements, and as supporting information generated by evaluators. However, due to wide variations in the classes and complexities of substances assessed, the New Substances Program advocates the judicious use of modelled data, predicting properties of well understood classes of chemicals using robust models with strong training sets.

Existing Substances

33. The focus of the Existing Chemicals program in Canada has shifted from assessment of a limited number of substances, for which considerable data are available, to a systematic review of all Existing Substances. This necessitates increased reliance on (Q)SAR modeling, in at least the early stages, to set priorities for additional testing or for further assessment.

34. Each of the 23,000 + substances on the Domestic Substances List (DSL), an inventory of chemicals and biological agents that were in commerce in Canada between January 1984 and December 1986, must be "categorized" by September, 2006. The purpose of "categorization" is to determine which substances on the DSL may have the "greatest potential for exposure" to the general population or are persistent (P) or bioaccumulative (B) and "inherently toxic" (iT) to human beings or to non-human organisms. Environment Canada considers persistence, bioaccumulation and "inherent toxicity" to environmental organisms while Health Canada considers "greatest potential for exposure" and "inherent toxicity" to humans. Due to relative availability of relevant data and the importance of water as a receiving medium for environmental chemicals, categorization of "inherent toxicity" to environmental organisms is focussed primarily on the aquatic compartment. Categorization, as well as additional priority setting, will identify substances for screening assessment and, full assessment where warranted, in an iterative approach of identifying candidates for risk management.

35. The types of substances on the DSL include discrete organic chemicals (which make up half of the DSL substances), polymers, inorganic substances, products of biotechnology, and substances that are of "Unknown or Variable Composition, complex reaction products, or Biological materials" (referred to as UVCBs). Generally, only the discrete organic substances are amenable to (Q)SAR modelling. Even within this class, there are certain types of substances for which some (Q)SARs are unsuitable.

36. However, given the paucity of experimental data available for the large number of substances on the DSL, Environment Canada uses (Q)SAR models and read-across data to help provide information for physical chemical properties, fate and toxicity endpoints within the categorization and priority setting exercises, where experimental data is lacking. "Rules of thumb" in the application and interpretation of the output of multiple and diverse (Q)SAR models to predict persistence, bioaccumulation and acute aquatic toxicity in categorization have been developed by Environment Canada based, in part, on the results of a workshop of global experts held in November, 1999. A guidance document outlining the models used in categorization by Environment Canada as well as all the (Q)SAR predictions for P, B and iT (aquatic organisms) for most of the organic (or organic components of) DSL substances are available by request from Environment Canada (DSL.Surveyco@ec.gc.ca).

37. (Q)SAR, read-across and experimental data are used for both priority setting as well as for risk assessment. Results from (Q)SARs and analogue data and are viewed as supplementary to experimental data. Experimental data are gathered from the literature, in-house/external databases or supplied in new substance notification packages. Environment Canada, in particular, uses (Q)SAR and analogue data to form a weight of evidence for a given assessment endpoint.

38. In order to address the technically demanding and precedent setting DSL mandate, Health Canada has developed a series of simple and complex tools for both exposure and hazard to efficiently identify substances that represent the highest priorities for additional consideration in screening assessments from a human health perspective. One of the tools developed is the complex hazard tool (ComHaz) which involves a comparison of information on a hierarchical series of toxicological endpoints relevant to human health with specific criteria that can be qualitative (e.g., carcinogenicity/genotoxicity) or quantitative (e.g., repeated dose toxicity).

39. Within each endpoint in the ComHaz tool, substance specific sources of information are considered in a hierarchical manner including empirical data, predictions from (Q)SAR) models and comparisons with analogues. Decisions are conservative in the absence of data, and for endpoints where data are limited and confidence in the outcome of (Q)SAR models is low, substances are retained for additional consideration in screening. (Q)SAR models applicable to ComHaz include a variety of models included in the TOPKAT (Accelrys Inc.), CASETOX (Multicase Inc.) and Derek for Windows (LHASA Ltd.) programs covering a variety of endpoints (e.g., carcinogenicity, genotoxicity, acute toxicity, etc.).

40. The first stage of ComHaz is based on a very conservative “first hit” approach such that if any of the available data satisfy endpoint specific criteria then the substance is prioritized for further assessment. This “first hit” approach has resulted in a significant proportion of those substances considered being captured on two endpoints with qualitative criteria, carcinogenicity and genotoxicity. Therefore, Health Canada is developing a preliminary weight of evidence framework to additionally focus on true priorities for these endpoints. Also, carcinogenicity and genotoxicity are endpoints for which (Q)SAR is likely to contribute to the greatest extent to weight of evidence determinations based on higher confidence in relevant predictive methods due to the generally larger, more diverse training sets available (e.g., Ames test models), the potential for combining information from models for related endpoints and the relevance of some of the endpoints of these models to specific modes of action (e.g., clastogenicity versus mutagenicity).

41. The preliminary weight of evidence framework for carcinogenicity and genotoxicity involves an iterative consideration of three lines of evidence for each endpoint starting with empirical data, followed by (Q)SAR predictions and analogue comparisons. The need to consider each successive line of evidence depends on the overall result (i.e., positive or negative) and the degree of confidence in the available information. Confidence in each line of evidence is primarily based on a comparison of the available positive and negative results. Because it is a preliminary framework used in the context of prioritization, individual studies are not critically evaluated at this stage; rather defensibility relates more to consistency across each line of evidence and between lines of evidence. The framework includes a weighting system to take into account the multiple study types and predictive models available for carcinogenicity and genotoxicity. The weighting for individual studies and (Q)SAR model predictions is largely based on the predictive power of the underlying assays. Additional weighting of (Q)SAR predictions based on available information on the domains of applicability and performance (e.g., sensitivity, specificity, etc.) in relation to the types of substances on the DSL is also being considered. Approaches to the identification of appropriate structural analogues to enable the extrapolation of toxicological data to data poor and model difficult substances are also under development.

42. Depending on the outcome for both carcinogenicity and genotoxicity, substances are prioritized for further assessment, or considered against the remaining endpoints in the first stage of ComHaz to (e.g., regulatory/reference values, developmental toxicity, etc.)

43. As a basis to transparently document output of the (Q)SAR models, Health Canada has developed "robust study summary templates". These templates include the identity of the substance tested and any available information on the individual model such as version number, size of the database, distribution of positive and negative compounds and criteria for inclusion of compounds in database. The templates also include details on the prediction (e.g., probabilities, results of coverage or other validation analyses, etc.), the criteria used to interpret predictions and information on similar compounds in the model database where available.

44. More information on the assessment approaches used by Environment Canada and Health Canada can be found at:

Health Canada Categorisation and Screening Health Risk Assessment:

http://www.hc-sc.gc.ca/ewh-semt/contaminants/existsub/index_e.html [accessed 1 June 2006]

Environment Canada Categorisation and Screening Risk Assessment:

http://www.ec.gc.ca/substances/ese [accessed 1 June 2006]

Health Canada New Substances:

http://www.hc-sc.gc.ca/ewh-semt/contaminants/chem-chim/index_e.html
[accessed 8 August 2006]

Environment Canada New Substances:

http://www.ec.gc.ca/substances/nsb/eng/index_e.htm [accessed 1 June 2006]

45. Health and Environment Canada strongly and actively support development and refinement of (Q)SAR models, through continued international discussions among experts.. In view of the critical importance of (Q)SAR to Canadian programs, an easily accessible global repository of toxicity testing results, physico-chemical data and information on molecular structure from the private and public sectors is desirable. Compilation or linking of such data in a central repository would reduce the need for animal experiments and associated costs by contributing to model development. It would also avoid duplication of effort and promote better-informed, more discriminating animal testing. It is envisaged that OECD could play a critically important coordinating role in this context.

Case Study II: Czech Republic

46. Commercial expert system HAZARDEXPERT SYSTEM is using in the National Institute of Public Health (governmental institute of the Ministry of Health) as a subsidiary material for Material Safety Data Sheets (MSDS) of chemical substances and/or preparations or in the process of notification of new chemicals, if no toxic data are available. The system is also used to judge whether a choice of a set of mandatory and additional tests of toxicity for risk assessment is sufficient, superfluous or deficient and/or in consequence to biotransformation and toxicokinetic studies as far as no knowledge of toxicity and kinetics of metabolites exist. The QSAR-based system reveals a possibility of a hazardous property in the molecule of chemicals and its result may to direct an attention to necessity to provide additional tests of toxicity. The legislative of Czech Republic makes it possible to use alternative methods recommended by OECD as far as no other suitable methods exist. The results of *in silico* methods QSAR and SAR are, thus, possible to be used in these cases in accordance with the Act no.356/2003 and Decrees no. 443/2004, no. 426/2004 and no. 434/2005 Collection of Laws of Czech Republic based on OECD Council Directives 67/548/EC, 98/8/EC and associated Technical Notes for Guidance.

47. Results of the expert system are used by researchers of the National Institute of Public Health in peer reviews for WHO if a conclusion or a speculation in the review is debatable.

48. Laboratories and organizations accredited for toxicity testing of chemicals purchase or consider to purchase commercial QSAR technology to make its inner validation to be prepared to fulfil demands of costumers when occur.

49. QSAR analysis and methodology is used in research by academic and university teams in solving molecular biological or chemical tasks (mechanisms of toxicity of contaminants *in vitro* or structure – activity relationships of enzymic catalysis at Masaryk University in Brno, a system to estimate a hazard of mixtures of chemicals or development of QSAR model for acute toxicity at Charles University and the National Institute of Public Health in Praha).

50. Information on QSAR and SAR technologies and their usage in estimating toxic properties by calculation is included in lectures on “Toxicology” of some universities. There is a separate course dealing especially with QSAR technology and its usage for regulatory purposes organized by Czech chemical and medical societies. Contributions involving a usage of QSAR or SAR analysis are a part of some national conferences and meetings. All mentioned activities serve to educate the professional community in possibilities of alternative methods *in silico* in testing toxicity of chemicals with the aim to extend applications of QSAR and SAR technology especially for legislative and laboratory usage. Lectures on QSAR technologies, usage and advantages are included in expert chemical symposia in context of REACH.

Case Study III: Denmark

51. The Danish Environmental Protection Agency (EPA) has evaluated (Q)SAR models for important test endpoints needed for both human health and ecological hazard assessments. The Danish EPA (2001) developed an advisory list for self-classification of dangerous substances using (Q)SAR models. Of approximately 47,000 substances which were examined, 20,624 substances were identified as requiring hazard classification for one or more of the following dangerous properties: acute oral toxicity, sensitisation by skin contact, mutagenicity, carcinogenicity and danger to the aquatic environment. The Danish EPA concluded that the (Q)SAR models used here are now so reliable that they are able to predict whether a given substance has one or more of the properties selected with an accuracy of approximately 70-85. For further information on specific endpoints see www.mst.dk/chemi/01050000.htm [accessed 1 June 2006]

52. The Danish EPA also demonstrated for the first time that automated approaches for applying (Q)SARs systematically to large chemical inventories is feasible, and that the resulting databases of non-test data offer important opportunities to review the hazards of chemicals produced in lesser amounts. The Danish EPA database is comprised of the estimations from more than 70 (Q)SAR models on endpoints for physico-chemical properties, fate, eco-toxicity, absorption, metabolism and toxicity. The database uses a chemical inventory of 166,000 chemicals and is constantly growing as new models are obtained, developed or refined. More than half of all the estimates are for human-health related endpoints found in the SIDS dossier.

53. Like many other Member countries, the (Q)SAR database and other (Q)SAR tools are being used by the Danish EPA in daily work to fill in data gaps when there are no experimental test results or where (Q)SAR gives information beyond the experimental test data. In connection with international work, information from the (Q)SAR database on environmental- and health effects has been used to give input on substances undergoing evaluation in the OECD High Productions Volume Chemicals program (ENV/JM/TG (2004)26/REV1), and where relevant also on substances undergoing risk assessment in the EU.

54. The Danish EPA has shared (Q)SAR predictions on discrete organic chemicals on the agenda since SIAM 11 in 2001, and has reviewed the (Q)SAR approach at the Special Session on the use and regulatory acceptance of (Q)SARs at the OECD Joint Meeting of the Chemicals Programme in November 2002. To provide Member countries with a sense of the state-of-the-science for important endpoints, the Danish EPA conducted an evaluation from a regulatory perspective of a family of (Q)SAR models used to estimate biodegradability, acute aquatic toxicity and mutagenicity for the chemicals included in the OECD HPV programme in SIAM 11-18 for selected endpoints.

55. The purpose of that evaluation was to examine the comparison between SIDS assessments and Danish EPA (Q)SAR predictions and discuss the outcome with regulators, industry and other NGOs. This case study with 184 discrete organic chemicals identified practical limitations, possibilities and advantages of using (Q)SAR models as alternative and/or supplementary information sources on the intrinsic properties of chemicals.

56. The comparison between biodegradability (Q)SAR predictions and SIDS data on the selected chemicals seems to be in accordance with results found in previous and comprehensive validations of the models; namely, that these models have a very high specificity and a moderate to high sensitivity for identifying chemicals which are not ready-biodegradable.

57. The evaluation of SIDS data included aquatic toxicity measures for the LC_{50} for fish (96 hrs), EC_{50} for Daphnia (48 hrs) and EC_{50} for algae (normally 72 hrs). Even though the SIDS data include only internationally validated test data, the data within the same taxa differ significantly and limited the evaluation of (Q)SAR model predictions. Nonetheless, (Q)SAR models for fish estimated acute toxicities for which 80%, 93% and 100% of the SIDS chemicals fell within a factor of 10,100 and 1000, respectively of the measured values.

58. The Danish EPA models include 13 *in vitro* mutagenicity models, 6 *in vivo* mutagenicity models and 19 carcinogenicity models. For the SIDS evaluation, the predictions for the Ames test (Q)SAR models and the accepted SIDS data seemed to indicate a very high specificity and also a high sensitivity. The specificity and sensitivity for the Ames test mutagenicity models were for this SIDS data set 95 % and 60-80%, respectively. However the sensitivity is uncertain because only few positive (mutagenic) test chemicals were included in the SIDS test data set.

Case Study IV: Netherlands

59. The Netherlands has taken the initiative of a so-called EU (Q)SAR Experience Project aimed at exchange of experience/knowledge in the use of QSAR predictions and to provide input in the (EU)-regulatory needs in decision making. This work is described in the section EU (Q)SAR Experience Project (see para 196).

60. In recognizing the importance of alternative methods in the evaluation of substances, The Netherlands compares the outcomes of several (Q)SAR models with the experimental data of New Chemicals (and occasionally also for the OECD/SIAM chemicals), mainly for experience and capacity building for most endpoints that needs to be assessed. The models EPISUITE, TSCA List for categories and DEREKfW are used. The Netherlands has presented a poster at the SETAC-Lille symposium 2005 on their experiences for irritation, sensitisation and fish toxicity.

61. Also for experience and capacity building the DEREKfW models have been screened for irritation, sensitisation, genotoxicity and carcinogenicity using 50 IPCS chemicals and circa 50 New Chemicals (Hulzebos and Posthumus, 1993). The predictions of BIOWIN models have been evaluated with the "Ready Biodegradability test' outcomes for 70 new chemicals (Posthumus et al., 2005). For reproductive toxicity DEREKfW models and the TSCA List on chemical categories were screened with chemicals classified in the EU for reproductive toxicity (Maslankiewicz et al., 2005). Also for the reproductive toxic group of phthalates a category was defined (Fabjan, 2005 submitted).

62. In the development of (Q)SARs work has been done in cooperation with John Walker (TSCA of US-EPA) and Ingrid Gerner BfR and co-workers, which is also mentioned in case study VI: Germany. A testing strategy for skin irritation, including mechanism and structural alerts has been proposed that potentially needs no in-vivo testing (Walker et al., 2004, 2005, Gerner et al., 2004b and Hulzebos et al., 2003 and 2005, Rorije and Hulzebos, 2005).

63. Earlier work has been done on ecotoxicity and biodegradation. Many of the applications of (Q)SARs for ecotoxicity in The Netherlands were reported by van Leeuwen and Hermens (1995). An approach was developed to classify environmental pollutants according to mode (or mechanism) of actions and subsequently to make predictions of toxicity (Verhaar et al. 1992; 1996; 2000). The approach allows for the calculation of a 'baseline' toxicity and then a classification scheme for four classes of toxicity: Class 1 (inert chemicals, baseline toxicity, typically non-polar narcosis); Class 2 (relatively inert chemicals, typically polar narcosis); Class 3 (reactive chemicals, typically compounds capable of covalent electrophilic and nucleophilic interactions with biological macromolecules); Class 4 (specifically acting chemicals such as pesticides). The classification approach was applied to predict the toxicity of HPVCs (Bol et al. 1993; Verhaar et al. 1994). For inert toxicants (i.e. non-polar narcotics) (Q)SAR were published for the no-observable effect concentration values for 19 different species of bacteria, algae, fungi, protozoa, fish, amphibians and other invertebrates (van Leeuwen et al. 1992). For biodegradation also validation studies were carried out and a new model developed (Langenberg et al., 1996; Rorije et al., 1999; Loonen et al., 1999).

Case Study V: United States

U.S. Environmental Protection Agency

Regulatory Use of (Q)SAR by the U.S. EPA's Office of Pollution Prevention and Toxics (EPA/OPPT)

[Based upon OECD, 2004d]

64. In order to assess the risk of a new chemical, the U.S. EPA makes predictions concerning chemical identity, physical/chemical properties, environmental transport and partitioning, environmental fate, environmental toxicity, engineering releases to the environment, and environmental concentrations. The Agency uses a variety of methods to make predictions which include SARs, nearest analogue analysis, chemical class analogy, mechanisms of toxicity, chemical industry survey data, and professional judgement. The Agency uses these methods to fill data gaps in an assessment and to validate submitted data in notifications (Nabholz, 2001). Predictions are made by the U.S. EPA's Office of Pollution Prevention and Toxics (OPPT) under TSCA (Zeeman 1995; Zeeman et al., 1995). The OPPT has routinely used (Q)SARs to predict ecological hazards, fate and risks of new industrial chemicals, as well as identifying new chemical testing needs for more than two decades.

65. As a result of this experience, the U.S. EPA offered to lead the OECD effort in development of guidance for regulatory applications of (Q)SARs. The U.S. proposed an approach that emphasized the use of program-specific case studies to highlight the importance of legal and practical constraints and information requirements of regulatory programs in applying (Q)SAR approaches. The U.S. EPA has presented a case study based on the experience of the U.S. EPA in the application of (Q)SARs to the assessment and management of chemicals. The study design is based upon the lengthy and extensive use of (Q)SAR within the U.S. EPA's specific regulatory framework for the assessment of new and existing industrial chemicals, available case-studies on the use of such U.S. EPA (Q)SAR methods (U.S. EPA, 1994a; OECD, 1994; Zeeman et al, 1999) and the application of the OECD principles for (Q)SAR validation to U.S. EPA methods as proposed by the OECD Expert workgroup.

66. A recent study by U.S. EPA of some commercial (Q)SAR methods revealed difficulties associated with several of the Setubal and the OECD's principles for (Q)SAR validation which should be addressed in future OECD guidance for (Q)SAR applications. An evaluation of these principles led to the conclusion that explanatory notes attached to the OECD's principles [See Annex] for the interpretation and application of Principle 3 (a defined domain of application) and Principle 4 (appropriate measures of goodness-of-fit, robustness and predictivity) are not sufficiently developed and the results reveal potential conflicts under particular regulatory constraints. It is possible that a similar assessment conducted by other Member countries under different regulatory frameworks may (or may not) lead to analogous results.

67. The U.S. EPA's methodology for hazard and risk assessment of new industrial chemicals, which integrates (Q)SAR models into the hazard and exposure analysis, has been used for over 20 years and reflects several specific regulatory requirements that define the framework under which the U.S. EPA New Chemicals Program must operate. The assessment of new chemicals and the retrospective assessment of an inventory of existing chemicals are within the purview of U.S. EPA's Office of Pollution Prevention and Toxics (OPPT), formerly the Office of Toxic Substances. The OPPT administers the Toxic Substances Control Act (TSCA) which was passed in 1976 to regulate all industrial chemicals in U.S. commerce. Under TSCA, U.S. EPA is charged with assessing, and if necessary, regulating all phases of the life cycle of industrial chemicals including manufacturing, processing, use, and disposal.

68. In 1979, almost 62,000 industrial chemical substances were reported to be in commerce in the U.S., and these were "grandfathered" as the TSCA inventory of existing industrial chemicals. Chemicals not included on this inventory before 1979 were considered "new" industrial chemicals and these substances had to be submitted to U.S. EPA for review within the OPPT New Chemicals Program prior to commencing commercial manufacture or import activities (Zeeman et al. 1995, 1999). More than 42,000 of these new chemicals have been submitted to OPPT by industry for assessment since July 1979. Via the inclusion of about 18,000 new industrial chemicals that have been assessed for risk and are now in commerce, the TSCA inventory has increased to more than 80,000 chemical substances.

69. TSCA requires OPPT to regulate two categories of industrial chemicals: (1) existing chemicals in commerce on the TSCA Chemical Substances Inventory, and (2) new chemicals. Section 5 of TSCA requires manufacturers and importers of new chemicals to submit a premanufacture notice (PMN) to U.S. EPA/OPPT 90 days before they intend to begin manufacturing or importing a new chemical. U.S. EPA/OPPT must evaluate whether the substance may present an unreasonable risk of injury to human health or the environment. OPPT must make a risk-based decision on the regulatory outcome of the chemical within these 90 days. The PMN is, otherwise, automatically approved. In addition to this demanding 90-day review period, two other major programmatic constraints are associated with evaluating PMN chemicals under the current TSCA framework.

70. The first constraint is that of the large number of PMN chemicals submitted each year (40-50 per week or up to 2000 per year), approximately 65% of the substances are being submitted with no test data. Under TSCA, the notifier is not required to conduct any ecological or human health testing before submitting a PMN. In the U.S., in the absence of a risk-based finding to support regulatory action after the specified 90-day review, the chemical is automatically cleared for commercialization with no legal requirement of more information unless the U.S. EPA takes regulatory action (Auer et al., 1994). Because of this constraint, U.S. EPA must assess each new chemical submitted, regardless of the level of understanding concerning the specific chemical or chemical class.

71. The second constraint is that TSCA places the burden of proof on the U.S. EPA to determine whether the manufacture of a new chemical "may present" an unreasonable risk to human health or the environment. EPA cannot require the notifier to submit additional information about the new chemical unless there is adequate data to support an unreasonable risk finding. Because of this statutory requirement, and the demonstrated lack of experimental data submitted with the PMNs, the U.S. EPA was faced with the need to predict over 150 attributes for a large number of chemicals in a very short period of time in order to make rapid decisions regarding the risk associated with manufacturing a PMN chemical.

72. From these constraints, it was obvious that the methods of risk assessment utilized by U.S. EPA in the New Chemicals Program had to be scientifically sound and very pragmatic. In response to this data-poor situation, U.S. EPA/OPPT developed predictive approaches (preferably referred to as "estimation methods" within OPPT) which are used to fill data gaps where little or no experimental data exists. These approaches include (Q)SARs, nearest analog analysis, chemical class analogy, mechanisms of toxicity, and professional judgement. In order to quickly complete an assessment for each new chemical, the Agency uses modeling to make predictions concerning physical/chemical properties, environmental transport and partitioning, environmental fate, environmental toxicity, engineering releases to the environment, and environmental concentrations to fill data gaps left by the notifier in the PMN (U.S. EPA, 2003). Those predictions are used to support the U.S. EPA/OPPT chemical management decisions within the TSCA framework and to assist the Agency in determining the most appropriate regulatory decisions for each new chemical based on the potential risks. For example, see the OPPT PMN case study presented below (Zeeman et al., 1999). If a risk is predicted, more detailed analyses are performed and the notifier may be asked for additional information. If the risk can be managed by the notifier, then the new chemical may be

regulated to ensure that the potential risk remains low. If the risk cannot be mitigated, then the new chemical may be subject to further regulatory controls (U.S. EPA, 2003).

73. For the purposes of chemical risk assessment under the New Chemicals Program, U.S. EPA/OPPT has had to develop its assessment methods within the programmatic constraints of TSCA. TSCA forces the U.S. EPA/OPPT to effectively deal with the very real and difficult problem of assessing hazard, fate, and risk for hundreds to thousands of chemicals in the absence of sufficient data on the chemical and within a short time period. The predictive models employed in the U.S. EPA/OPPT assessment process have been the subject of extensive review and some of the past case study material is presented below.

74. An OECD workshop was held in 1989 on notification schemes for new chemicals. One of its recommendations focused upon the need to evaluate the predictive power of the (Q)SARs used by the U.S. EPA/OPPT by comparing the results of the OPPT's (Q)SAR assessment with those obtained from the base-set testing of new chemicals (physical/chemical, environmental fate, toxicological properties/health effects, and ecotoxicity parameters) required by the European Union (EU, then the European Commission/EC).

75. The relative accuracy of these two diverse systems became the subject of a joint U.S. EPA/EC evaluation called the "Structure Activity Relationship/Minimum Premarketing Dataset" (SAR/MPD) study. This comparison was undertaken from 1991-1993 and it contrasted the OPPT (Q)SAR predictions with the new chemical base-set test data for 175 chemicals that had been submitted to, and selected by, the EC. The results of the SAR/MPD study were subsequently jointly published in 1994 by the U.S. EPA and the EC (OECD, 1994), and some of the conclusions are presented below.

76. Overall, some (Q)SARs for physical/chemical properties appeared to not be as reliable as had been hoped (e.g., water solubility and vapor pressure) and some that were, within specific limits, reasonably accurate [i.e, (K)ow]. The OPPT estimation methods for biodegradation were likely to identify slowly degrading chemicals, but were less likely to identify those which readily degrade.

77. For human health hazards, qualitative SAR was generally found useful and adequate for some endpoints (i.e., acute lethality and mutagenicity) and not for others (i.e., eye/skin irritation and skin sensitization). For ecotoxicity, the EU base-set only required acute toxicity data for fish and daphnids. The EPA experts reported that "the agreement between the U.S. predicted values and the EC measured values is 87% for fish acute toxicity and 79% for daphnid acute toxicity." Although the EU experts estimated a slightly lower concordance, they nevertheless concluded that the U.S. EPA/OPPT (Q)SAR methods "performed extremely well in predicting acute toxicity to fish and daphnia." (U.S. EPA, 1994; OECD, 1994).

78. In 1994, the OPPT New Chemical Program finalized a Policy Paper responding to the conclusions of the SAR/MPD study. The Policy Paper, in response to the findings of high overall agreement on aquatic toxicity results states that ecotoxicity is "probably the area in the New Chemicals Program where (Q)SAR predictive capabilities have reached their greatest refinement and improved most over time. Factors contributing to this success include: Careful development and analysis of chemical categories ... thoroughness and diligence in adding new data points to established categories, attempting to complete the toxicity profile and refine the analysis ... The experience in ecotoxicity predictions should serve as a model for other disciplines."

79. The U.S. EPA/OPPT PMN case study was originally prepared to illustrate the consistency between the OPPT ecological risk assessment approach and the U.S. EPA's Framework for Ecological Assessment. That OPPT/PMN case study was subsequently updated for inclusion in a U.S. government

publication and became a relatively comprehensive example of OPPT's capabilities in using (Q)SAR to conduct hazard (and ecological risk) assessment for specific new chemical substances (Zeeman et al., 1999).

80. The PMN submitter identified processing, use, and disposal sites adjacent to rivers and streams and it was expected that the PMN substance would be discharged into such environments. Therefore, the endpoints of concern were the mortality, growth and development, and reproduction in aquatic organisms such as fish, aquatic invertebrates, and algae.

81. Initial exposure concentrations were estimated using a simple dilution model that divided releases (kg/day) by stream flow (millions of liters/day). Subsequent exposure analyses used a probabilistic dilution model (PDM) and the exposure analysis modeling system (EXAMS). PDM3 was used to estimate the number of days a particular effect concentration would be exceeded in 1 year, and EXAMS II was used to estimate concentrations in the water column and in sediments using site-specific data.

82. This study focused on the assessment of a specific PMN substance, i.e., an alkylated diphenyl, that is a neutral organic compound. Chemicals belonging to this class of compounds elicit a nonspecific and simple form of toxicity known as narcosis. The toxicity of neutral organic compounds were estimated through (Q)SARs that correlate toxicity with the octanol-water partition coefficient (Kow) and molecular weight. The subject PMN substance had a predicted log Kow of 6.7. Compounds with such a high log Kow are not expected to be acutely toxic (i.e., no acute effects at saturation over short-term exposure durations), but are expected to elicit chronic effects following longer-term exposures. Actual testing of the PMN substance subsequently confirmed the OPPT predictions of aquatic toxicity.

83. Thus, aquatic toxicity was initially estimated by the use of (Q)SAR. Acute aquatic toxicity test data accompanying the PMN submission and later receipt of additional chronic aquatic toxicity test data confirmed the accuracy of the OPPT (Q)SAR predictions. Uncertainty factors were used to estimate a concern level or concern concentration (CC) in the receiving stream. This stream water-column CC was set at 1 μg/L (ppb). When the OPPT risk assessment determined that this CC was exceeded for more than 20 days, a potential unreasonable chronic risk was assumed to be expected if this PMN chemical substance was allowed to be used.

84. This PMN evaluation ultimately resulted in five iterations of hazard and exposure analysis and risk characterization. The first four iterations identified an ecological risk and resulted in the collection of additional and more specific ecological effects test data and more detailed information on potential exposures to the PMN substance. The accuracy of the OPPT's (Q)SAR assessments were essential in identifying potential hazards and risks, and in the development of chronic aquatic toxicity test data on the PMN chemical. The final outcome was that the PMN substance could only be used at the identified sites because there was uncertainty as to whether the concern level (1 μg/L) might be exceeded at sites not identified and characterized by the submitter.

85. An Examination of the Difficulties Associated with Applying Universal (Q)SAR Principles to Existing Regulatory Frameworks

86. Using the explanatory notes for OECD Principles 3 and 4 for (Q)SAR validation as guidance, examples from current research and 20 years of experience in (Q)SAR application within a regulatory framework, it is possible to illustrate why it may be difficult to overlay pre-defined universal (Q)SAR principles to all types of regulatory/decision making frameworks.

Domain of Applicability (OECD principle #3 for (Q)SAR validation)

87. U.S. EPA agrees that (Q)SARs are reductionist models that are inevitably associated with limitations in terms of the types of chemical structures, physiochemical properties, and mechanisms of action for which they may be appropriate. However, it has been the experience of U.S. EPA/OPPT and other researchers in the field of (Q)SAR validation that rigorously defining the domain of applicability for a (Q)SAR, as described in the interpretive guidance for the OECD principles for (Q)SAR validation, can be a difficult process that often does not prevent misuse. There are a number of mathematical methods that describe the domain of applicability for (Q)SAR models (Eriksson et al., 1997 and 2003; Netzeva et al., 2005; Tong et al., 2004; Klopman et al., 2003; Walker et al., 2003), which would satisfy the needs described by Principle 3. However, to-date, there are still relatively few models that implement a mathematical assessment of model domain in a way that does not dramatically decrease the number of chemicals that can be assessed by a model and/or that maintains transparency in scientific reasoning.

88. In general, the structural domain of a (Q)SAR model is assumed to be defined by the structural composition of the training set chemicals. Simply put, if the chemical of interest has structural features that are consistent with those features of the training set chemicals, then the chemical of interest is said to be well-defined by the model and subsequent predictions may be considered appropriate. However, studies indicate that defining a structural domain by analyzing the components of the training set and devising inclusion and/or exclusion rules for those corresponding functionalities, as described by the OECD interpretive guidance provided for Setubal Principle 4, is an imprecise activity. Often times it is difficult to determine an appropriate level of structural coverage and/or it is difficult to capture the modulatory effects of mechanistic considerations, the use of alternative regression methods, or variations in descriptor calculations. All of this is illustrated by the following examples.

89. Under its Chemical Substances Law, the Japanese government has tested approximately 900 discrete substances utilizing the Ministry of International Trade and Industry (MITI)-I test, a protocol for determining ready biodegradability that is one of the six officially approved ready biodegradation test guidelines under the OECD. This data set constitutes what is probably the largest available collection of measured biodegradability data obtained using a single defined test protocol. Over the past few years, this database, or subsets of it, have been used as the training set for several new predictive models (Tunkel et al., 2000, 2004, 2005; Rorije et al., 1999; Loonen et al., 1999; Gamberger et al., 1996), including BIOWIN 5, BIOWIN 6, TOPKAT, and MCASE. Each of these 4 (Q)SAR models is derived using the MITI data set, but each method employs different statistical techniques and descriptors in the derivation of the regression equations. The BIOWIN models were developed by the U.S. EPA for the prediction of ready biodegradation and are currently used in the U.S. EPA/OPPT's New Chemicals Program to aid in the determination of environmental fate. Recently, research was conducted (Tunkel et al., 2004, 2005) on these 4 ready biodegradation models (BIOWIN 5, BIOWIN 6, TOPKAT, and MCASE), and since these training sets were identical, the research provided a unique opportunity to address the difficult issues associated with defining a model's domain of applicability as set forth in the OECD principles for (Q)SAR validation.

90. Results from a recent study by Tunkel et al. (2005) indicate that the predictive accuracies determined for each of the ready biodegradation models vary widely. Additionally, for the two models that employ internal domain analysis, the number and types of chemicals for which a prediction could not be made (flagged as being outside of the domain) varied also. This suggests that not only do the structural features within the training set chemicals influence the parameters of domain, but the descriptors and algorithms used to derive the models may also significantly influence the parameters of the domain of applicability. There are many scientifically acceptable methods that use such things as structural fragments, descriptor ranges, and confidence intervals to make an assessment of the domain of applicability, however, application of each of these methods leads the assessor to different conclusions regarding the domain of

applicability. Thus, this could easily lead to disagreement with regard to which methods should be considered scientifically appropriate for a given substance.

91. For example, one approach for identifying the domain of applicability employs statistical methods that quantify variation in the model descriptors (mimicking structural variations) called the K-nearest neighbors (KNN) approach. For the K-nearest neighbors approach, the method adds a new data point (chemical of interest) to the Euclidean space which is defined by the range of the model descriptors (based on the training set) and finds its K-nearest neighbors in that space (Cash 1999). To demonstrate, if you are using log Kow, molar volume, HOMO-LUMO gap, and solvent-accessible surface area, you have a 4 dimensional Euclidean space with those axes. If all K of the nearest neighbors lie in the same direction as one or more descriptors (variables), that is a red flag that you may be outside the model's domain of applicability.

92. This approach creates bounds for the model that are strictly defined by the quantitative range of the model descriptors. It does not account for permutations in domain parameters created by the use of different regression methods, nor does it tolerate any potential external mitigating factors which may alter the definition of the domain of applicability. An example of such potential external considerations can be demonstrated using ECOSAR, U.S. EPA/OPPT's set of models to predict aquatic toxicity (see Table 1). These models were designed by the U.S. EPA/OPPT under the New Chemical Program for the purpose of predicting the aquatic toxicity of industrial premanufacture chemicals from chemical structure. ECOSAR can predict toxicity of chemicals to aquatic organisms such as fish, daphnids, and algae, often by using class specific (Q)SARs based on log Kow.

Table 1. Classes of Chemicals for which Toxicity Models exist in ECOSAR

Acid Chloride/Halide	Acrylamides	Acrylates
Aldehydes	Aliphatic Amines	Anilines (amino-meta)
Anilines (amino-ortho)	Anilines (amino-para)	Aromatic Amines
Aziridines	Benzotriazoles	Benzyl Alcohols
Benzyl Amines	Benzyl Halides	Diazoniums
Diepoxides	Diketones	Dinitro Aromatic Amine
Dinitrobenzenes	Epoxides	Esters
Esters (phosphate)	Haloacetamides	Hydrazines
Imides	Isocyanates	Malononitriles
Methacrylates	Neutral Organics	Peroxy Acids
Phenols	Phenols (dinitro)	Propargyl Alcohols
Propargyl Ethers	Quinone	Salicylates
Salicylic Acid	Schiff Bases	Silamines
Silanes (alkoxy)	Surfactants-anionic	Surfactants-cationic
Surfactants-nonionic	Thiazolidinones	Thiazolinone (iso-)
Thiocyanates	Thiols(mercaptans)	Thiophenes
Triazines	Ureas(substituted)	Vinyl/Allyl Alcohols
Vinyl/Allyl Ethers	Vinyl/Allyl Halides	Vinyl/Allyl Ketones
Vinyl/Allyl Sulfones		

93. In the development of the neutral organic (Q)SARs, generally chemicals with log Kow values in the range of -3 to 8 and molecular weights less than 1000 are used in the training sets. However, the domain of the SAR is in fact larger than the descriptor range of the training set of chemicals because any neutral organic chemical with log Kow greater than 8 (and belonging to no other SAR class) will be

predicted to have acute and chronic effective concentrations equivalent to "no effects at saturation" or NES (ECOSAR, 1996). Therefore, the domain of this SAR-based methodology can be larger than the domain(s) of the algorithms that predict toxicity as a continuous variable over the specified ranges of Kow.

94. Additionally, it is very difficult to retain the robust quality of a (Q)SAR model while implementing a rigorous domain evaluation as the domain assessments quite often limit the number of chemicals that can be predicted by the model (OECD, 2004b). Given the regulatory needs of U.S. EPA/OPPT under the New Chemicals Program to make an assessment of all chemicals submitted, the OPPT has not found a suitable method for defining a domain of applicability that does not dramatically sacrifice the robust nature of the model. Many of the (Q)SAR programs utilized within OPPT incorporate atom/fragment methods or chemical class specific SAR and, oftentimes, the compounds of interest will have some structural features that can be recognized by the (Q)SAR, even if all structural components of the molecule are not completely represented. When a molecule contains one or more fragments generally considered to contribute significantly to the endpoint of interest, this may be sufficient to permit a prediction for that endpoint.

95. Due to the programmatic need to make a decision for all chemicals submitted and because there is currently no consensus on one single approach for the evaluation of the domain of applicability, it is the practice of the U.S. EPA/OPPT to implement external domain evaluations on a case-by-case basis. These evaluations are completed using professional judgement by U.S. EPA scientists to identify chemicals structures for which the use of particular models may lead to estimations with unreasonable uncertainty. In these particular cases where the chemicals may appear to be outside of the domain, the potential uncertainty in predictions is not quantified by timely mathematical and statistical evaluations of domain, but rather, the potential uncertainty in the prediction is assessed qualitatively by staff and managers.

96. OECD Principle 3 certainly appears to possess strong scientific validity, however, the interpretations brought forth by the OECD expert group do not define what is a scientifically acceptable assessment method for domain evaluation, which leaves the principle rather subjective considering the number of available approaches. From OPPT's perspective, an appropriate assessment of the domain of applicability is best characterized by simultaneously considering the sum of a program's quantitative and qualitative methods, and currently, there are no agreed upon methods that can encompass both characteristics. Additionally, regulatory/decision making authorities must consider the external regulatory constraints they may need to perform under, such as the constraints faced by U.S. EPA/OPPT in the new chemicals review process. These elements support the need to maintain flexibility in the application of Principle 3.

Internal Performance and Predictivity (OECD principle #4 for (Q)SAR validation)

97. Ideally, a (Q)SAR model should be accompanied by full disclosure of the internal performance information for the training set chemicals including chemical names, structural formulae, raw data, data for descriptor variables, data quality, data processing methods, methods for selection of variables, and any statistical methods employed in the derivation of the (Q)SAR (OECD, 2004a&b). However, it is often difficult for OPPT to assemble and release all of the information regarding internal performance of a (Q)SAR in an effort to promote transparency of the model. Additionally, some of the information may actually be confidential or protected and therefore, restricted from being disseminated to a third party. In fact, this is the case in many instances for some of the (Q)SAR models used in the U.S. EPA/OPPT New Chemicals Program.

98. The U.S. EPA affords Confidential Business Information (CBI) protection to submitters under TSCA in an effort to help companies maintain their competitive advantage in the marketplace. Under this protection, manufacturers and importers can designate many characteristics of their material as CBI such as

chemical name and structure. Only personnel with TSCA CBI clearance and members of Congress can access the information, thereby prohibiting dissemination of the information to the public, and, therefore, to the PMN submitter's potential competitors.

99. Ideally, methods should be continually updated in an effort to fill data gaps and/or enhance predictions for particular classes of chemicals. In an effort to utilize all existing data, U.S. EPA/OPPT often uses validated CBI data submitted by Industry to perform these update/enhancement activities. This creates a situation in which, by law, full disclosure of the internal performance information for some models is strictly prohibited. Does that mean each of those models should be considered "inappropriate" simply because the developer is restricted by law from disseminating the information regarding internal performance?

100. Consider the outcomes of a study completed by the Dutch National Institute of Public Health and Environment (Hulzebos and Postumus, 2003) that demonstrated an OPPT (Q)SAR model (ECOSAR) could perform with acceptable predictivity even though the program was not accompanied by full disclosure of the internal performance parameters. The study was designed to determine if all (Q)SARs within ECOSAR conformed to the recommended acceptability criteria for (Q)SAR application within Dutch risk assessment. These Dutch criteria very closely parallel some of OECD's proposed guidance for the interpretation of Principle 4 for the evaluation of internal performance (OECD, 2004a&b). In the study, a comparison was made of these Dutch applicability criteria against 123 (Q)SARs within ECOSAR, representing 46 chemical classes. This activity resulted in what was defined as 96 "unreliable" (Q)SARs and 27 "reliable" (Q)SARs within ECOSAR. However, further results of the study indicate that upon validation of these "reliable" and "unreliable" (Q)SARs, both performed equally well. It was concluded that within ECOSAR, it was possible to obtain accurate and useful predictions using (Q)SARs that were originally considered to be "unreliable" when evaluated against predefined internal performance standards (Huzbelos and Posthumus, 2003). In this case, it seems unsuitable to make the applicability of a particular (Q)SARs totally contingent upon data availability.

101. From the results of this study it was demonstrated that applying a set of extensive applicability criteria by which "reliable" (Q)SARs can be defined may result in the exclusion of many potentially useful (Q)SARs. In addition, it was noted that misinterpretations were made during the study regarding the derivation of the regression equations and internal methods of ECOSAR. For example, it was noted that (Q)SARs for which it "appeared" that chemicals outside the log Kow domain of the method were used in the training set were considered "unreliable". However, those data are supplied in the ECOSAR Technical Reference Manual (ECOSAR, 1996) simply for reference to support the Log Kow cut-offs for those chemical categories, and not used directly in the development of the regression equations. Therefore, the (Q)SARs should not have been disregarded based upon that specific performance standard. This, in turn, appears to be another potentially confounding factor in the proper application of global internal performance standards. Without the benefit of extensive experience with a model, it may be difficult to delineate all of the interpretive guidance material and clearly understand the internal methods of a program.

102. An objective external evaluation of the predictive accuracy of a model is always desirable when determining its usefulness within a specified framework. However, it is often difficult to perform a truly representative evaluation of the predictivity using standard external performance measures without first considering the context within which a (Q)SAR model will be used to support chemical management decisions. It is important to understand these parameters before commencing an external evaluation, as different situations may lead the assessor to different conclusions regarding the appropriateness of a particular model.

103. In its most simple design, an external evaluation uses chemicals not employed in the development of the model and takes the form of a direct comparison between the experimental and estimated values for the chemicals. When the predicted endpoint is quantitative (provides a numeric value), a regression analysis is performed comparing the experimental and estimated data to ascertain the coefficient of determination (r2) for the model. This coefficient of determination is used as a surrogate measure for the predictivity. The higher the r2 value, the greater the correlation between experimental and estimated values, the better the predictive accuracy of the model. This is an appropriate scientific measure for the predictivity of a model; however, in some cases it may not reflect the true predictive power of a (Q)SAR within a particular decision making framework.

104. For example, regulatory/decision making bodies often use a set of preliminary classification criteria to make decisions regarding the potential fate and effects of chemicals and may not actually require the use of the discrete experimental or estimated values themselves. These classification schemes typically define ranges to allow the assessors to make more qualitative calls regarding the chemical of interest. Within the U.S. EPA/OPPT New Chemicals Program, (Q)SARs and classification schemes are used in screening and priority setting to identify potentially hazardous chemicals of concern from the universe of industrial chemicals. When looking at aquatic toxicity for example, a compound is considered highly acutely toxic to fish if the predicted LC50 value for the chemical is less than 1.0 mg/l. During the initial hazard assessment of a PMN review, it is immaterial whether the predicted value is in fact 0.100 mg/l or 0.0001 mg/l, as it will still lead the assessor to the same conclusions regarding the potential aquatic hazard of the chemical (high hazard concern) and the substance will thus go to the next step for further risk/hazard review. Therefore, within the context of specific regulatory/decision making frameworks, the predictivity of the model seems more appropriately measured when the quantitative values are overlaid on the respective classification schemes in order to truly represent how many times the predictions led the assessor to the right conclusions within that framework. Unlike the more traditional statistical approaches, this classification technique allows the models to be evaluated directly for their applicability within a given regulatory/decision-making framework.

105. This type of classification approach was used by Tunkel et al. (2004, 2005) to assess the predictivity of four fish aquatic toxicity models (ECOSAR, OASIS, MCASE, TOPKAT) and dramatic differences were seen between the conclusions drawn from a purely statistical assessment of the models, compared to those drawn from the overlay of the predicted values to a classification scheme. The classification scheme used in the assessment was derived from the U.S. EPA/OPPTs New Chemicals Program criteria for low, moderate, and high aquatic toxicity hazard. The external validation set was derived from chemicals submitted under the U.S. HPV Challenge Program. First, a traditional statistical assessment was made by determining the coefficient of determination (r2) for each model and secondly, an additional classification assessment was completed by determining the number of times a correct prediction occurred where the experimental and predicted value would receive the same hazard call within the U.S. EPA/OPPTs New Chemicals Program. Results of the assessments revealed that the two models with the highest r2 values, ECOSAR (r2 = 0.74) and OASIS (r2 = 0.74), had different predictive accuracies when using the classification assessment methodology. Even though these models originally had the same r2 value, ECOSAR predicted the hazard category correctly for 71% of the chemicals, while OASIS predicted less than half (40%) of the chemicals correctly. Additionally, changing the criteria of the classification scheme also changed the relative predictivity of the aquatic toxicity models. Using the OECD classification criteria for aquatic hazard (in place of the U.S. EPA/OPPTs New Chemicals criteria), ECOSAR now predicted 67% of the chemicals correctly while OASIS dropped to 27%.

106. These results demonstrate that the types of methods used to evaluate a model's predictivity will ultimately affect the conclusion concerning a model's suitability within a decision-making framework. OECD may provide guidance on proper methods for external validation, but the regulatory authorities

ultimately employing the (Q)SAR should complete a validation of the model under the auspices of their own chemical management framework in order to assess a method's ultimate applicability.

107. (Q)SAR models are inevitably associated with limitations in terms of the types of chemical structures, physiochemical properties, and mechanisms of action for which they may be appropriate. However, defining domain of applicability by investigating the structural components of the training set and associated descriptors to devise inclusion and/or exclusion rules for a model appears to be an activity with several different "precise" outcomes. Additionally, implementing domain evaluations that single out specific (Q)SARs for evaluation can dramatically restrict the number of chemicals able to be run through a program and, often, does not necessarily prevent misuse.

108. There are many scientifically acceptable methods that define domain using such parameters as structural fragments, descriptor ranges, and confidence interval parameters in the respective assessments. However, each method may lead the assessor to different conclusion regarding an appropriate domain of applicability for a model. Without consensus on which particular methods are considered acceptable and understanding that some Member countries such as the U.S. may be faced with inherent legislative constraints, the interpretive guidance proposed for Principle 3 to determine the domain of applicability remains rather vague and difficult to implement.

109. Additionally, in an effort to promote transparency of a model, it can be difficult to assemble the complete set of information regarding internal performance standards of a (Q)SAR as defined under the OECD's interpretive guidance for Principle 4 (OECD, 2004a&b). Some of the information regarding the derivation of the model may actually be confidential or protected (as seen with CBI data submitted to U.S. EPA/OPPT under the New Chemicals Program) and therefore, restricted from being disseminated to a third party. Due to these restrictions, it seems unsuitable to make the usefulness of particular (Q)SARs contingent upon availability of all internal performance standards. Studies have shown that even models initially considered "inappropriate" when evaluated under guidance similar to that proposed for OECD Principle 4, provided useful hazard predictions.

110. On the subject of predictive power, U.S. EPA's experience over the last 20 years regarding (Q)SAR integration into a regulatory framework shows that the level of acceptable uncertainty and/or predictivity is best defined starting from within the constraints of the particular regulatory/decision making framework. Clearly, a (Q)SAR estimation tool can not be expected to exceed the accuracy of results obtained from experimental test data. Furthermore, it is difficult to perform a representative evaluation of a model's predictivity using external test data without first considering the nature of the framework within which the (Q)SAR model will be used. When providing interpretive guidance for principle 4 to determine predictivity, OECD committee members should consider that each Member country may use (Q)SARs for different types of chemical management decisions and/or priority setting and that each may have their own unique system for classification of chemicals. If there are multiple estimation methods available for a specific endpoint, then the most appropriate method should be selected based on the results of an external evaluation conducted specifically under the classification scheme used to support chemical management decisions within the framework to which the (Q)SAR will be applied.

111. It has been the experience of the U.S. EPA/OPPT that building an effective regulatory program which utilizes (Q)SAR approaches in chemical management decisions begins from within the constraints of the specific framework. Thus, it will be very difficult to rigorously define universal (Q)SAR principles for acceptability that will fit the demands of all regulatory frameworks because of distinct programmatic needs. This case study suggests that Member countries may wish to consider the Setubal criteria, and subsequent OECD principles for (Q)SAR validation, as appropriate scientific goals that provide generic base-line guidance for integrating (Q)SARs into regulatory frameworks but not an inflexible standard for the global adoption of (Q)SARs. Flexibility will be needed in the interpretation and application of each

principle because ultimately, the proper integration of (Q)SARs into any type of regulatory/decision-making framework depends upon the needs and constraints of the specific regulatory authority.

112. Highlighting another use of (Q)SAR within the agency, the U.S. EPA established and chartered the Endocrine Disruptors Screening and Testing Advisory Committee (EDSTAC) in October 1996 and they developed a conceptual framework to screen and test chemicals for endocrine disruption as part of this effort. The Agency attempted to validate existing-prototype QSAR models against several hundred TSCA HPV chemicals whose ER binding affinity had been measured. The models ability to predict binding affinity was limited, which was not surprising since the models had been developed with training sets that were not representative of the TSCA high production volume chemicals - the regulatory domain (inventory) of interest. Examples of existing (Q)SARs for predicting the binding affinities of chemicals to the estrogen receptors in multiple species have been reported. (Schmieder et al., 2003).

U.S. EPA/OPPT Voluntary Programs

113. While not a regulatory approach, voluntary programs have been another way to promote the application of these tools in the United States. Voluntary activities provide a vehicle to help educate both regulatory authorities, as well as chemical industry stakeholders, in the use and interpretation of these tools for chemical assessments. For example, U.S. EPA's High Production Volume (HPV) Chemical Challenge Program ("Challenge Program") was developed to challenge the chemical industry to voluntarily compile a Screening Information Data Set (SIDS) for chemicals on the U.S. HPV list. The SIDS set provides basic screening data needed for an initial assessment of the physico-chemical properties, environmental fate, and human and environmental effects of chemicals. The information used to complete the SIDS can come from either existing data, new tests conducted as part of the Challenge Program, or for some endpoints, they may be generated through the use of (Q)SAR. The Challenge Program chemical list, available online (U.S. EPA, 2002c), consists of about 2,800 HPV chemicals reported under the Toxic Substance's Control Act's 1990 and 1994 Inventory Update Rule (IURs). The large number of chemicals on the list makes it important to reduce the number of tests to be conducted, where this is scientifically justifiable. Structure-activity relationships [such as (Q)SARs] may be used to reduce testing in at least three different ways. First, by identifying a number of structurally similar chemicals as a group or category, and allowing selected members of the group to be tested with the results applying to all other category members. Second, by applying SAR principles to a single chemical that is closely related to one or more better characterised chemicals ("analogues"). Or third, a combination of the analogue and category approaches may be used for individual chemicals.

114. The U.S. EPA has drafted a guidance document to assist sponsors and others in constructing and supporting SAR arguments for potential application in the Challenge Program (U.S. EPA, 2002d). The final guidance will draw on experience from the OECD SIDS program, the U.S. EPA Premanufacture Notification (PMN) program, and other sources available in the literature. OECD guidance on the use of (Q)SARs is also available (OECD, 2000). The environmental fate and aquatic toxicity (Q)SARs rely heavily on physico-chemical properties as inputs, and are similarly structured in terms of models, chemical classes, and regression equations. However, "accepted (Q)SARs" (cases in which ample data are available for a given chemical class) are not available for certain chemical classes for ecotoxicity endpoints. For biodegradation (estimated using BIOWIN), many chemical substructures are poorly or not represented in the various BIOWIN training sets.

115. Additionally, as part of its continued interest in increasing efficiency through innovative processes and voluntary partnerships, the Agency, in 2002, launched the Sustainable Futures Initiative (http://www.epa.gov/oppt/newchems/pubs/sustainablefutures.htm: accessed 1 June 2006) which is a voluntary program designed to assist industry in learning how to pre-screen their chemicals using EPA/OPPT predictive models and methods. A combination of training and technical assistance in the use

of EPA chemical risk screening tools and regulatory incentives are used to promote the development of safer chemicals and the identification of safer chemical alternatives. Industry response to Sustainable Futures has been very strong and equally positive. Well over 300 chemical companies and other stakeholders have taken Sustainable Futures hands-on training and OPPT has seen a significant increase in the number of new chemical submissions from Industry that contain predictions from these Sustainable Futures chemical risk screening methods.

116. A number of companies participating in Sustainable Futures have conducted case studies which illustrate various applications of the U.S. EPA/OPPT chemical risk screening tools [Find at OPPT Sustainable Futures website: http://www.epa.gov/oppt/newchems/pubs/sustainablefutures.htm: accessed 1 June 2006]. For example:

- Eastman Kodak sponsored a cost accounting study by the Tellus Institute that compared Kodak's operations both before and after using the screening models (Votta and White, 2000). Kodak's Tellus Report showed that by prescreening their chemicals at the R&D stage, Kodak reduced product development costs by 13% to 100%, reduced time to market, and reduced generation of chemical waste.

- PPG Industries, working with OPPT, compared measured acute aquatic toxicity data with Structure Activity Relationship (SAR) predictions for 38 compounds that had already been commercialized at PPG. Results showed 91% agreement between predictions and measured data (Chun et al. 2000). In another PPG case study, a PPG Industries toxicologist compared measured data on seven selected chlorobenzenes with OPPT's PBT Profiler predictions, and also got similar positive results (Burleigh-Flayer, 2000).

- SC Johnson used the OPPT PBT Profiler to evaluate over 2000 of the chemicals in their inventory for potential PBT concerns and concluded that they needed to assess only 173 of them (Weeks, et al. 2002).

- The Bayer Corporation also used the OPPT PBT Profiler to evaluate nine materials under consideration for commercialization and concluded that this tool helped them to evaluate and prioritize such chemicals at the R&D stage (Ruppel-Kerr, 2003).

- Eaton Aeroquip Inc. applied the OPPT PBT Profiler to two chemicals associated with their extrusion processes and four different cutting fluids. As a result of their assessment, Eaton Aeroquip will now use the PBT Profiler as a first check for new chamicals and processes before they enter their facilities and will incorporate OPPT predictive models into their current Environmental Management System (EMS) (Rippeon, 2004).

U.S. Food and Drug Administration

Use of (Q)SAR and In Silico Technology within the FDA

117. This case study is based on the experiences of the US Food & Drug Administration (FDA) in the application of (Q)SARs to provide decision support for regulatory and research issues, and includes recommendations with regard to the current OECD principles to better address these specific regulatory needs. The study is in part based upon the use of SAR and (Q)SAR models in use within: 1) a wide variety of decision support applications at the FDA Center for Drug Evaluation and Research (CDER); 2) the FDA Center for Food Safety and Applied Nutrition's (CFSAN) specific regulatory framework for the assessment of food contact substances (FCS), and for cosmetic ingredients; and 3) FDA's National Center

for Toxicological Research (NCTR), Center for Toxicoinformatics (CT) has developed (Q)SAR models to identify chemicals with endocrine disrupter properties.

118. The mission of the FDA is, in part, to protect the public health by assuring the safety, efficacy, and security of human and veterinary drugs, biological products, medical devices, the US food supply, cosmetics, and products that emit radiation. The FDA is also responsible for advancing the public health by helping to speed innovations that make medicines more effective, safer, and more affordable; and helping the public get the accurate, science-based information they need to use medicines to improve their health.

119. The FDA has historically been a data-rich organization that has relied upon the results of *in vivo* animal toxicology studies and human clinical trials to assess the safety and efficacy of FDA regulated substances, and these data come from FDA-regulated industry. However, over the past decade there has been a growing problem in which the applied sciences used to produce FDA regulated products have lagged behind the basic sciences. The following documents the emergence of SAR, (Q)SAR, and other *in silico* technologies into applied science of FDA regulated industry, and into a host of applications within FDA.

Critical Path (FDA/CDER)

120. In keeping with its mission, the FDA has issued a report to address the growing problem of moving basic discoveries to the market in a timely manner where they can be made available to patients (U.S. FDA 2004). The report evaluates how this problem came about and offers a way forward. It highlights examples of FDA efforts that have improved the critical path and discusses opportunities for future efforts. "Not enough applied scientific work has been done to create new tools to get fundamentally better answers about how the safety and effectiveness of new products can be demonstrated, in faster time frames, with more certainty, and at lower costs" (page ii). "A new product development toolkit — containing powerful new scientific and technical methods such as animal or computer-based predictive models, biomarkers for safety and effectiveness, and new clinical evaluation techniques — is urgently needed to improve predictability and efficiency along the critical path from laboratory concept to commercial product" (page ii). In Figure 6 of this report, the *'old path'* of development of drugs [basic research preclinical development (*in vitro* and animal testing) clinical development (human clinical trials) drug approval, marketing and post-market surveillance] has been modified to reflect new basic science opportunities. The '*new critical path*' now incorporates structure activity relationships (SARs) and computer models to select the best lead substances for preclinical development and animal testing.

121. The need for and the emergence of *in silico* science has also been discussed at a recent FDA Science Forum (2001). PriceWaterhouseCoopers predicted that electronic research and development (e-R&D) will become the primary science in this decade and supercede the current primary science of laboratory animals and patients (PriceWaterhouseCoopers 2001). Likewise, Stanski predicted that *in silico* modeling and simulation of human clinical trials are projected to supercede standard clinical trials (Stanski 2001). This paradigm shift in using *in silico* technologies first to determine the most efficient biological testing strategy has been suggested by others (Dearden 2003; McGee 2005; Wilson *et al.* 2003).

Computational Toxicology Program and Computational Toxicology Consulting Service (FDA/ICSAS)

122. The FDA/CDER's Informatics and Computational Safety Analysis Staff (ICSAS) has a Computational Toxicology Program and Computational Toxicology Consulting Service to apply and develop (Q)SAR software programs that is providing decision support information for a variety of Agency regulatory and research activities (Matthews, *et al*. 2000; http://www.fda.gov/cder/Offices/OPS_IO/ICSAS.htm; accessed 1 June 2006). The Computational Toxicology Program was established in 1994 to evaluate the predictive performance of commercially

available (Q)SAR software programs to determine whether one or more of the programs could provide reliable estimates of the potential toxicological activities of chemicals not present in the training data sets. The FDA Computational Toxicology Consulting Service was established in 1997 to provide regulatory *in silico* toxicology decision support information in situations where the results of *in vivo/in vitro* non-clinical toxicology studies were either inadequate or unavailable. At FDA/CDER this includes screening of lead pharmaceutical candidates and the evaluation of contaminants and degradation products of pharmaceutical products for hazard identification. The *in silico* program and service have grown and through leveraging agreements a battery of different (Q)SAR software programs have been developed and validated (Contrera *et al.* 2004, 2005; Matthews *et al.* 1998, 2000, 2004, 2006a, 2006b).

Food Contact Notification (FCN) Program (FDA/CFSAN/OFAS)

123. FDA/CFSAN's Office of Food Additive Safety (OFAS) is responsible for ensuring the safe use of U.S. food ingredients and food packaging (Bailey *et al*. 2005). OFAS evaluates safety information in industry submissions for the use of various categories of food substances; historically these submissions were primarily petitions for food and color additives and generally recognized as safe (GRAS) substances. In 1997, Section 309 of the FDA Modernization Act amended the US Federal Food, Drug, and Cosmetic Act (or the US FD&C Act) to establish a new process, referred to as the food contact notification (FCN) process, as the primary method of authorizing new uses of food additives that are food contact substances (FCS).

124. Under the FCN process the submitter bears the burden of demonstrating that the intended use of the FCS is safe. The safety standard is the same for all food additives and a FCN for a FCS is effective only for the manufacturer identified in the FCN. A FCN becomes effective and the FCS may be legally marketed for the proposed use 120 days after acceptance of the submission, unless FDA concludes that the safety has not been demonstrated and so informs the notifier. In order to meet this deadline, CFSAN/OFAS instituted a FCN administrative process in 1999 that makes use of SAR and (Q)SAR analysis in the safety evaluation of FCSs (Bailey *et al*. 2005). In practice the majority of the FCSs are industrial chemicals that are used in the manufacture of packaging materials which transfer to food products, and industrial chemicals that leach from equipment used in the manufacture, transportation, and holding of food products. As of March 2005, FDA had received and processed over 500 FCNs.

125. Although used in past safety evaluations, (Q)SAR became an integral part of the review process with the inception of the FCN program. Moreover, FDA/CFSAN has a SAR team that performs preliminary, qualitative SARs on every chemical with an exposure submitted in an FCN. The need for a quantitative SAR is determined following the full evaluation of all safety data (submitted and available to FDA). SAR analysis is performed using a variety of commercial and in-house software, including *MC4PC* (MultiCASE, Inc.), *OncoLogic®*, and *Enterprise* (Leadscope, Inc.). As the majority of exposures resulting from food contact use are below 50 ppb in the diet, SAR analysis in the review of FCNs is highly beneficial to the completion of a thorough, robust, and efficient safety assessment.

Cosmetic Ingredients (FDA/CFSAN/OCAC)

126. The US FD&C Act does not require that cosmetic products and their constituent raw materials (except color additives) pass the rigorous pre-market FDA approval or notification processes used for color additives, pharmaceuticals, and food additives. However, cosmetic ingredients and cosmetic products are expected to undergo "adequate safety substantiation" prior to their being offered for retail sale to the consumer, and manufacturers have the burden of assuring that such safety substantiation has, in fact, been performed for each cosmetic product and ingredient formulated therein prior to marketing FDA does not, however, specify whether such safety substantiation packages are to be comprised of historical proprietary data on similar formulations in company files, authoritative literature (peer review or trade), preclinical

study results (animal, *in vitro*, or *in silico*), human clinical protocol study results, or post-marketing surveillance data.

127. The lack of FDA authority to require the submission of safety data from cosmetics manufacturers forces the FDA to rely heavily on published reports, voluntary data submissions, and other available resources to make its safety assessments concerning cosmetic raw materials (ingredients); QSAR and quantitative structure toxicity relationship (QSTR) approaches are viewed in this context as a valuable "bridging" technology, providing predictive toxicology insights concerning the more than 12,500 cosmetic ingredients currently available (Gottschalck and McEwen 2004), in the absence of experimental or clinical data. *TOPKAT*® (Accelyrs, Inc., formerly Health Design Inc.), one of several available computer-assisted predictive toxicology (QSTR) packages, was evaluated in 1996 by the FDA/CFSAN Office of Cosmetics and Colors (OCAC) as an *in silico* QSTR approach for the safety assessment of structurally-defined single cosmetic ingredients and color additives (Yourick *et al.* 1996). In addition, OCAC has also explored the use of (Q)SARs to evaluate human percutaneous absorption data of chemicals using physicochemical data (Bronaugh and Barton 1993). Additionally, *TOPKAT*® and other QSTR approaches (Barratt and Basketter 1994) may enhance the ability to respond to safety-related questions concerning functional and incidental cosmetic ingredients and environmental trace-level contaminants from other regulatory competent authorities and stakeholders. In 1997 the Cosmetic Ingredient Review Expert Panel, a Cosmetic, Toiletry, and Fragrance Association-funded industry initiative originated in the 1970s to conduct authoritative peer-review level safety assessments of cosmetic ingredients [on which OCAC serves in an *ex-officio* (non-voting) liasion capacity] began using QSTR data from the *TOPKAT*® software program to assist in the prioritization of cosmetic ingredients for safety assessment evaluation (Bergfeld and Andersen 1998).

Endocrine Disrupters (FDA/NCTR/CT)

128. FDA's National Center for Toxicological Research (NCTR), Center for Toxicoinformatics (CT) has an established program for the investigation of (Q)SAR models to identify chemicals with endocrine disrupter properties (http://www.fda.gov/nctr/science/centers/toxicoinformatics/edkb/; accessed 1 June 2006). First, the group applied three different QSAR methodologies (Tong, *et al*. 1998), Comparative Molecular Field Analysis (CoMFA), classical QSAR (utilizing the CODESSA program), and Hologram QSAR (HQSAR) to screen large data sets of chemicals as endocrine disrupting compounds (EDCs). Given the fact that any (Q)SAR approach has strengths and weaknesses and using a single model for priority setting poses difficulties for regulatory application, the group has adopted an approach to rationally combine different (Q)SAR models into a sequential "Four-Phase" scheme according to the strength of different (Q)SAR models (Tong *et al*. 2003a).

129. A progressive Phase paradigm is used as a screen to reduce the number of chemicals to be considered in each subsequent Phase. Therefore, these four phases work in a hierarchical way to incrementally reduce the size of a data set while increasing precision of prediction. Within each phase, different models have been selected to work complementarily in representing key activity-determining structure features to minimize the rate of false negatives. More recently this group has developed a novel (Q)SAR method, Decision Forest, which combines the results of multiple heterogeneous but comparable Decision Tree models to produce a consensus prediction (Tong *et al*. 2003b). In Decision Forest, the prediction confidence and applicability domain of a model can be realized (Tong *et al*. 2004), which has been demonstrated as one of the most important characteristics for determining the utility of a (Q)SAR model (Tong *et al*. 2005).

An Examination of the OECD Principles from an FDA/ICSAS Perspective

130. FDA/ICSAS has no specific comments on OECD Principles for (Q)SAR validation # 1 and #5. However, it may be difficult for FDA/ICSAS to apply OECD (Q)SAR Principles #2, #3, and #4 to the Agency's existing regulatory and decision making framework. The following illustrates why.

Intent of the OECD Principles

131. The OECD principles for (Q)SAR validation are intended to identify the types of information that are considered useful for the regulatory review of (Q)SARs. Taken together, the principles represent a conceptual framework to guide the validation of (Q)SARs, but they are not intended to provide criteria for the regulatory acceptance of (Q)SARs. The definition of acceptance criteria, where considered necessary, is the responsibility of individual authorities within the Member Countries. To facilitate the consideration of a (Q)SAR model for regulatory purposes, it should be associated with the following information:

1) a defined endpoint;
2) an unambiguous algorithm;
3) a defined domain of applicability;
4) appropriate measures of goodness-of–fit, robustness and predictivity; and
5) a mechanistic interpretation, if possible.

132. FDA/ICSAS believes that Principle #2 is extremely important when (Q)SARs are considered for, and used in, regulatory applications. Many commercially-developed models do not have adequate information on the (Q)SAR algorithm. Nevertheless, FDA/ICSAS has not ruled out commercially developed models and believes they can be useful if they meet most of the OECD Principles for (Q)SAR validation and contain (Q)SAR algorithms and training data sets that are sufficiently described and adequately supported with validation test data. This information allows users to replicate validation statistics, evaluate the domain of applicability, and consider the suitability of the software for their application.

133. FDA/ICSAS has a policy of developing and improving upon commercially-developed (Q)SAR models to meets the needs of our scientists and our regulated industry clients. In our experience, many commercial and free-ware (Q)SAR models can have serious flaws in the algorithm of the (Q)SAR model, logic of the software program platform, or scoring of toxicological activity of the training data set which need to be addressed before the (Q)SAR software programs and (Q)SAR models can be used internally for FDA regulatory applications. Flexible software platforms are available that allow module building and manipulation of data sets and decision rules while still maintaining the proprietary status of model algorithms. The following are some specific examples in which sufficient details on the construction of a commercial (Q)SAR model were not supplied and upon use of the (Q)SAR, the programs were found unsuitable for FDA/ICSAS purposes:

- FDA/ICSAS discovered that one commercial (Q)SAR model for carcinogenicity in rodents had not disclosed that only chemicals that caused malignant tumors were scored as positive in the training data set. Since the FDA and EPA regulate chemicals that cause a significant increase in benign tumors as carcinogens, this commercial (Q)SAR model was inappropriate for our applications.
- FDA/ICSAS discovered that one commercial (Q)SAR model for developmental toxicity in rodents had not disclosed that fetal survival, fetal growth, and classical morphogenesis (teratology birth defects) were given equal weight in the (Q)SAR model. Since FDA/CDER reports morphogenesis findings separately from non-specific chemical toxicity to the fetus (fetal

survival and fetal growth), this commercial (Q)SAR model was inappropriate for our applications.

- FDA/ICSAS has discovered that many commercial (Q)SAR models have not disclosed how 'weak and equivocal findings,' or chemicals with structure activities that were outliers, are treated in the (Q)SAR model. Since the FDA and EPA regulate all types of chemicals, and many regulated chemicals have unusual structure activity relationship findings, it is paramount that they disclose such information to ensure that the (Q)SAR model algorithm is unambiguous.

These scenarios illustrate why it is important for (Q)SARs to be adequately described and reported in the literature.

OECD Principle #3: A Defined Domain of Applicability

134. OECD Principle #3 requires a domain of applicability of the (Q)SAR model be defined in terms of molecular coverage of the test molecule relative to the molecules in the training data set of the model. The domain of applicability is essentially the suitability of the software for a specific application of the software user. The direct effect of having a specified domain of applicability is to identify chemicals that have molecular properties similar to those in the training data set and can be reliably predicted, and to identify those that are not covered and may result in poor predictions.

135. FDA/ICSAS uses a battery of (Q)SAR models for the "high-end human health effect" endpoints, and the same exact training data set is used with different SAR and (Q)SAR software platforms. From FDA/ICSAS' perspective, an appropriate assessment of the domain of applicability is best characterized by simultaneously considering the sum of a program's quantitative and qualitative methods. Currently, there are no agreed upon methods that can encompass both characteristics for all of the SAR/(Q)SAR software platforms being evaluated. Additionally, FDA/CDER and FDA/CFSAN must consider the different regulatory constraints under which they may need to perform their assessments. Taken together, it is strongly recommended that there is flexibility in the application of Principle #3. For example, FDA/ICSAS and FDA/CFSAN/OFAS have the dilemma of being asked to make (Q)SAR assessments on all chemicals submitted; FDA/CFSAN/OFAS must make a decision within 120 days on FCSs, and FDA/ICSAS attempts to complete consultation reports within 14 days. Unfortunately, all of the (Q)SAR models that meet our four acceptance criteria have inherent restrictions on the domain of applicability of the model, and none of the models cover all substances submitted for analysis. This is one important reason why a battery of validated predictive software platforms employing different model algorithms is being developed by FDA/ICSAS.

OECD Principle #4: Appropriate Measures of Goodness-of–fit, Robustness and Predictivity

136. FDA/ICSAS could comply with portions of Principle #4 requirements and routinely reports the validation statistics for the different (Q)SARs and (Q)SAR programs used in computational toxicology applications.

137. However, FDA/ICSAS cannot comply with the OECD Principle for (Q)SAR validation #4 requirement for complete transparency in evaluating the internal predictive performance of a (Q)SAR model. This principle requires that the (Q)SAR model be accompanied by full disclosure of the internal performance information for the training set chemicals including chemical names, structural formulae, raw data, data for descriptor variables, data quality, data processing methods, methods for selection of variables, and any statistical methods employed in the derivation of the (Q)SAR. It is impossible for FDA/ICSAS to release all of the information regarding internal predictive performance of (Q)SAR model it develops for the following reasons:

1. Many of the FDA (Q)SAR models use knowledge derived from proprietary studies and the identity of the substances and their chemical structures are confidential. Proprietary animal toxicology and human clinical information for pharmaceuticals is regarded as proprietary because it is information for non-marketed and/or non-approved pharmaceuticals, and cannot legally be made available to the public by FDA. Furthermore, any such information used as part of an internal or external validation exercise to evaluate a model cannot be disclosed.

2. The Agency has developed (Q)SAR models using leveraged agreements with commercial software companies, and it is contractually restricted under these licensing agreements not to reveal the identity and activities of chemicals in the training data sets of the program. These licensing restrictions extend to significant amounts of confidential information derived from these training data, such as lists of structural alerts or proprietary molecular descriptors used by the (Q)SAR software and developed by the company.

3. Details of the (Q)SAR software program logic used by commercial software companies are usually described in principle in the literature and company publications, but the logic may not always be explained in sufficient detail to meet the needs of all investigators. Commercial software companies have a vested interest in protecting their software technology and computer code and are unwilling to divulge this intellectual property.

138. Since one of the most important goals of the FDA/ICSAS Computational Toxicology Program was to develop (Q)SAR models that could be used both internally and by regulated industry, and enable both parties to get identical predictions for the same test chemicals, FDA/ICSAS had to address the issues of proprietary study data and transparency of the (Q)SAR models. Likewise, the FDA wanted to develop (Q)SAR models with the best predictive performance and the widest chemical domain space possible. The non-proprietary data include both data from public sources and information for marketed pharmaceuticals that is available from the Center's Freedom of Information Office and has been published in drug labeling, the *Physicians' Desk Reference*, or other sources. In order to include knowledge derived from proprietary studies, the FDA/ICSAS and its collaborators have invented technical procedures in which the proprietary data are converted into non-traceable information before the FDA (Q)SAR modules are made publicly available.

Uncontrolled External Predictive Performance

139. FDA/ICSAS also has concerns for the overall value of using uncontrolled external data sets to evaluate the predictive performance of certain (Q)SAR models. In our experience, we routinely use all available experimental data at the time the model is constructed in order to create the best possible (Q)SAR model for non-congeneric chemical data sets that has: 1) the best predictive performance in external cross-validation experiments; and, 2) the widest possible chemical space domain. Independent external validation data sets are constructed at a later date and use new studies that are discovered after the model has been built. In our experience independent external validation data sets are uncontrolled data sets and poor tools to evaluate the predictive performance of models. The independent external data sets invariably exhibit the following:

1) they contain unbalanced chemical structure classes;

2) they contain only a few chemicals due to the low numbers of new chemicals being developed and marketed for which new test data could be obtained;

3) they are unbalanced in terms of chemical toxicities and frequently have too many inactive molecules; and they contain some novel new chemical structures that are not structurally similar to chemicals in the database and are outside of the domain of the model.

Case Study VI: European Commission

140. Under current European legislation, Directive 67/548/EEC requires new substances to be tested and assessed for possible risks to human health and the environment before they are marketed in volumes of 10 kg or more.

141. Existing substances are assessed under the provisions of Regulation (EEC) 793/93, known as the Existing Substances Regulation (ESR). This introduced a comprehensive framework, foreseeing that the evaluation and control of the risks posed by approximately 100,000 existing substances should be carried out in four steps: data collection, priority setting; risk assessment and risk reduction.

- Data Collection. The Regulation was initially concerned, in phases I and II of the data collection step, with the so-called High Production Volume Chemicals (HPVCs). i.e. substances which have been imported or produced in quantities exceeding 1000 tonnes per year and produced/imported between March 23, 1990 and March 23, 1994. In phase III of the data collection step, companies which produce or import existing substances in quantities between 10 and 1000 tonnes per year (Low Production Volume Substances or LPVCs) were required to submit a reduced data-set by June 4, 1998. All the data had to be submitted in a specific electronic format, the Harmonised Electronic DataSET (HEDSET) and is incorporated in the International Uniform ChemicaL Database (IUCLID).

- Priority Setting. In consultation with the Member States, the Commission regularly produces lists of priority substances that require immediate attention because of their potential effects to human health or the environment. The Commission and Member States utilise the information collection as a basis for selecting priority substances. Since 1994, four such priority lists have been published.

- Risk Assessment. Substances on priority lists must undergo an in-depth risk assessment covering the risks posed by the priority chemical to man and the environment. This risk assessment follows the framework set out in Commission Regulation (EC) 1488/94 and implemented in the detailed Technical Guidance Documents (TGD) on Risk Assessment for New and Existing Substances (EC, 1996, 2003). The EU Member States act as rapporteurs in the drafting of the risk assessment reports, and the European Commission mediates meetings, which attempt to reach consensus on the conclusions of the risk assessments.

- Risk Reduction. One of the possible outcomes of the risk assessment is that the chemical is considered to be a "substance of concern" and that "further risk reduction measures, beyond those already in place, are required." In such cases, a risk reduction strategy is developed and implemented by means of appropriate legal instruments, such as Directive 76/769/EEC on the restrictions in marketing and use of dangerous substances.

142. Since comprehensive risk assessments (and therefore testing requirements) have focused on priority existing substances, increasing concerns over the lack of data for existing substances, and thus lack of regulatory consideration on the vast majority of existing substances in commerce, led to various EU Member States initiating the development of a new Chemicals Policy, REACH (see Case Study XII).

143. In this section, some examples are given to illustrate how (QSARs and other non-test methods have been used in the EU in the assessment of New and Existing Substances. A more comprehensive overview is given in the final report of the TAPIR project (ECB, 2005), developed in the context of one of the REACH-implementation projects.

Priority Setting

144. The priority setting procedure focused on ranking the approximately 2500 High Production Volume Substances on the EU list of 101,195 Existing Substances (EINECS). For this purpose, the EU Risk Ranking Method (EURAM) was developed (van der Zandt and van Leeuwen, 1992; Hansen et al. 1999; van Haelst et al 2000). EURAM was designed to select suitable data from the IUCLID (International Uniform Chemical Information) database and to provide a ranking of substances based on their potential risk to humans and the environment. It calculates an Environment Score (ES), based on environmental exposure and effects scores, and a Human Health Score (HS), based on human exposure and effects scores. The EURAM rankings were not directly used to set priorities for testing, but they were used as the basis for technical discussions leading to the preparation of the third and fourth priority lists, which also took into account national priorities of the Member States.

145. For the purposes of the REACH regulation, the development of priority setting procedures is foreseen for implementation in the Evaluation and Authorisation procedures. The scientific and technical preparations are being carried out in the context of the REACH-implementation projects.

Classification and Labelling

146. The "EU Labelling Guide" (Annex VI of Directive 67/548/EEC) contains criteria that are based largely on the interpretation of experimental test results. Nevertheless, Section 1.6.1 of the Annex recognises that "validated" QSARs can be used for the classification and labelling of substances with the following wording:

> "For substances the data required for classification and labelling may be obtained: … The results of validated structure-activity relationships and expert judgement may also be taken into account where appropriate."

147. The use of a QSAR in Annex VI can be illustrated by the use of predicted logKow values in the classification of long term aquatic hazard (bioaccumulation). When valid test data on the preferred predictor of bioaccumulation (fish BCF) are not available, the BCF value can be calculated by using a QSAR or by using a decision rule based on the (experimental or calculated) log Kow value, provided that the QSAR is considered valid for the chemical in question. Classifications based on logKow values are more conservative than those based on experimental BCF data (i.e. application of logKow-based trigger results in the classification of more chemicals).

148. The use of SARs in Annex VI is illustrated can be illustrated by the fact that isocyanates are included in the EU List of Dangerous Substances under the assumption that an isocyanate is likely to be a respiratory sensitiser, unless there is evidence to the contrary. Similary, organic peroxides are assumed to be skin irritants, unless evidence suggests otherwise. In addition, read-across from structural analogues that are known sensitisers or carcinogens can be used as supporting evidence for classifications based on sensitisation or carcinogenicity, respectively.

149. The EU List of Dangerous Substances, Annex I of Directive 67/548/EEC, also contains a significant number of group entries, in which a classification is assigned to the entire group. There are more than 90 group entries covering more than 1900 chemicals. In addition, by the time of 21st Adaptation to Technical Progress (ATP) of the Directive, 149 coal-derived complex substances had been assigned to 41 groups, and 543 oil-derived complex substances had been assigned to 22 groups in Annex I.

150. Official EU classifications in Annex I are produced according a consensus process in which the EU Member State authorities agree on the classification. However, the classification criteria in Annex VI are also implemented by the manufacturer and/or importer to provisionally classify and label chemicals,

and a number of industry sectors have published guidance for the "self-classification" of chemicals within their responsibility.

151. To support the self-classification process, the Danish EPA published an "advisory list for self-classification of dangerous substances". The list of suggested hazard classifications was derived by using predictions from (Q)SAR models obtained or developed by the Danish EPA for the following endpoints: acute oral toxicity, skin sensitisation, mutagenicity, carcinogenicity and danger to the aquatic environment. The models were used to make predictions for the approximately 47,000 discrete organic substances on the EINECS list. The Danish Advisory List contains 20,624 chemical substances with suggested classifications for one or more of the dangerous properties, and is searchable via the internet (Danish EPA, 2001). The Danish QSAR database is also accessible via the ECB website (http://ecb.jrc.it/QSAR).

Risk Assessment

152. In order to ensure consistency in the Environmental Risk Assessment (ERA) process, the European Commission produced, in 1996, a comprehensive Technical Guidance Document (TGD) to support the Directive on New Substances and the Regulation on Existing Substances (EC, 1996). This document includes guidance on the use of (Q)SAR models in the ERA process in terms of where they should be used, how they should be used and which ones should be used. Whilst considerable information is provided in the TGD regarding the prediction of ecological effects and environmental fate, no formal recommendations are given on the use of (Q)SARs for the prediction of human health effects. The TGD was updated in 2003 (EC, 2003), but only editorial revisions were made to the section on QSARs.

153. The four major uses for (Q)SARs in ERA are identified in the TGD are:

- Data Evaluation. Acceptable (Q)SARs may be used as a supporting tool to evaluate the adequacy of the available experimental data, for example when the validity of the test data are not obvious. This may occur when incomplete data are available on the test and/or the test guideline differs in some way from current OECD test guidelines.

- Decision for Further Testing. In those cases where a Predicted Environmental Concentration (PEC)/Predicted No Effect Concentration (PNEC) ratio established using test data is greater than one, there will be a requirement to determine whether or not additional testing is needed to allow a refinement of the PEC/PNEC ratio. In order to facilitate this decision, all available test data should be reconsidered along with estimates established using acceptable (Q)SARs. If PEC/PNEC ratios derived using the (Q)SARs suggest that further testing is required, then generally a chronic test should be conducted on the species which showed the lowest estimated No Observable Effect Concentration (NOEC).

- Establishing Specific Parameters. Acceptable (Q)SARs can be used for the estimation of specific (input) parameters used in the risk assessment, particularly in the exposure assessment when no measured data are available to enable derivation of the PEC.

- Identifying Data Gaps. Acceptable (Q)SARs can be used for preliminary assessment of endpoints which are not part of the base set of data and for which information is not available.

154. Rules for the use of (Q)SARs in environmental effects assessment are provided in the TGD. These simply state that a (Q)SAR is considered to be acceptable for a particular use within the risk assessment process if: a) the (Q)SAR applied has been validated by an appropriate process; and b) the estimate possesses the necessary accuracy for the intended use.

155. The TGD gives general guidance and rules, but most examples can be found in the risk assessment reports produced within the context of the ESR. From these reports, it has been concluded that (Q)SARs have generally not been used as stand-alone methods, but in conjunction with available test data, although the reliance on (Q)SAR estimates has depended on the nature of the endpoint.

156. For physicochemical properties, predictions have seldom been made, because experimental data (which are generally preferred) have been available. In a few cases, physicochemical properties have been predicted. For example, a QSAR was used to estimate the vapour pressure of V6 due to practical difficulties in performing the test. The validity of the QSAR estimate was established by using measured data on TCPP and TDCP.

157. QSAR estimates have been used routinely for predicting certain environmental fate parameters of organic substances, partly because the experimental determination of these parameters can be difficult and/or expensive, and partly because the information is not normally required in the regulatory submissions. For example, the AOPWIN program has been used to derive atmospheric degradation rate constants, and log Kow has been used as a predictor of the solid-water partitioning coefficient. For a few chemicals (e.g. trichloroethylene, nonylphenol), QSAR-generated BCF values have been used instead of a range of measured values.

158. There are a few cases in which aquatic toxicity has been predicted. For example, no aquatic toxicity data were available for 1,3-butadiene due to the physical nature of the substance (volatile, carcinogenic and flammable), so the toxicity was estimated by using a QSAR from the Technical Guidance Document (TGD) on Risk Assessment (EC 2003). The validity of this estimate was established by comparing predictions made by the QSAR for two structurally similar substances, isoprene (2-methyl-1,3-butadiene) and 1,3-pentadiene, for which experimental data were available. In another case, QSARs were used to estimate the acute and chronic aquatic toxicity for octabromodiphenyl ether and decabromodiphenyl ether. The estimates suggested that toxicity would not be expressed below the limit of water solubility, which provided argumentation against the need to perform chronic toxicity tests on aquatic organisms on these two substances.

159. Grouping approaches have been used in the context of the Existing Substances Regulation for both registration and risk assessment. Examples include metals and their compounds (e.g. chromium (V)I, nickel, cadmium and zinc) as well as petroleum substances. The Hydrocarbon Block Method is a grouping method for evaluating the environmental fate and effects of complex hydrocarbon mixtures. Individual hydrocarbons with similar properties are grouped into “Hydrocarbon Blocks” and a surrogate chemical from the block is selected to represent the properties of the whole block. The properties of individual blocks are then used to predict the environmental fate and effects of the complex hydrocarbon substance. The method has been used to predict both environmental fate properties and effects on environmental species, for example in risk assessments for gasoline and naphthas.

PBT Assessment

160. In this section, PBT assessment includes the assessment of both PBT and vPvB potential. In the EU, PBT assessment has been carried out in the framework of the Commission’s “interim strategy for the management of PBT and vPvB substances” (EC, 2001), and in accordance with the strategy and criteria proposed in the TGD on risk assessment (EC 2003). The work has been carried out by the PBT Working Group, which is a subgroup of the Technical Committee on New and Existing Substances (TC NES). The activities of this Working Group have also served to gain experience in PBT assessment and to contribute to the development of guidance for PBT assessment under REACH.

161. In general, QSARs have been used in combination with experimental data, but have been used on their own for the selection of PBT candidates (where experimental data did not exist or was considered unreliable), and alongside experimental data for the confirmation of PBT status. An initial screening exercise, based on the use of both experimental and QSAR data for persistency, bioaccumulation and toxicity (aquatic and mammalian), led to the selection of 125 candidate PBTs with tonnages in the range 10-1000 metric tonnes.

162. The subsequent assessment of the candidate PBTs, using both existing experimental data and QSAR predictions in a weight-of-evidence approach, has resulted in some chemicals being deselected from the list, whereas others have been confirmed as PBTs, or targeted for further assessment. For persistence, the BIOWIN models have frequently been used in screening for persistency. In addition, a MultiCASE model developed by the Danish EPA, BIOHCWIN and CATABOL have been used in a few specific cases. For bioaccumulation, the BCFWIN model has been used, in addition to the TGD BCF model. For toxicity, QSARs for short-term aquatic toxicity to algae, fish and Daphnia have been used, generally when test data were available for one or more of the three organisms, but lacking for the remaining ones. These QSARs have included the ECOSAR models, the QSARs recommended by the TGD, as well as MultiCASE models developed by the Danish EPA. QSARs for chronic mammalian toxicity, reproductive toxicity and mutagenicity have been proposed for use in the screening for PBT candidates, but have not been decisive for T assignment. Read-across has been used on a case-by-case basis (e.g. read-across from 1,2,4-trichlorobenzene to 1,2,3 trichlorobenzene) and grouping approaches have also been used (e.g. diarylide pigments, in which different functional groups attached to a common substructure are proposed to account for differences in bioconcentration). In addition to use for predicting the PBT properties of individual substances, QSARs (and experimental data) have frequently been used to evaluate whether constituents of multi-component chemicals (mixtures) fulfill the PBT screening criteria.

163. The PBT Working Group also provides input to, and takes account of, international activities on PBTs and on the protection of the marine environment. Such activities include the UNEP Stockholm Convention on Persistant Organic Pollutants (POPS), and the Oslo-Paris (OSPAR) Convention for the Protection of the Marine Environment. The OSPAR Commission has developed a strategy with regard to Hazardous Substances that involved the application of a (Q)SAR-based prioritisation scheme (DYNAMEC) to produce an updated list of priority chemicals (OSPAR, 2000).

PROSPECTIVE USES OF (Q)SAR WITHIN OECD MEMBER COUNTRIES:

Case Study VII: Australia

164. Australian regulatory authorities do not currently use (Q)SARs in relation to the listing of chemicals on the Australian Inventory of Chemical Substances (AICS). With regard to new chemical assessments, limited use may be permitted in the National Industrial Chemicals Notification and Assessment Scheme (NICNAS). For example, data from chemical analogues may be used from structurally similar chemicals if toxicological and ecotoxicological data are unavailable on the notified chemical. From the analogue data, conclusions may be drawn on the health and environmental effects of the notified chemical. However, Australian authorities do not routinely accept SAR predictions without support from actual test data, e.g. physico-chemical properties and ecotoxicological effects.

Case Study VIII: Germany

Federal Institute for Risk Assessment (BfR)

165. In Germany, the assessment of new and existing industrial chemicals with regard to human health hazard is performed at the Federal Institute for Risk Assessment (BfR). To provide a tool for the evaluation of physico-chemical and toxic properties of notified substances, a computerised database of physico-chemical and toxicological data obtained from the new chemicals notification process was developed at the BfR (formerly the Federal Institute for Health Protection of Consumers and Veterinary Medicine (BgVV)). Since its first publication (Gerner *et al.* 2000), the database was updated to include now about 2000 entries. It has been used to develop a so-called 'Decision Support System' (DSS) containing a set of (Q)SAR tools for predicting the presence or absence of a substance's potential to cause local toxic effects. It was also used as a basis to evaluate the structural alerts for sensitisation incorporated into the DEREK expert system (Zinke *et al.* 2000). In detail, the following (Q)SAR tools were developed based on the database:

- A set of structural alerts for skin sensitisation (Gerner *et al.* 2004a);
- A set of structural alerts for skin irritation/corrosion potential (Hulzebos *et al.* 2005);
- A set of structural alerts for eye irritation/corrosion potential (Gerner *et al.* 2005);
- A set of physico-chemical exclusion rules for predicting the absence of skin irritation/corrosion potential (Gerner *et al.* 2004b);
- A set of physico-chemical exclusion rules for predicting the absence of eye irritation/corrosion potential (Gerner *et al.* 2005).

A mechanistic background for some of these tools was proposed in Walker *et al.* 2004.

166. Both the structural alerts and the exclusion rules were submitted for formal validation to the ECB in 2004. Within this project, the section on the validation of the skin irritation/corrosion rules and alerts can be regarded as a model for how to circumvent the problems associated with the validation of (Q)SARs built from confidential data: While the rules/alerts had been developed at a regulatory authority in one EU member state (BfR/DE), the project was supervised by an international body (the ECB), but independent validation was contracted to a regulatory authority in a second EU member state (RIVM/NL) with all parties having legal access to the confidential training set data.

167. Reports on the validation of the physico-chemical exclusion rules for skin irritation/corrosion and eye irritation/corrosion have already been published (Rorije and Hulzebos 2005, Tsakovska *et al.* 2005) and have demonstrated both the validity and relevance of the validated rules as well as their potential for the avoidance of a great number of animal tests for the respective local effects.

168. As soon as more experience will have been gained with the application of these tools in routine hazard assessment, it is envisaged that they could be used in the framework of so-called integrated testing strategies (Hoefer *et al.* 2004, Gerner *et al.* 2004c, Walker *et al.* 2005). To this end, current testing strategies would have to be revised in order to use (Q)SARs also for the prediction of the absence of relevant toxicity.

Federal Environment Agency (UBA)

169. The German Federal Environment Agency (UBA) uses (Q)SARs both to fill missing data needs and to give some assurance of the quality of the available experimental test data (Lange and Vormann 1995). UBA developed a (Q)SAR-based software system to assess the validity of predictions made for data

supplied with newly notified chemicals. The system predicted a wide range of fate and effect endpoints including fish and *Daphnia* acute toxicity. However in an evaluation of 64% of the substances assessed, (Q)SARs for some endpoints were not applicable because the substances were ionic, contained heavy elements, hydrolysed rapidly or were reaction mixtures which are outside the applicability domain.

Case Study IX: Italy

170. The Istituto Superiore di Sanita' (ISS) (National Institute of Health) is the main governmental regulatory authority involved in national and international regulatory programmes. Up to now no systematic use of (Q)SAR concepts and models has been made by the Italian authority for regulatory purposes, nor dossiers based on (Q)SAR evidence have been presented by notifiers. Only recently, read-across has been used to a limited extent in the assessment of biocides.

171. In recognizing the importance of alternative methods in the evaluation of substances, the ISS has developed and assessed (Q)SAR approaches, mainly for mutagenicity and carcinogenicity.

172. One area of activity at the ISS is the development of QSAR models for individual chemical classes of mutagens and carcinogens of primary environmental and industrial importance. This includes: a) the induction of aneuploidy by halogenated aliphatics (Benigni *et al.* 1993); b) the carcinogenicity of aromatic amines (Benigni & Passerini 2002); and c) the genotoxicity / carcinogenicity of aliphatic aldehydes (Benigni *et al.* 2005).

173. Another area of activity regards the development and evaluation of "general", predictive (Q)SAR models for noncongeneric sets of mutagens and carcinogens. This includes: a) the participation to two comparative exercises on the prediction of rodent carcinogenicity, held under the aegis of the US National Toxicology Program; b) the detailed analysis of the outcomes of the exercises; c) the evaluation of the performance of the main commercial prediction systems. These various activities are summarized in (Benigni 2005).

174. As a continuation of the latter activity, at present the ISS is involved in an EU Project on the Collection and Evaluation of (Q)SAR Models for Mutagenicity and Carcinogenicity. The project is supported by the European Chemicals Bureau, Institute for Health and Consumer Protection, Joint Research Centre, European Commission, and has the specific goal of evaluating the non-commercial (Q)SAR models for mutagenicity and carcinogenicity.

175. An area of activity is the revision and integration of the Structural Alerts for mutagenicity and carcinogenicity from both mechanistic and machine-learning studies (Benigni & Bossa 2006).In addition to its role for predictive toxicology, this research is aimed at contributing to the development of chemical categories, and to the clarification of chemical similarity concepts.

176. Finally, the ISS has developed, and made freely available at its web-site a database of chemical carcinogens (ISSCAN). It contains information on chemical compounds tested with the long-term carcinogenicity bioassay on rodents (rat, mouse); in addition to the toxicological data, it provides the structures of the chemicals in a machine-readable format, and thus can be used as a Chemical-Relational Database. The two main aims of this activity are to provide: a) revised, high-quality toxicological data; b) data coded in a QSAR-ready format. The ISSCAN database can be accessed from:

http://progetti.iss.it/ampp/hhhh/hhhh.php?id=233 [accessed 1 June 2006]

or from

http://www.epa.gov/nheerl/dsstox/ExternalPublicDatabases.html#ISSCAN [accessed 1 June 2006].

Case Study X: Japan

177. In order to prevent environmental pollution by chemical substances with persistent harmful properties that may affect human health and living organisms in the environment, chemical substances newly manufactured in or imported into Japan are subject to premarketed evaluation under the Law Concerning the Evaluation of Chemical Substances and Regulation of their Manufacture (Chemical Substances Control Law: CSCL).

178. Biodegradation and bioconcentration are important items of environmental fates in this evaluation scheme. The Ministry of Economy, Trade and Industry (METI) is conducting examination of biodegradability and bioconcentration of existing chemicals, and data for more than 1400 existing chemicals are available. As for new chemical substances, approximately 3000 data have been obtained. Making use of these data and related empirical knowledge, METI has been developing two SAR models for biodegradation and bioconcentration in Comprehensive Chemical Substances Risk Assessment and Management Program to evaluate chemical substances effectively and efficiently. In addition, a hydrolysis prediction model based on quantum chemistry calculation has been recently started to develop for supporting to biodegradability prediction models.

179. In April 2004, National Institute of Technology and Evaluation (NITE), an incorporated administrative agency under METI, supporting CSCL examination, has established an in-house committee for evaluating the approach to use (Q)SARs under CSCL. At present, the committee is making its efforts on the screening and prioritization of untested existing chemicals by using three biodegradability prediction models (METI model, CATABOL and BIOWIN5), one bio-concentration prediction model (BCFWIN) and one LogP prediction model (CLOGP). These five models were validated based on OECD principles for (Q)SAR validation. The validation reports for these models and the advisory list showing the prediction for existing chemicals by these models will be disclosed at NITE's web site.

180. The Ministry of Health, Labour and Welfare (MHLW) has been taking responsibility for the human health part of the Chemical Substances Control Law. Data have been generated on, mutagenicity (Ames test and chromosomal aberration test), 28days repeated dose toxicity for 275 chemicals. National Institute of Health Science (NIHS) is studying application of (Q)SAR software (e.g. DEREK, MultiCASE, and ADMEWorks) for evaluating chemical genotoxicity by comparing the results of (Q)SARs with the results of the Salmonella/microsome assay.

181. The Ministry of the Environment (MoE) is in charge of evaluation of adverse effects of chemicals on ecosystems which are determined by ecotoxicity tests using specific organisms such as fishes, daphnia and algae. MoE has evaluated ecotoxicity of approximately 440 existing chemicals so far. As for applying (Q)SAR model to the ecotoxicity evaluation, National Institute for Environmental Studies (NIES) has conducted relevant studies including development and verification of a (Q)SAR program.

182. (Q)SAR is considered to be one of the efficient measures to sort chemicals with a specific biological activity from the enormous number of chemicals. Accordingly, (Q)SAR method, which evaluates the hormonal activity of chemicals in terms of endocrine disruption, is defined as a tool for obtaining information on *in vitro* mechanistic data in the OECD conceptual framework for the testing of endocrine disrupting chemicals. When (Q)SAR methods are fully established, certain *in vitro* and *in vivo* testing may be replaced and relevant chemical management program of endocrine disrupting chemicals will be further facilitated. For screening of the chemicals having hormonal effects, MHLW and METI have been developing 3-D (Q)SAR models using *in vitro* data of more than 900 chemicals.

Case Study XI: United Kingdom

183. The UK Health and Safety Executive (HSE) is a governmental regulatory authority involved in national and international regulatory programmes, most notably the EU New Substances scheme, the EU Existing Substances Regulation (ESR), the EU Classification and Labelling scheme (C&L) and the OECD SIDS/HPV programme, in relation to the hazard identification and risk assessment of the human health effects of chemicals. (Q)SARs are an important part of the broader discipline of "read-across", rather than as a stand-alone technique which can be used in complete isolation to underpin any regulatory decisions. For the purposes of this paper read-across is defined as the process of using the totality of data on one chemical or group of chemicals (physicochemical properties, toxicology profile, structural or chemical analogy etc) to predict the properties of another related chemical or group of chemicals for which data are not available or the available data are unreliable. The UK has indicated that in general, it is unlikely that (Q)SAR predictions alone would be used to underpin any regulatory position (although it should be noted that in the EU classification and labelling scheme an isocyanate is regarded as a respiratory sensitiser unless there is evidence to the contrary). However, read-across has been used, and will increasingly be used, as a means to fill data gaps, exonerate industry from testing requirements and highlight chemicals of particular concern.

184. At the moment read-across is used to a limited extent in EU chemical programmes, mostly in the New Substances scheme and ESR, and generally on a case by case basis. There is little, if any, formal use of (Q)SAR systems in support of regulatory decisions relating to health effects at present. The EU Technical Guidance Document (TGD), which aims to ensure consistency in hazard identification and risk assessment within the new and existing substances programmes, alludes to the use of read-across and (Q)SAR but doesn't present any general strategy or criteria for their use in relation to human health endpoints.

185. One of the more notable uses of read-across has been in the EU New Substances scheme where it has been used to support exonerations from testing. For all new industrial chemicals supplied in the EU above 1 tonne/annum, a 'base-set' of good quality toxicological data must be provided, as specified in the regulations, on acute and repeat-dose toxicity, irritancy, skin sensitisation and mutagenicity. Over the last few years, in relation to human health, the UK has been active in developing a strategy for reading across from such data to exonerate from testing or reduce the testing requirements for new structurally related substances or groups of substances. One of the drivers for this work has been our commitment to animal welfare issues and the reduction of animals used in testing. This work has highlighted some important principles such as the value of physicochemical properties and acute oral toxicity and mutagenicity data to confirm the validity of the read-across between two chemicals. It has also shown the potential of the 'group' approach, where full testing is only required for chemicals at each end of a series of related chemicals. The DEREK system has been used on an ad-hoc basis to identify possible alerts and positive predictions have been used to inform on the requirements for testing. We have taken a similar approach in the OECD HPV programme. We have not utilised computerised Quantitative SAR technology in our read-across approach in any programme.

186. Under the new EU chemicals policy ("REACH"), currently being developed, all chemicals must be registered on a central database, and a standard set of basic hazard information will be required on each chemical. We envisage that mainly *in vitro* data, probably including (Q)SAR predictions, will be required for chemicals supplied at less than 10 tonne/annum. At higher tonnage levels additional data will be required. This large data-requirement is likely to create an impetus to promote read-across/(Q)SAR as a way of generating the required data and also may provide an effective way of rapidly identifying substances of high concern, such as carcinogens, mutagens and reproductive toxins, for priority regulatory

attention. Given these considerations, the use of read-across/(Q)SAR is likely to become increasingly important in the EU in the future.

187. The UK Department for Environment, Food, and Rural Affairs (DEFRA) has established a Chemical Stakeholders Forum (CSF) (DEFRA, 2001). The CSF has established criteria for identifying chemicals of concern based on specific Persistence, Bioaccumulation and Toxicity values. The approach has been endorsed by DEFRA's Advisory Committee on Hazardous Substances (ACHS, 2001) who also have agreed that appropriate (Q)SAR models can be used to fill data gaps. The Environment Agency has carried out an initial screen of the International Uniform Chemicals Information Database (IUCLID) database for substances that fulfil the CSF PBT criteria and reported the initial findings from this study to the ACHS in September 2001 (ACHS, 2001).

Case Study XII: European Commission

The REACH proposal

188. In 2003, the European Commission (CEC 2001a) published a White Paper setting out a 'Strategy for a Future Chemicals Policy' which proposed a new system, called REACH (Registration, Evaluation and Assessment of Chemicals) for managing new and existing chemicals in a single regulatory framework. The 30,000 existing substances affected will be processed on a phased basis over a period of 11 years (expected from 2007 to 2018) starting with those manufactured or imported in the highest volumes (>1000 metric tonnes p.a), as well as those of very high concern (e.g. CMR substances). An important part of this policy is the fostering of research on the development and validation of alternative (non-animal) methods, including (Q)SAR models and in vitro test methods. The White Paper was subsequently discussed by the European Parliament (2001a, b) who requested 'the use of screening procedures based on simplified risk assessment using data modelling, e.g., quantitative structure activity relationships ((Q)SARs) and use patterns to prioritise substances of possible concern …in order to speed up risk assessments…'. Following an extensive political and stakeholder consultation, the Commission's initial proposal for REACH was adopted in October 2003 (CEC, 2003). During 2006, the REACH proposal is undergoing the co-decision procedure between the Council of Ministers and the European Parliament. Further information can be found at:

http://europa.eu.int/comm/enterprise/reach/index_en.htm [accessed 1 June 2006].

189. Under the draft REACH regulation, it is foreseen that non-testing methods, including (Q)SARs, read-across and chemical category approaches will be used more extensively and more systematically than under previous EU legislation on chemicals. Of particular importance for (Q)SAR applications is Annex IX in the draft regulation, which outlines the general rules for using (Q)SARs and grouping approaches (read-across approaches and chemical categories) as a means of adapting the standard testing requirements, which are largely tonnage-dependent and specified in Annexes V to VIII.

190. The regulatory use of these non-testing methods is based on the premise that structurally similar chemicals will have similar physical attributes and biological (including toxicological) effects. The concept of similarity can be used in hazard and risk assessment to estimate regulatory endpoints when test data are either inadequate or unavailable. Approaches for describing such relationships between similar chemicals include:

- SAR and (Q)SAR. A (Q)SAR consists of a relationship between the chemical structure, or physicochemical representations thereof, and the outcome in a laboratory measurement for a test endpoint (biological or other physicochemical. property). SARs are qualitative relationships in the form of structural alerts that incorporate molecular substructures or fragments related to the presence or absence of activity. QSARs are quantitative models based on numerical descriptors of chemical structure and/or physicochemical properties. Often, QSARs estimate the relative chemical activity of chemicals presumed to behave according to the same mechanisms.

- Analogues/read-across. The use of read across/nearest analogue analysis is a possible approach to obtain relevant data when there are no experimental studies on the compound of interest, and/or to evaluate the reliability of predicted estimates for a particular substance. Read across is the process by which one or more properties of a given chemical are inferred by comparison of that chemical with a chemical(s) of similar molecular structure(s) and physicochemical properties, for which the properties of interest are known. In principle, the read-across approach can be used to assess physicochemical properties, ecotoxicity, toxicity, and environmental fate.

- Qualitative read-across can be regarded as the application of SAR. Quantitative read-across involves the identification of a chemical substructure that is common to the two substances, and the assumption that the known value of a property for one substance can be used to estimate the unknown value of the same property for another substance.

- Chemical Categories. A chemical category is a group of chemicals whose physicochemical and toxicological properties are likely to be similar or follow a regular pattern as a result of structural similarity. These structural similarities may create a predictable pattern in any or all of the following parameters: physicochemical properties, environmental fate and environmental effects, and/or human health effects.

REACH-Implementation Projects

191. Within the context of the EU Chemicals Policy, the European Commission has initiated a number of REACH Implementation Projects (RIPs) with the intention of developing tools and guidance for the new legislation. The RIPs, numbered from 1-7, are coordinated in close collaboration with the main stakeholders, namely Member States, Industry and NGOs.

192. The overall aim of RIP 3 is to develop guidance documents and tools for industry, to facilitate a smooth implementation of the legislation. The regulatory use of (Q)SARs and other non-testing approaches is part of RIP 3.3, a sub-project focusing on developing guidance on Information Requirements on Intrinsic Properties of substances (or Integrated Testing Strategies).

193. A scoping study for RIP3.3 called TAPIR (Three point three – A Project for the Information Requirements of REACH) was undertaken between January and July 2005 (ECB, 2005). This was co-ordinated by CEFIC (the contractor) and carried out by experts from CEFIC, Danish EPA, Environment Agency (England and Wales), ECETOC, INERIS, KemI and TNO. The over-arching scientific objective of the project was to explore how certain types of information (i.e. in vitro, SAR, QSAR, read-across, categories) could help to fill data gaps, when adequate substance-specific information was not available, while keeping confidence in the assessment of the substance. Within the project, Information Working Groups (IWGs) reviewed the state of the science in three relevant areas: exposure considerations, testing methods and non-testing methods. In addition, four Endpoint Working Groups (EWGs) on aquatic toxicity, degradation, reproductive/developmental toxicity and irritation/corrosivity, developed endpoint-specific Integrated Testing Strategies (ITS), which were tested on a limited number of chemicals. A generic decision-making framework was also developed. The final reports of this scoping study are available from the ECB website (http://ecb.jrc.it/REACH/, accessed 1 June 2006).

194. One of the conclusions that arose from the first phase of RIP 3.3 was with respect to read across and chemical categories, in terms of how to carry out a read-across or build a category, and how to justify and report the read-across / category proposal. These issues have been taken up in the second phase of RIP 3.3, which is being coordinated by CEFIC and steered by a multi-stakeholder Project Management Group (PMG). Within the second phase of RIP 3.3, the ECB and the OECD secretariat are jointly coordinating an activity on the development of guidance for categories (including read-across).

EU Working Group on (Q)SARs

195. At the beginning of 2005, a (Q)SAR subgroup was established under the Technical Committee on New and Existing Chemicals (TC NES). The Working Group, which is coordinated by the ECB, has the remit to carry out, at a technical level, the following activities on (Q)SARs, read-across methods and chemical categories in preparation for the eventual use of these approaches under REACH:

1) to provide a forum for discussing the results of ongoing and future projects within the EU which aim at promoting the regulatory acceptance of (Q)SARs;
2) to develop guidance and criteria for establishing the regulatory acceptability of (Q)SAR models and their estimates. These criteria will eventually be used by industry when submitting (Q)SAR estimates to meet information requirements, and will be used by the Competent Authorities to make case-by-case decisions on the acceptability of (Q)SAR model estimates for specific substances;
3) to develop guidance for the formation of chemical categories, and criteria for assessing the applicability of chemical categories under REACH;
4) to provide advice on the needs for further development and validation of (Q)SAR models;
5) to provide advice on the development of a dynamic compendium of accepted (Q)SARs and chemical category proposals.

196. Under the umbrella of the QSAR Working Group, the Netherlands is leading the EU (Q)SAR Experience Project aimed at gaining experience with (and confidence in) the use of (Q)SAR predictions in the toxicological evaluation of chemical substances. The project, which started in 2005 consists of "eyes-on" and "hands-on" experience of using (Q)SAR models. In the first phase of the project, a set of 177 SIDS chemicals were used to evaluate the performance of various (Q)SAR models on three selected endpoints: acute toxicity to fish, biodegradability, and mutagenicity in the Ames test. The set of 177 SIDS chemicals were chosen because they had already been used in a previous comparative analysis by Denmark [ENV/JM/TG(2004)26/REV1].

(Q)SAR Activities of the Joint Research Centre

197. The European Commission's Joint Research Centre has a QSAR Group within the European Chemicals Bureau (ECB). The mission of the QSAR Group is to promote the regulatory use of valid (Q)SARs and related estimation methods. In addition to coordinating the activities of the EU QSAR Working Group, the ECB has initiated a wide range of in-house and collaborative activities in the area of non-testing approaches, including:

1) reviewing the use of read-across and chemical categories by the US EPA, Canadian authorities, OECD, and EU Member States
2) reviewing the status of (Q)SARs for endpoints of regulatory relevance (e.g. Lessigiarska et al, 2005)
3) development and characterisation of (Q)SARs in the areas of topical toxicity (skin and eye irritation), skin sensitisation, skin penetration, acute fish toxicity, steroid hormone-receptor binding, persistence, bioaccumulation, mutagenicity and carcinogenicity (e.g. Gallegos Saliner et al, 2006; Pavan et al 2005a, 2005b, 2006; Netzeva et al, 2006).
4) research on methods for assessing chemical similarity (Gallegos, 2005)
5) research on the applications of neural network approaches, e.g. to the grouping of chemicals
6) research on chemical ranking methods
7) contribution to the development of guidance on the validation of (Q)SARs (Worth et al, 2005; Netzeva et al 2005)
8) contribution to the development of guidance on the use of batteries of (Q)SARs
9) contribution to the development of standardised formats for reporting QSARs and their estimates, and controlled vocabularies for reporting, storing and retrieving chemical information

10) contribution to the development of the (Q)SAR Application Toolbox, under the OECD
11) implementation of an internet-based version of the Danish (Q)SAR database
12) development of a (Q)SAR Inventory
13) contribution to the EU (Q)SAR Experience Project
14) contribution to the development of integrated testing strategies for regulatory endpoints under REACH
15) contribution to the development of guidance on grouping approaches (read-across and categories)
16) contribution to the development of guidance on priority setting under REACH
17) development of training materials and organisation of training courses on (Q)SARs
18) assessments of the possible impacts of using non-testing methods on costs and animal testing.

198. Some of the ECB (Q)SAR activities are carried out entirely in-house, others are carried out in small collaborative groups, whereas others are carried out in the context of REACH-Implementation Projects or under the umbrella of the OECD. In addition, the ECB QSAR Group provides an advisory service to the ECB and other Directorates General of the Commission (especially DGs Environment and Enterprise) by providing QSAR input to technical discussions on classification and labelling, risk assessment and PBT assessment. These technical discussions take place under the umbrella of Commission Working Groups such as the Technical Committee on New and Existing Substances (TC NES), the Technical Committee for Classification and Labelling (TC C&L), and the PBT Working Group. The duties of these working groups will eventually be taken over by the European Chemicals Agency and its committees, being established in Finland.

199. As an example of an activity carried out under the umbrella of the OECD, the ECB has taken a lead for the development of guidance on how to apply the OECD validation principles to (Q)SAR models that show promise in regulatory applications.

200. Further information on ECB (Q)SAR activities can be found at the following website: http://ecb.jrc.it/QSAR/, accessed 7 August 2006.

Projects Funded by DG Research of the European Commission

201. DG Research of the Commission (http://ec.europa.eu/research/index_en.cfm, accessed 1 June 2006) provides funding for collaborative research projects in the EU in the context of the Commission's Framework Programmes (e.g. the 6th Framework Programme from 2003-2006). Many of these projects have been focused on broad themes, involving the development, optimisation and validation of diverse technologies, whereas other projects have been specifically aimed at (Q)SARs.

202. As an example, the EU is funding a project on the Development of Environmental Modules for Evaluation of Toxicity of Pesticide Residues in Agriculture, DEMETRA, to develop software which will give a quantitative prediction of toxicity for pesticides, candidate pesticides and their metabolites. DEMETRA will finish in 2006, and will provide software on the web for the estimation of toxicity of chemicals to trout, daphnia, bees, and quail (www.demetra-tox.net, accessed 1 June 2006).

203. A four-year multi-partner integrated project on the development of integrated testing strategies for REACH, called OSIRIS ("Optimized Strategies for Risk assessment of chemicals based on Intelligent testing"), has been selected for funding by the European Commission. If the contract negotiations are successful, it is expected that this research project will start in the fourth quarter of 2006 or the first quarter of 2007.

OTHER OECD ACTIVITES IN THE AREA OF (Q)SAR

204. As a part of the OECD activities on hazard assessment, the OECD Workshop on Quantitative Structure-Activity Relationships (QSARs) in Aquatic Effects Assessment was held in Utrecht, the Netherlands, in 1990. The report from the workshop was published as OECD Environment Monograph No. 58 (OCDE/GD(92)168) in 1992. The workshop recommended the development of guidance on the use of (Q)SARs especially in particular classes of organic chemicals that act by narcosis, and for phenols and primary aromatic amines.

205. Use of (Q)SARs in predicting environmental fate properties was discussed at the OECD Workshop on the Use of Simple Models in Exposure Assessment, which was held in Berlin in 1991. The outcome from this workshop, as well as from further work was published as two OECD monographs: "Application of Structure-Activity Relationships to the Estimation of Properties Important in Exposure Assessment" (OECD Environment Monograph N0. 67, OCDE/GD(93)125) and "Structure-Activity Relationships for Biodegradation" (OECD Environment Monograph N0. 68, OCDE/GD(93)126).

206. These meetings established the basis for the use of (Q)SAR approaches in the Guidance Document for Aquatic Effects Assessment (OECD Environment Monograph No. 92, OCDE/GD(95)18), and other guidance documents that followed.

207. The US Environmental Protection Agency and the European Commission also undertook a joint project to evaluate (Q)SARs, and the outcome of this project was published from OECD in 1994 (OECD Environment Monograph No. 88, OCDE/GD(94)28). In this project, predictions for 144 substances from (Q)SAR models by the US EPA were compared with European pre-marketing laboratory data. The study covered physical/chemical properties (boiling point, vapour pressure, water solubility, water-octanol partition coefficient and biodegradation), ecotoxicity (acute toxicity to fish and daphnia) and human health effects (acute toxicity, skin and eye irritation, sensitisation, repeated dose toxicity and mutagenicity). The study identified promising areas for wider use of (Q)SAR predictions in chemicals regulation, such as biodegradation and acute toxicity to fish and daphnia, and also recommended further work for building (Q)SARs into a future battery of assessment approaches.

208. The OECD is facilitating the development of harmonised criteria for health and environmental hazards, as a part of the Globally Harmonised System for Hazard Classification and Communication (GHS), within the framework of the IOMC (Inter-organisation programme for the Sound Management of Chemicals). The GHS includes guidance on the use of (Q)SARs for human health and environmental classification endpoints. For many of the human health endpoints, SARs are referred to as a factor that can be considered in the absence of test data or human or animal experience. For hazards to the aquatic environment, more detailed guidance has been developed as an annex to the GHS document, for using validated (Q)SAR models for aquatic toxicity and partition co-efficient, where no experimental data are available.

209. The SIDS Manual (OECD 2002a) guidance on the use of SAR in the OECD SIDS program outlines the use of SAR/(Q)SAR in the HPV Challenge Program which is expected to decrease the number of new tests required to develop a SIDS for each HPV chemical. Their use, by either the category or individual chemical approach, will necessarily be limited by the nature of the SIDS endpoint, the amount and adequacy of the existing data, and the type of SAR/(Q)SAR analysis performed. Measured data

developed using acceptable methods are preferred over estimated values. The development and use of SAR/(Q)SAR in the Challenge Program will be different for each of the major categories of SIDS (i.e., physicochemical properties, environmental fate, ecotoxicity, and health effects). Finally, the US FDA, DK EPA, and Health Canada have used (Q)SAR model predictions for various purposes on a range of human health endpoints.

210. Hulzebos et al. (1999) noted that more validation is needed to correlate SAR with individual health endpoints. Given the complexity of health endpoints and the amount of uncertainty in many models, some regulators have historically used an expert judgement/ nearest analogue approach to SAR for predicting some endpoints when assessing new chemicals. It has been suggested that a similar approach could be applied in the Challenge Program, where the goal is to find toxicity data for an analogue that can be used to address the testing needs of an HPV chemical, on an endpoint-by-endpoint and case-by-case basis.

211. The guidance document for the development and use of chemical categories within the OECD HPV Chemicals Programme was revised during 2004 and 2005. There are no plans for further revisions in the immediate future, but given the rate of submission of chemical categories to the Programme, it is anticipated that, in the light of additional experience gained during use, a further update of the guidance document might become necessary within approximately two years. In the meantime, the OECD Secretariat is co-leading with the European Commission (ECB) a project aimed at the development of guidance on read-across and categories in the context of both the OECD Existing Chemicals Programme and the REACH-Implementation Projects. The current OECD guidance is taken as the starting point.

212. In parallel to the guidance document on the formation and use of chemical categories, the development of a (Q)SAR Application Toolbox is foreseen within the OECD work programme on (Q)SARs (see para 8-17). While the scope and use of the (Q)SAR Application Toolbox has not yet been finalised in the OECD, it is likely to contain a module aimed at facilitating the development of categories for organic chemicals with multiple functional groups. The Toolbox could also identify chemicals which might be placed in a particular category but which might have significant metabolic pathways from the other substances in that category. These new approaches will be discussed and evaluated over the next three years or so.

REFERENCES

ACHS (Advisory Committee on Hazardous Substances) (2001), Minutes of the Second Meeting of the Advisory Committee on Hazardous Substances, ACHS, held on 3 September 2001, DEFRA, Ashdown House, London

Allanou, R., B.G. Hansen and Y. van der Bilt (1999), *Public Availability of Data on EU High Production Volume Chemicals*, EUR 18996 EN, http://ecb.jrc.it/, accessed 1 June 2006.

Auer, C., J. Nabholz and K. Baetcke (1990), "Mode of Action and the Assessment of Chemical Hazards in the Presence of Limited Data: Use of Structure-Activity-Relationships (SAR) Under TSCA, Section 5", *Environmental Health Perspectives*, 87: 183-197.

Auer, C., *et al.* (1994), "SAR - The U.S. Regulatory Perspective"*, SAR and (Q)SAR in Environmental Research,* 2: 29-38.

Bailey, A.B., *et al.* (2005), "The Use of Structure-Activity Relationship Analysis in the Food Contact Notification Program", *Regulatory Toxicology and Pharmacology*, 42(2): 225-235.

Barratt, M.D. and D.A. Basketter (1994), "Structure-Activity Relationships for Skin Sensitization: An Expert System", in A. Rougier, A.M. Goldberg and H.I. Maibach (eds.), "In Vitro Skin Toxicology, Irritation, Phototoxicity, Sensitization"*, Alternative Methods in Toxicology*, Vol. 10, Mary Ann Liebert , Inc. Publishers, New York, pp. 293-301.

Benigni, R. (2005), "Structure-Activity Relationship Studies of Chemical Mutagens and Carcinogens: Mechanistic Investigations and Prediction Approaches", *Chem. Revs.*, 105, 1767-1800.

Benigni, R., *et al.* (1993), "Quantitave Structure-Activity Relationship Models Correctly Predict the Toxic and Aneuploidizing Properties of Six Halogenated Methanes in Aspergillus Nidulans", *Mutagenesis*, 8, 301-305.

Benigni, R. and C. Bossa (2006), "Structural Alerts of Mutagens and Carcinogens", *Curr. Comput.-Aid. Drug Des.*, in press.

Benigni, R., *et al.* (2005), "Simple and α–β -unsaturated Aldehydes: Correct Prediction of Genotoxic Activity through Structure-Activity Relationship Models", *Environ. Mol. Mutagen.*, 46, 268-280.

Benigni, R. and L. Passerini (2002) "Carcinogenicity of the Aromatic Amines: from Structure-Activity Relationships to Mechanisms of Action and Risk Assessment", *Mutat. Res. Revs.*, 511, 191-206.

Bergfeld, W.F. and F.A. Andersen (1998), "The Cosmetic Ingredient Review", *Cosmetic Regulation in a Competitive Environment,* Marcel Decker, Inc. pp. 1-21.

Bol, J., *et al.* (1993a), *Predictions of the Aquatic Toxicity of High-Production Volume Chemicals. Part A: Introduction and Methodology*, Research Institute of Toxicology, Utrecht University, Utrecht.

Publikatiereeks Stoffen, Veiligheid, Straling, 1993/9A, Ministerie VROM, Den Haag. http://ecb.jrc.it/, accessed 1 June 2006.

Bol, J., *et al.* (1993b), *Predictions of the Aquatic Toxicity of High-Production Volume Chemicals. Part B: Predictions*, Research Institute of Toxicology, Utrecht University, Utrecht. Publikatiereeks Stoffen, Veiligheid, Straling, 1993/9B, Ministerie VROM, Den Haag, http://ecb.jrc.it/, accessed 1 June 2006.

Bradbury, S.P., T.C.J. Feijtel and C.J. van Leeuwen (2004), "Meeting the Scientific Needs of Ecological Risk Assessment in a Regulatory Context", *Environmental Science and Technology,* 38(23): 463A-470A.

Bronaugh, R.L., and C.N. Barton (1993), "Prediction of Human Percutaneous Absorption with Physiochemical Data", in R.G.M. Wang , J.B. Knaak and H.I. Maibach (eds.), *Health Risk Assessment*, CRC Press, New York, pp. 117-126.

Burleigh-Flayer, H., (2000), *Comparison of Environmental Fate Predictions from the PBT Profiler with Data found in the Literature*, PPG Industries, Inc., Pittsburgh, PA. (1-page table), http://www.epa.gov/oppt/newchems/pubs/sustainable/profiler-predictions1.pdf, accessed 7 August 2006.

Cash, G. (1999), "Prediction of Physicochemical Properties from Euclidean Distance Methods based on Electrotopological State Indicies", *Chemosphere*, 39(14): 2583-2591.

CEC (Commission of the European Communities) (2001a), *White Paper on the Strategy for a Future Chemicals Policy*, COM(2001) 88 final.

CEC (2001b), *Joint Meeting of the Competent Authorities for the Implementation of Council Directive 67/548/EEC (New Substances) and Council Regulation (EEC) 793/93 (Existing Substances). An Interim Strategy for Management of PBT and VPVB Substances*, ENV/D/432048/01. NOTIOF/36/2001.

CEC (2003), *Proposal for a Regulation of the European Parliament and of the Council concerning the Registration, Evaluation, Authorization and Restriction on Chemicals (REACH), establishing a European Chemicals Agency and amending Directive 1999/45/EC and Regulation (EC) {on Persistent Organic Pollutants},* COM(2003) 644 final / *Proposal for a Directive of the European Parliament and of the Council amending Council Directive 67/548/EEC in order to adapt it to Regulation (EC) of the European Parliament and of the Council concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals,* Brussels, 29 October 2003, Brussels, Belgium.

CEFIC Long Range Research Inititaive (LRI) (2002), *(Q)SARs for Human Health and the Environment*, Report from the Workshop on Regulatory Acceptance held in Setubal, Portugal, 4-6 March 2002, 58 pp.

Chun, J., J.V. Nabholz and M.J. Wilson (2000), *Comparison of Measured Aquatic Toxicity Data with EPA, OPPT SAR Predictions*, PPG Industries, Inc., Pittsburgh, PA and U.S. EPA, OPPT, Washington, DC., 6 pages, http://www.epa.gov/oppt/newchems/pubs/sustainable/ppg-sar-study-1999-2000.pdf, accessed 7 August 2006.

Clements, R.G. (ed.) (1988), *Estimating Toxicity of Industrial Chemicals to Aquatic Organisms Using Structure Activity Relationships*, U.S. EPA, Office of Pollution Prevention and Toxics (OPPT),

Washington, D.C. EPA-560-6-88-001. 286 pp. Available from U.S. EPA, National Service Center for Environmental Publications, tel.: 1-800-490-9198.

Clements, R.G., *et al.* (1995), "The Application of Structure-Activity Relationships (SARs) in the Aquatic Toxicity Evalutation of Discrete Organic Chemicals", *SAR & QSAR in Environmental Research,* 3(3): 203-215.

Contrera, J.F., *et al.* (2005), "QSAR Modeling of Carcinogenic Risk Using Discriminant Analysis and Topological Molecular Descriptors", *Current Drug Discovery Technologies,* 2: 55-57.

Contrera, J.F., E.J. Matthews and R.D. Benz (2003), "Predicting the Carcinogenic Potential of Pharmacueticals in Rodents Using Molecular Structural Similarity and E-State Indices", *Regulatory Toxicology and Pharmacology,* 38: 243-259.

Contrera, J.F., *et al* (2004), "Estimating the Safe Starting Dose in Phase I Clinical Trials and No Observed Effect Level Based on QSAR Modeling of the Human Maximum Recommended Daily Dose", *Regulatory Toxicology and Pharmacology,* 40: 185-206.

Cronin, M.T.D., *et al.* (2003a), "Use of (Q)SARs in International Decision-Making Frameworks to Predict Ecologic Effects and Environmental Fate of Chemical Substances", *Environmental Health Perspectives,* 111(10): 1376-1390.

Cronin, M.T.D., *et al.* (2003b), "Use of (Q)SARs in International Decision-Making Frameworks to Predict Health Effects of Chemical Substances", *Environmental Health Perspectives*, 111(10): 1391-1401.

Danish EPA (Environmental Protection Agency) (2001), *Draft Advisory List for Self-Classification of Dangerous Substances*, Environmental Project No. 636, http://www.mst.dk/chemi/01050000.htm, accessed 1 June 2006.

Dearden, J.C. (2003), "In Silico Prediction of Drug Toxicity", *Journal of Computer-Aided Molecular.Design.,*17(2-4): 119-27.

DEFRA (Department for Environment, Food, and Rural Affairs) (2001), *UK Chemical Stakeholders Forum. First Annual Report, 2000 - 2001,* November, 2001, DEFRA, London, ISBN 1 85112 700 3.

ECB (European Chemicals Bureau) (2005), *Scoping Study on the Development of a Technical Guidance Document on Information Requirements on Intrinsic Properties of Substances (RIP 3.3-1)*, Report prepared by CEFIC, Danish EPA, Environmental Agency of Wales and England, ECETOC, INERIS, KemI and TNO.
European Commission, Joint Research Centre, Ispra, Italy, http://ecb.jrc.it/, accessed 1 June 2006.

ECOSAR (1996). *Technical Reference Manual* [for Estimating Toxicity of Industrial Chemicals to Aquatic Organisms Using Structure-Activity Relationships, R.G. Clements (ed.)]. U.S. EPA, OPPT, Washington, D.C. 413 pp.
http://www.epa.gov/oppt/newchems/tools/sarman.pdf, accessed 7 August 2006.

Environment Canada (2002), *Existing Substances Evaluation,* Ottawa, ON: Environment Canada, http://www.ec.gc.ca/substances/ese/eng/esehome.cfm#autop, accessed 1 June 2006.

Eriksson, L., E. Johansson and S. Wold (1997), "Quantitative Structure-Activity Relationship Model Validation", in F. Chen and G. Schuurmann (eds.), *Quantitative Structure-Activity Relationships in Environmental Sciences - 7,* SETAC: Pensacola, FL, USA, pp. 381-397.

Eriksson, L., *et al.* (2003), "Methods for Reliability and Uncertainty Assessment and for Applicability Evaluation of Classification- and Regression-Based (Q)SARs", *Environmental Health Perspectives,* 111(10): 1361-1375.

European Commission (1996), *Technical Guidance Document 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*, European Commission, Office for Official Publications of the European Communities, Luxembourg, http://ecb.jrc.it/, accessed 1 June 2006.

European Commission (2000), *Towards the Establishment of a Priority List of Substances for Further Evaluation of Their Role in Endocrine Disruption*, Document Number M0355008/1786Q/10/11/00, http://ec.europa.eu/environment/docum/01262_en.htm, accessed 7 August 2006.

European Commission (2001), *An Interim Strategy for Management of PBT and vPvB Substances, Joint Meeting of the Competent Authorities for the Implementation of Council Directive 67/548/EEC (New Substances) and Council Regulation (EEC) 793/93 (Existing Substances)*, ENV/D/432028/01. NOTIO/36/2001, DG Environment, Brussels.

European Commission (2003), *Technical Guidance Documents on Risk Assessment in Support of Commission Directive 93/67/EEC on Risk Assessment for New Notified Substances, Commission Regulation (EC) No. 1488/94 on Risk Assessment for Existing Substances, and Directive 98/8/EC of the European Parliament and of the Council concerning the Placing of Biocidal Products on the Market*, EUR report 20418 EN, European Chemicals Bureau, Ispra, Italy, http://ecb.jrc.it/, accessed 1 June 2006.

European Parliament (2001a), *Compromise Amendments 1-23. Draft Report on the Commission White Paper on Strategy for a Future Chemicals Policy*, PE 304.693.

European Parliament (2001b), *Draft Report by Inger Schorling on the Commission White Paper on Strategy for a Future Chemicals Policy,* PE 304.693/COMP. 1-23.

Fabjan, E., E. Hulzebos and W. Mennes (2005), "Category Approach for Reproductive Toxic Phthalates", submitted.

Gallegos, A. (2005), *A Similarity Based Approach for Chemical Category Classification*, EUR 21867 EN. Joint Research Centre, European Commission, Ispra, Italy, http://ecb.jrc.it/, accessed 1 June 2006.

Gallegos Saliner, A., T.I. Netzeva and A.P. Worth (2006), "Prediction of Estrogenicity: Validation of a Classification Model", *SAR and QSAR for Environmental Research,* in press.

Gamberger, D., *et al.* (1996), "Application of Artificial Intelligence in Biodegradation Modeling", in W.J.G.M. Peijnenburg and J. Damborsky (eds.), *Biodegradability Prediction, Vol. 23,* NATO ASI Series 2, Kluwer, Dordrecht, The Netherlands, pp. 41-50

Gerner, I., *et al.* (2000), "Development of a Decision Support System for the Introduction of Alternative Methods into Local Irritancy/Corrosivity Testing Strategies. Development of a Relational Database", *Alternatives to Laboratory Animals,* 28, 11-28.

Gerner, I., *et al.* (2004a) "Development and Pre-validation of a List of SAR Rules to be Used in Expert Systems for the Prediction of Skin Sensitising Properties of Chemicals", *Alternatives to Laboratory Animals,* 32, 487-509.

Gerner, I., *et al.* (2004b), "Use of Physicochemical Property Limits to Develop Rules for Identifying Chemical Substances with No Skin Irritation or Corrosion Potential", *QSAR and Combinatorial Science*, 23, 726-733.

Gerner, I., *et al.* (2004c), "Regulatory Use of (Q)SARs in Toxicological Hazard Assessment Strategies", *SAR and QSAR in Environmental Research*, 15, 359-366.

Gerner, I., M. Liebsch and H. Spielmann (2005), "Assessment of the Eye Irritating Properties of Chemicals by Applying Alternatives to the Draize Rabbit Eye Test: the Use of QSARs and In Vitro Tests for the Classification of Eye Irritation", *Alternatives to Laboratory Animals*, 33, 215-237.

GESAMP (Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection) (2002), *The Revised GESAMP Hazard Evaluation Procedure for Chemical Substances Carried by Ships*, GESAMP Reports and Studies, No. 64, 126 pp., IMO, London.

Gottschalck, T.E. and G.N. McEwen (eds.) (2004), *CTFA International Cosmetic Ingredient Dictionary and Handbook*, 10th Edition, The Cosmetic, Toiletry, and Fragrance Association, Washington, DC.

Hansen, B. and C.J. van Leeuwen (1995), "US-EU-Japanese Collaborative Project on Fate and Effects Predictions for High Productive Volume Chemicals. An Overview", report to the OECD Hazard Assessment Advisory Body, Ispra, Italy and The Hague, the Netherlands.

Hansen, B.G., *et al.* (1999), "Priority Setting for Existing Chemicals: European Union Risk Ranking Method", *Environ. Toxicol. Chem*., 18, 772-779.

Hoefer, T., *et al.* (2004), "Animal Testing and Alternative Approaches for the Human Health Risk Assessment under the Proposed New European Chemicals Regulation", *Archives of Toxicology,* 78, 549-564.

Hulzebos, E., *et al.* (2005) "Validity and Validation of Expert (Q)SAR System", *SAR and (Q)SAR in Environmental Research*, 16 (4): 385-401.

Hulzebos, E, *et al.* (2005), "Use of Structural Alerts to Develop Rules for Identifying Chemical Substances with Skin Irritation or Skin Corrosion Potential", *QSAR and Combinatorial Science,* 24, 332-342.

Hulzebos, E.M. and R. Posthumus (2003), "(Q)SARs: Gatekeepers against risk on chemicals?", *SAR and QSAR in Environmental Research*, 14, 285-316.

Hulzebos, E.M., L. Maslankiewicz and J.D. Walker (2003), "Verification of Literature-derived SARs for Skin Irritation and Corrosion", *QSAR Comb. Sci.,* 22, 351-363.

Hulzebos, E.M., P.C.J.I. Schielen and L. Wijkhuizen-Maslankiewicz (1999), *(Q)SARs for Human Toxicological Endpoints: A Literature Search,* report by the RIVM (Research for Man and Environment), The Netherlands, RIVM Report 601516.001

Jaworska, J.S., *et al.* (2003), "Summary of a Workshop on Regulatory Acceptance of (Q)SARs for Human Health and Environmental Endpoints", *Environmental Health Perspective*, 111/10, 1358-1360

Karcher, W., *et al.* (1995), "Predictions for Existing Chemicals- A Multilateral QSAR Project", *SAR, QSAR Environ. Res*., 3, 217-221.

Klopman, G., *et al.* (2003), "In-Silico Screening of High Production Volume Chemicals for Mutagenicity Using the MCASE (Q)SAR Expert System", *SAR and (Q)SAR in Environmental Research,* 14(2): 165-180.

Lange, A.W. and K. Vormann (1995), "Experiences with the Application of (Q)SAR in the Routine of the Notification Procedure", *SAR and (Q)SAR in Environmental Research,* 3: 171-177.

Langenberg, J.H., W.J.G.M. Peijnenburg and E. Rorije (1996), "On the Usefulness and Reliability of Existing QSBRs for Risk Assessment and Priority Setting", *SAR QSAR Environ. Res.*, 5, 1-16.

Lessigiarska I., A.P. Worth and T.I. Netzeva (2005), *Comparative Review of QSARs for Acute Toxicity*, JRC report EUR 21559 EN. Joint Research Centre, European Commission, Ispra, Italy, http://ecb.jrc.it/, accessed 1 June 2006.

Lipnick, R.L., *et al.* (1985), "Comparison of Fish Toxicity Screening Data for 55 Alcohols with the Quantitative Structure Activity Relationship Predictions of Minimum Toxicity for Non-Reactive Nonelectrolyte Organic Compounds", *Environmental Toxicology and Chemistry,* 4: 281-296.

Loonen, H., *et al.* (1999), "Prediction of Biodegradability from Chemical Structure: Modeling of Ready Biodegradation Test Data", *Environ. Toxicol. Chem.*, 18, 1763-1768.

Loonen, H., *et al.* (1996), "Use of (Q)SARs for Classification and Labelling of Substances as Dangerous for the Environment", presentation at (Q)SAR 96. 7th. International Workshop on (Q)SARs in Environmental Sciences. Helsingør, Denmark, 24-28 June 1996.

Loonen, H., *et al.* (1997), "Application of (Q)SAR in Hazard and Risk Assessment", presentation at Seventh Annual Meeting of SETAC-Europe, Amsterdam, the Netherlands, 6-10 April 1997.

MacDonald, D., *et al.* (2002), "Uses and Limitations of Quantitative Structure-Activity Relationships ((Q)SARs) to Categorize Substances on the Canadian Domestic Substance List as Persistent and/or Bioaccumulative, and Inherently Toxic to Non-Human Organisms", *SAR and (Q)SAR in Environmental Research,* 13: 43-55.

Maślankiewicz, L. *et al.* (2005), *Can Chemical Structure Predict Reproductive Toxicity?*, RIVM report 601200 005, Bilthoven, The Netherlands.

Matthews, E.J. and J.F. Contrera (1998), "A New Highly Specific Method for Predicting the Carcinogenic Potential of Pharmacueticals in Rodents Using MCASE QSAE-ES Software", *Regulatory Toxicology and Pharmacology*, 28: 242-264.

Matthews, E.J., R.D. Benz and J.F. Contrera (2000), "The Use of Toxicological Information in Drug Design and Application of Informatics, e-TOX, and e-ADME for Regulatory Decision Support at the FDA", *Journal of Molecular Graphics and Modeling,* 18: 605-613.

Matthews, E.J., *et al.* (2004), "Assessment of Health Effects of Chemicals in Humans. (Q)SAR Estimation of the Maximum Recommended Therapeutic Dose (MRTD) and No Effect Level (NOEL) of Organic Chemicals Based on Clinical Trial Data", *Current Drug Discovery Technologies,* 1: 61-76.

Matthews, E.J., *et al.* (2006a), "An Analysis of Genetic Toxicity, Reproductive and Developmental Toxicity, and Carcinogenicity Data: I. Identification of Carcinogens Using Surrogate Endpoints", *Regulatory Toxicology and Pharmacology*, 44(2): 83-96.

Matthews, E.J., *et al.* (2006b), "An Analysis of Genetic Toxicity, Reproductive and Developmental Toxicity, and Carcinogenicity Data: II. Identification of Genotoxicants and Reprotoxicants Using In Silico Methods", *Regulatory Toxicology and Pharmacology*, 44(2): 97-110.

McGee, P. (2005), "Modeling Success with In Silico Tools", *Current Opinions in Drug Discovery and Development,* 8(4): 24-28.

Nabholz, J.V. (2001), "Ecological Hazard/Risk Assessment and Risk Management of Chemicals under TSCA in USA", invited presentation at the International Symposium on Chemical Management for the Protection of Ecosystems on 9 November 2001 in Tokyo, sponsored by the Ministry of the Environment, Government of Japan.

National Research Council (1984), *Toxicity Testing. Strategies to Determine Needs and Priorities*, National Academic press, Washington, DC.

Netzeva, T.I., *et al* (2005), "Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure-Activity Relationships. The Report and Recommendations of ECVAM Workshop 52", *Alternatives to Laboratory Animals,* 33, 155−173.

Netzeva T.I., A. Gallegos Saliner and A.P. Worth (2006), "Comparison of the Applicability Domain of a QSAR for Estrogenicity with a Large Chemical Inventory", *Environmental Toxicology & Chemistry*, 25 (5): 1223-1230

OECD (1992a), *Report of the OECD Workshop on Quantitative Structure Activity Relationships (QSARs) in Aquatic Effects Assessment*, OECD Environment Monographs No. 58, OECD, Paris, France, 100 pp.
http://www.oecd.org/document/30/0,2340,en_2649_34365_1916638_1_1_1_1,00.html, accessed 1 June 2006.

OECD (1992b), *Report of the OECD Workshop on the Extrapolation of Laboratory Data to the Real Environment*, OECD Environment Monographs No. 59, OECD, Paris, France, 45 pp.
http://www.oecd.org/document/30/0,2340,en_2649_34365_1916638_1_1_1_1,00.html, accessed 1 June 2006.

OECD (1993a), *Report of the OECD Workshop on Application of SARs to the Estimation of Properties Important in Exposure Assessment*, OECD Environment Monographs No. 67, OECD, Paris, France, 67 pp.
http://www.oecd.org/document/30/0,2340,en_2649_34365_1916638_1_1_1_1,00.html, accessed 1 June 2006.

OECD (1993b), *Structure-Activity Relationships for Biodegradation*, OECD Environment Monographs No. 68, OECD, Paris, France, 105 pp.
http://www.oecd.org/document/30/0,2340,en_2649_34365_1916638_1_1_1_1,00.html, accessed 1 June 2006.

OECD (1994), *US EPA/EC Joint Project on the Evaluation of (Quantitative) Structure Activity Relationships*, OECD Environment Monographs No. 88, OECD, Paris, France, 367 pp.
http://www.oecd.org/document/30/0,2340,en_2649_34365_1916638_1_1_1_1,00.html, accessed 1 June 2006.

OECD (1995), *Guidance Document for Aquatic Effects Assessment*, OECD Environment Monographs No. 92, OECD Environment Health and Safety (EHS) Publications, EHS Series on Testing and Assessment, No. 3, OECD, Paris, France, 116 pp. http://www.oecd.org/document/30/0,2340,en_2649_34365_1916638_1_1_1_1,00.html, accessed 1 June 2006.

OECD (2000), "Chapter 3. Guidance for the Use of Structure-Activity Relationships (SAR) in the HPV Chemicals Programme", 22 pp in *Manual for Investigation of HPV Chemicals*, OECD, Paris, France, http://www.oecd.org/dataoecd/60/24/1947517.pdf, accessed 1 June 2006.

OECD (2001), "Chapter 6. Use of (Q)SAR", pp. 89-96, in *Guidance Document on the Use of the Harmonized System for the Classification of Chemicals Which Are Hazardous for the Aquatic Environment*, OECD Environment, Health and Safety Publications, EHS Series on Testing and Assessment, No. 27, OECD, Paris, France, 115 pp. http://www.oecd.org/document/30/0,2340,en_2649_34365_1916638_1_1_1_1,00.html, accessed 1 June 2006.

OECD (2002), *OECD's Database on Chemical Risk Assessment Models,* OECD, Paris, France, http://webdomino1.oecd.org/comnet/env/models.nsf, accessed 1 June 2006.

OECD (2004a), *The Report From the Expert Group on (Quantitative) Structure-Activity Relationships [(Q)SARs] on the Principles for the Validation of (Q)SARs*, OECD Environment, Health and Safety Publications, EHS Series on Testing and Assessment, No. 49, OECD, Paris, France, 206 pp. http://www.oecd.org/document/30/0,2340,en_2649_34365_1916638_1_1_1_1,00.html, accessed 1 June 2006.

OECD (2004b), "The Principles for Establishing the Status of Development and Validation of (Quantitative) Structure-Activity Relationships [(Q)SARs]", meeting document for the second meeting of the Expert Group on (Q)SARs, 21-21 September 2004, OECD, Paris, France. [ENV/JM/TG(2004)27; 8 July 2004; 17 pp.].

OECD (2004c), "Comparison of SIDS Test Data with (Q)SAR Predictions for Acute Aquatic Toxicity, Biodegradability and Mutagenicity on Organic Chemicals Discussed at SIAM 11-18", meeting document for the second meeting of the Expert Group on (Q)SARs, 21-21 September 2004, OECD, Paris, France. [ENV/JM/TG(2004)26/REV1; 22 October 2004; 35 pp.].

OECD (2004d), "Regulatory Application of (Q)SARs: a U.S. EPA Case Study", meeting document for the second meeting of the Expert Group on (Q)SARs, 21-21 September 2004, OECD, Paris, France. [ENV/JM/TG(2004)25/REV2; 10 September 2004; 14 pp.]

OECD (2004e), *Manual for Investigation of HPV Chemicals*, OECD, Paris, France, http://www.oecd.org/document/7/0,2340,en_2649_34379_1947463_1_1_1_1,00.html, accessed 1 June 2006.

OSPAR Commission (2000), Briefing Document on the Work of DYNAMEC and the DYNAMEC Mechanism for the Selection and Prioritisation of Hazardous Substances.

Pavan, M., A.P. Worth and T.I. Netzeva (2005a), *Preliminary Analysis of an Aquatic Toxicity Dataset and Assessment of QSAR Models for Narcosis*, EUR 21749 EN. Joint Research Centre, European Commission, Ispra, Italy, http://ecb.jrc.it/, accessed 1 June 2006.

Pavan, M., A.P. Worth and T.I. Netzeva (2005b), *Comparative Assessment of QSAR Models for Aquatic Toxicity*, JRC Report EUR 21750 EN. Joint Research Centre, European Commission, Ispra, Italy, http://ecb.jrc.it/, accessed 1 June 2006.

Pavan, M., T.I. Netzeva and A.P. Worth (2006), "Validation and Applicability Domain of a QSAR Model for Acute Toxicity", *SAR and QSAR for Environmental Research, in press.*

Pedersen, F., *et al.* (2003), *Assessment of Additional Testing Needs Under REACH: Effects of (Q)SARS, Risk Based Testing and Voluntary Industry Initiatives*, European Commission, Joint Research Centre, Institute for Health and Consumer Protection, Ispra, Italy. Report EUR 20863 EN. 36 pp., http://ecb.jrc.it/, accessed 1 June 2006.

PriceWaterhouseCoopers (2001), *An Industrial Revolution in R&D. Pharma 2005*, 22 pp. Find update at http://www.pwcglobal.com/gx/eng/about/ind/pharma/industrial_revolution.pdf, accessed 1 June 2006.

Posthumus, R., *et al.* (2005), "Validation of the BIOWIN Models with 110 EU New Notified Chemicals", *SAR and QSAR in Environmental Research*.

Rippeon, J. (2004), *PBT Profiler Case Study: Eaton Aeroquip Inc.'s Experience in the Use and Application of the PBT Profiler for Predicting Persistence, Bioconcentration, and Toxicity of Chemical Substances*, Eaton Aeroquip Inc., Williamsport, MD. 2 pages, http://www.epa.gov/oppt/newchems/pubs/sustainable/eaton-aeroquip-study-2004.pdf, accessed 7 August 2006.

Rorije, E., *et al.* (1999), "Evaluation and Application of Models for the Prediction of Ready Biodegradability in the MITI-Test", *Chemosphere*, 38, 1409-1417.

Rorije, E. and E.M. Hulzebos (2005), *Evaluation of (Q)SARs for the Prediction of Skin Irritation/Corrosion Potential*, National Institute of Public Health and Environment (RIVM-SEC)/European Chemicals Bureau (ECB), Bilthoven, The Netherlands, http://ecb.jrc.it/QSAR/, accessed 1 June 2006.

Ruppel-Kerr, R. (2003), *PBT Profiler Case Study: Bayer's Experience in the Use and Application of the PBT Profiler for Predicting Persistence, Bioconcentration and Aquatic Toxicity of Chemical Substances*, Bayer Corporation, Product Safety and Regulatory Affairs, Pittsburgh, PA, 2 pages, http://www.epa.gov/oppt/newchems/pubs/sustainable/bayer-study-2003.pdf, accessed 7 August 2006.

Schmieder, P.K., *et al.* (2002), "Quantitative Structure-Activity Relationship Models for Prediction of Estrogen Receptor Binding Affinity of Structurally Diverse Chemicals", *Environmental Toxicology and Chemistry*, 22: 1844-1854.

Stanski, D. (2001), "Modeling and Simulation of Clinical Trials in Drug Development", 2001 FDA Science Forum, Breakout Session B, http://vm.cfsan.fda.gov/~frf/forum01/absstan.htm, accessed 1 June 2006.

Tong, W., *et al.* (1998), "Evaluation of Quantitative Structure-Activity Relationship Methods for Large-scale Prediction of Chemicals Binding to the Estrogen Receptor", *Journal of Chemical Information and Computer Science*, 38(4): 669-77.

Tong, W., *et al.* (2003a), "Regulatory Application of SAR/QSAR for Priority Setting of Endocrine Disruptors - a Perspective", *Pure and Applied Chemistry*, 75(11-12): 2375-2388.

Tong, W., *et al.* (2003b), "Decision Forest: Combining the Predictions of Multiple Independent Decision Tree Models", *Journal of Chemical Information and Computer Science*, 43(2): 525-531.

Tong, W., *et al.* (2004), "Assessment of Prediction Confidence and Domain Extrapolation of Two Structure-Activity Relationship Models for Predicting Estrogen Receptor Binding Activity", *Environmental Health Perspectives*, 112(12): 1249-1254.

Tong, W., *et al.* (2005), "Assessing QSAR Limitations - a Regulatory Perspective", *Current Opinions in Computer Aided Drug Design,* 1: 65-72.

Tsakovska, I, T. Netzeva and A. Worth (2005), *Evaluation of (Q)SARs for the Prediction of Eye Irritation/Corrosion Potential. Physicochemical Exclusion Rules*, European Commission, DG JRC, IHCP, ECB, Ispra, Italy, http://ecb.jrc.it/QSAR/, accessed 1 June 2006.

Tunkel, J., *et al.* (2000), "Predicting Ready Biodegradability in the Japanese Ministry of International Trade and Industry Test", *Environmental Toxicology and Chemistry*, 19(10): 2478-2485.

Tunkel, J., *et al.* (2004), "Practical Considerations on the Use of Predictive Models for Regulatory Purposes", presented at the 11th International Workshop on (Q)SAR in the Human Health and Environmental Sciences ((Q)SAR 2004), 9-13 May 2004, Liverpool, England.

Tunkel, J., *et al.* (2005), "Practical Considerations on the Use of Predictive Models for Regulatory Purposes", *Environmental Science and Technology*, 39(7): 2188-2199.

Tyle, H. and J. Niemelä (1999), "Use of (Q)SARs for Selection of POPs", working paper to the participants of CEGII Meeting Vienna, Danish EPA, 14-18 June 1999, pp.38.

Tyle, H., *et al.* (2002), "Identification of Potential PBTs and vPvBs by Use of (Q)SARs", SHC/TS2-3/029, Danish EPA Copenhagen 2002-06-27 (pp. 38) and "Appendix I". SHC/TS2-2/30 (pp. 51).

U.S. EPA (Environmental Protection Agency) (1994a), *US EPA/EC Joint Project on the Evaluation of (Quantitative) Structure Activity Relationships*. EPA 743-R-94-001, U.S. EPA, OPPT (Office of Pollution Prevention and Toxics), Washington, D.C., 364 pp., http://www.epa.gov/oppt/newchems/tools/sar_report.pdf, accessed 7 August 2006.

U.S. EPA (1994b), *Estimating Toxicity of Industrial Chemicals to Aquatic Organisms Using Structure-Activity Relationships*, 2nd Edition. U.S. EPA, OPPT, Washington, DC., EPA 748-R-93-001, available from U.S. EPA, National Service Center for Environmental Publications, tel.: 1-800-490-9198.

U.S. EPA (1994c), *ECOSAR: A Computer Program for Estimating the Ecotoxicity of Industrial Chemicals*, EPA-748-R-93-002, available from U.S. EPA, National Service Center for Environmental Publications, tel.: 1-800-490-9198.

U.S. EPA (2002a), *ECOTOX Database*, U.S. EPA, Washington, DC., http://www.epa.gov/ecotox/, accessed 1 June 2006.

U.S. EPA (2002b), *Exposure Assessment Tools and Models*, U.S. EPA, Washington, DC., http://www.epa.gov/oppt/exposure/docs/episuitedl.htm, accessed 1 June 2006.

U.S. EPA (2002c), *High Production Volume (HPV) Challenge Program*, U.S. EPA, Washington, DC., http://www.epa.gov/chemrtk/volchall.htm, accessed 1 June 2006.

U.S. EPA (2002d), *High Production Volume (HPV) Challenge Program*, U.S. EPA, Washington, DC., http://www.epa.gov/chemrtk/sarfinl1.htm, accessed 1 June 2006.

U.S. EPA (2002e), *Minimum Pre-Market Data/Structure Activity Relationships Study*, U.S. EPA, Washington, DC., http://www.epa.gov/opptintr/MPD-SAR/, accessed 1 June 2006.

U.S. EPA (2002f), *New Chemicals Program*, U.S. EPA, Washington DC., http://www.epa.gov/oppt/newchems/, accessed 1 June 2006.

U.S. EPA (2002g), *The Office of Pesticide Programs Guidance Document on Methodology for Determining the Data Needed and the Types of Assessments Necessary to Make FFDCA Section 408 Safety Determinations for Lower Toxicity Pesticide Chemicals,* U.S. EPA, Office of Pesticide Programs, Washington, D.C., 52 pp. (plus Table). http://www.epa.gov/oppfod01/cb/csb_page/updates/lowertox.pdf , accessed 1 June 2006.

U.S. EPA (2003), *Toxicity Assessment, Risk Assessment, and Risk Management of Chemicals under TSCA in USA*, unpublished manuscript available from Dr. Vincent Nabholz, Risk Assessment Division (mail code: 7403M), Office of Pollution Prevention and Toxics, U.S. EPA, Washington, D.C.

U.S. EPA. (2004), *Status and Future Directions of the High Production Volume Challenge Program,* U.S. EPA, Office of Pollution Prevention and Toxics, Washington, D.C., EPA 743-R-04-001, November 2004, 108 pp, http://www.epa.gov/chemrtk/hpvstatr.htm, accessed 1 June 2006.

U.S. FDA (Food and Drug Administration) (2004), *Challenge and Opportunity on the Critical Path to New Medical Products,* pp. 1-21, http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.html, accessed 1 June 2006.

ITC (2002), *Toxic Substances Control Act (TSCA) Interagency Testing Committee (ITC)*, U.S. EPA, Washington, DC., http://www.epa.gov/opptintr/itc/, accessed 1 June 2006.

Van der Jagt, K., *et al.* (2004), "Alternative Approaches Can Reduce the Use of Test Animals under REACH", addendum to the report *Assessment of Additional Testing Needs under REACH. Effects of (Q)SARs, Risk Based Testing and Voluntary Industry Initiatives*, EUR 21405, European Commission, Joint Research Centre, Ispra, Italy, http://ecb.jrc.it/, accessed 1 June 2006.

Van der Zandt, P.T.J. and C.J. van Leeuwen (1992), *A Proposal for Priority Setting of Existing Chemical Substances*, VROM 92408/b/9-92, 1502/033. Ministry of Housing, Physical Planning and the Environment, Hague, Netherlands.

Van Haelst, A.G. and B.G. Hansen (2000), "Priority Setting for Existing Chemicals: Automated Data Selection Routine", *Env Toxicol Chem*, 19, 2372-2377.

Van Leeuwen, C.J., *et al.* (1996), "Risk Assessment and Risk Management of Chemicals: Status and Future Needs", *Environ. Toxicol. Pharmacol*., 2, 243-299.

Van Leeuwen, C.J., *et al.* (1992), "Application of (Q)SARs, Extrapolation and Equilibrium Partitioning in Aquatic Effects Assessment. I. Narcotic Industrial Pollutants", *Environmental Toxicology and Chemistry*, 11: 267-282.

Van Leeuwen, C.J. and J.L.M. Hermens (eds.) (1995), *Risk Assessment of Chemicals. An Introduction*, Dordrecht, The Netherlands: Kluwer Academic Publishers.

Verhaar, H.J.M., C.J. van Leeuwen and J.L.M. Hermens (1992), "Classifying Environmental Pollutants. 1. Structure-Activity Relationships for Prediction of Aquatic Toxicity", *Chemosphere*, 25: 471-491.

Verhaar, H.J.M., *et al.* (1994), "Application of (Q)SARs in Risk Management of Existing Chemicals", *SAR and (Q)SAR in Environmental Research*, 2: 39-58.

Verhaar, H.J.M., E.U. Ramos and J.L.M. Hermens (1996), "Classifying Environmental Pollutants, 2: Separation of Class 1 (Baseline Toxicity) and Class 2 ("Polar Narcosis") Type Compounds Based on Chemical Descriptors", *Journal of Chemometrics*, 10: 149-162.

Verhaar, H.J.M., *et al.* (2000), "Classifying Environmental Pollutants, Part 3: External Validation of the Classification System", *Chemosphere*, 40: 875-883.

Votta, T.J. and A.L. White (2000), *Design for Competitive Advantage: The Business Benefits of the EPA Pollution Prevention Assessment Framework in New Product Development*, Tellus Institute, 11 Arlington Street, Boston, MA. 50 pages. http://www.epa.gov/oppt/newchems/pubs/sustainable/kodak-tellus1b2c.pdf, accessed 7 August 2006.

Walker, J.D., *et al.* (2004), "(Q)SARs for Predicting Skin Irritation and Corrosion: Mechanisms, Transparency and Applicability of Predictions", *QSAR and Combinatorial Science*, 23, 721-725.

Walker, J.D., *et al.* (2005), "The Skin Irritation Corrosion Rules Estimation Tool (SICRET): Trilateral Cooperation in the Development of Rules for Predicting Skin Irritation and Corrosion", *QSAR and Combinatorial Science*, 24, 378-384.

Walker, J.D. (1993), "The TSCA Interagency Testing Committee's Approaches to Screening and Scoring Chemicals and Chemical Groups: 1977-1983" in P.Y. Lu (ed.), *Access and Use of Information Resources in Assessing Health Risks from Chemical Exposure*, Oak Ridge, TN: Oak Ridge National Laboratories, pp. 77-93.

Walker, J.D. (1995), "Estimation Methods Used by the TSCA Interagency Testing Committee to Prioritize Chemicals for Testing: Exposure and Biological Effects Scoring and Structure Activity Relationships", *Toxicol Modeling*, 1:123-141.

Walker, J.D., *et al.* (2003), "Guidelines for Developing and Using Quantitative Structure-Activity Relationships", *Environmental Toxicology and Chemistry*, 22: 1653-1665.

Weeks, J.A., F.H. Martin and J. Tunkel (2002), *Case Study: SC Johnson's Use of the EPA PBT Profiler to Screen SC Johnson's Chemical Inventory, a Joint Study Conducted by SC Johnson and Syracuse Research Corporation*, SC Johnson, Racine, WI. and the Syracuse Research Corporation, Syracuse, NY, 10 pages, http://www.epa.gov/oppt/newchems/pubs/sustainable/scjohnson-study-2002.pdf, accessed 7 August 2006.

Wilson, A.G., A.C. White and R.A. Mueller (2003), "Role of Predictive Metabolism and Toxicity Modeling in Drug Discovery--a Summary of Some Recent Advancements", *Current Opinions in Drug Discovery and Development*, 6(1): 123-128.

Worth, A.P., *et al.* (2005), *The Characterisation of (Quantitative) Structure-Activity Relationships: Preliminary Guidance*, EUR 21866 EN, Joint Research Centre, European Commission, Ispra, Italy, http://ecb.jrc.it/, 1 June 2006.

Worth, A.P., M.T.D. Cronin and C.J. van Leeuwen (2004a), "A Framework for Promoting the Acceptance and Regulatory Use of (Quantitative) Structure-Activity Relationships", in M.T.D. Cronin and D. Livingstone (eds.), *Predicting Chemical Toxicity and Fate*, pp. 429-440. CRC Press, Boca Raton, FL, USA.

Worth, A.P., C.J. van Leeuwen and T. Hartung (2004b), "The Prospects for Using (Q)SARs in a Changing Political Environment. High Expectations and a Key Role for the European Commission's Joint Research Centre", *SAR and QSAR in Environ. Res.*, 15: 331-344.

Yourick, J.J., *et al.* (1996), "Integrating the QSAR Paradigm into the Safety Assessment of Cosmetic Ingredients and Color Additives", 1996 FDA Science Forum, Abstract F16, http://vm.cfsan.fda.gov/~frf/forum96/sxabsnum.htm, accessed 1 June 2006.

Zeeman, M., *et al.* (1995), "U.S EPA Regulatory Perspectives on the Use of (Q)SAR for New and Existing Chemical Evaluations", *SAR and (Q)SAR in Environmental Research*, 3(3): 179-201.

Zeeman, M. (1995), "Ecotoxicity Testing and Estimation Methods Developed Under Section 5 of the Toxic Substances Control Act. Chapter 23 (pp. 703-715)" in G. Rand (ed.), *Fundamentals of Aquatic Toxicology: Effects, Environmental Fate, and Risk Assessment*, Taylor & Francis, Washington, DC.

Zeeman, M., D. Rodier and J.V. Nabholz (1999), "Ecological Risks of a New Industrial Chemical Under TSCA. Chapter 2 (pp. 2-1 to 2-30)" in *Ecological Risk Assessment in the Federal Government*, U.S. White House, National Science & Technology Council, Committee on Environment & Natural Resources (CENR), Washington, DC. CENR/5-99/001, 219 pp. [The CENR report is available at: http://yosemite.epa.gov/SAB/sabcvpess.nsf/0/b882baf473df807185256de4006a39a5/$file/ecorisk.pdf, accessed 1 June 2006.]

Zinke, S, I. Gerner and E. Schlede (2000), "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 Expert System", *Alternatives to Laboratory Animals,* 30, 285-298.

ABBREVIATIONS

Advisory Committee on Hazardous Substances, ACHS;
Agency for Toxic Substances and Disease Registry, ATSDR;
American Water Works Association, AWWA;
AQUatic toxicity Information REtrieval, AQUIRE;
ASsessment Tools for the Evaluation of Risk, ASTER;
Australian Inventory of Chemical Substances, AICS;
Bioconcentration Factor, BCF;
Boiling Point, BP;
Chemical Stakeholders Forum, CSF;
Chemical Substances Control Law, CSCL;
Chronic Value, ChV;
COmmon REactivity PAttern, COREPA;
Comparative Molecular Field Analysis, CoMFA;
Computational Neural Network, CNN;
Computer Automated Structure Evaluation, CASE;
Department for Environment, Food, and Rural Affairs, DEFRA, UK;
Domestic Substance List, DSL;
DYNamic selection and prioritisation MEChanism, DYNAMEC;
Endocrine Disruption Priority Setting Database, EDPSD;
Endocrine Disruptors Screening and Testing Advisory Committee, EDSTAC;
Environment Agency, EA, UK;
Environmental Risk Assessment, ERA;
European Centre for Validation of Alternative Methods, ECVAM, JRC, Italy;
European Commission, EC;
European Union Scientific Committee on Toxicity, Ecotoxicity and the Environment, CSTEE;
European Union, EU;
Federal Advisory Committee Act, FACA;
Food and Drug Administration, FDA, USA;
German Umweltbundesamt, UBA;
Henry's Law Constant, HLC;
High Production Volume, HPV;
Holographic Quantitative Structure-Activity Relationship, H(Q)SAR;
Integrated Scientific Information System, ISIS;
Interagency Coordinating Committee on the Validation of Alternative Methods, ICCVAM;
International Uniform ChemicaL Information Database, IUCLID;
Japanese Ministry of International Trade and Industry, MITI;
Linear Free Energy Relationship, LFER;
Logarithm of the octanol-water partition coefficient, log Kow;
Logarithm of the organic-carbon partition coefficient, logKoc;
Melting Point, MP;
Minimum Pre-market Data, MPD;
Molecular Design Limited, MDL;
National Industrial Chemicals Notification and Assessment Scheme, NICNAS, Australia;

No Observable Effect Concentration, NOEC;
No Observable Effect Level, NOEL;
Estrogen Receptor, ER;
Office of Pollution Prevention and Toxics, OPPT, USA;
Optimized Approach based on Structural Indices Set; OASIS;
Optimum Prediction Space, OPS;
Organisation for Economic Cooperation and Development, OECD;
Partial Least Squares, PLS;
Persistent Bioaccumulative and Toxic substances, PBTs;
Predicted Environmental Concentration, PEC;
Predicted No Effect Concentration, PNEC;
Pre-Manufacture Notification, PMN;
Probabilistic Neural Network, PNN;
Quantitative Structure-Activity Relationship, (Q)SARs;
Quantitative Structure-Biodegradation Relationship, QSBRs;
Quantitative Structure-Toxicity Relationship, QSTR;
Registration, Evaluation and Assessment of Chemicals, REACH, EU;
Relative Binding Affinities, RBAs;
Screening Information Data Set, SIDS, OECD;
Simplified Molecular Input Line Entry System, SMILES;
Substructure-based Computerized Chemical Selection Expert System, SuCCSES;
Syracuse Research Corporation, SRC, USA;
Technical Advisory Group, TAG, Canada;
Technical Guidance Document, TGD, EU;
Toxic Substances Control Act, TSCA, USA;
TSCA Interagency Testing Committee, ITC, USA;
U.S. Environmental Protection Agency, U.S. EPA;
Vapour Pressure, VP;
very Persistent and very Bioaccumulative substances, vPvBs;

ANNEX: OECD PRINCIPLES FOR THE VALIDATION, FOR REGULATORY PURPOSES, OF (QUANTITATIVE) STRUCTURE-ACTIVITY RELATIONSHIP MODELS

These principles were agreed by OECD member countries at the 37th Joint Meeting of the Chemicals Committee and Working Party on Chemicals, Pesticides and Biotechnology in November 2004. The principles are intended to be read in conjunction with the associated explanatory notes which were also agreed at the 37th Joint Meeting.

To facilitate the consideration of a (Q)SAR model for regulatory purposes, it should be associated with the following information:

1) a defined endpoint[1]
2) an unambiguous algorithm[2]
3) a defined domain of applicability[3]
4) appropriate measures of goodness-of–fit, robustness and predictivity[4]
5) a mechanistic interpretation, if possible[5]

Notes

1. The intent of Principle 1 (defined endpoint) is to ensure clarity in the endpoint being predicted by a given model, since a given endpoint could be determined by different experimental protocols and under different experimental conditions. It is therefore important to identify the experimental system that is being modeled by the (Q)SAR. Further guidance is being developed regarding the interpretation of "defined endpoint". For example, a no-observed-effect level might be considered to be a defined endpoint in the sense that it is a defined information requirement of a given regulatory guideline, but cannot be regarded as a defined endpoint in the scientific sense of referring to a specific effect within a specific tissue/organ under specified conditions.

2. The intent of Principle 2 (unambiguous algorithm) is to ensure transparency in the model algorithm that generates predictions of an endpoint from information on chemical structure and/or physicochemical properties. It is recognized that, in the case of commercially-developed models, this information is not always made publicly available. However, without this information, the performance of a model cannot be independently established, which is likely to represent a barrier for regulatory acceptance. The issue of reproducibility of the predictions is covered by this Principle, and will be explained further in the guidance material.

3. The need to define an applicability domain (Principle 3) expresses the fact that (Q)SARs are reductionist models which are inevitably associated with limitations in terms of the types of chemical structures, physicochemical properties and mechanisms of action for which the models can generate reliable predictions. Further work is recommended to define what types of information are needed to define (Q)SAR applicability domains, and to develop appropriate methods for obtaining this information.

4. The revised Principle 4 (appropriate measures of goodness-of–fit, robustness and predictivity) includes the intent of the original Setubal Principles 5 and 6. The wording of the principle is intended to simplify the overall set of principles, but not to lose the distinction between the internal performance of a model (as represented by goodness-of-fit and robustness) and the predictivity of a model (as determined by external validation). It is recommended that detailed guidance be developed on the approaches that could be used to provide appropriate measures of internal performance and predictivity. Further work is recommended to determine what constitutes external validation of (Q)SAR models.

5. It is recognised that it is not always possible, from a scientific viewpoint, to provide a mechanistic interpretation of a given (Q)SAR (Principle 5), or that there even be multiple mechanistic interpretations of a given model. The absence of a mechanistic interpretation for a model does not mean that a model is not potentially useful in the regulatory context. The intent of Principle 5 is not to reject models that have no apparent mechanistic basis, but to ensure that some consideration is given to the possibility of a mechanistic association between the descriptors used in a model and the endpoint being predicted, and to ensure that this association is documented.