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

**24-Feb-2009**

_____
English - Or. English

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

SERIES ON TESTING AND ASSESSMENT
Number 102

GUIDANCE DOCUMENT FOR USING THE OECD (Q)SAR APPLICATION TOOLBOX TO DEVELOP CHEMICAL CATEGORIES ACCORDING TO THE OECD GUIDANCE ON GROUPING OF CHEMICALS

JT03260091

Document complet disponible sur OLIS dans son format d'origine
Complete document available on OLIS in its original format
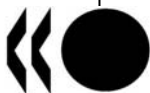
OECD Environment, Health and Safety Publications

Series on Testing and Assessment

No. 102

GUIDANCE DOCUMENT FOR USING THE OECD (Q)SAR APPLICATION TOOLBOX TO DEVELOP CHEMICAL CATEGORIES ACCORDING TO THE OECD GUIDANCE ON GROUPING OF CHEMICALS

IOMC

INTER-ORGANIZATION PROGRAMME FOR THE SOUND MANAGEMENT OF CHEMICALS

A cooperative agreement among **UNEP, ILO, FAO, WHO, UNIDO, UNITAR and OECD**

Environment Directorate

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

Paris 2009

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 (2007)*

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,   *Guidance Document on Mammalian Reproductive Toxicity Testing and Assessment (2008)*

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)*

No. 59, *Report of the Validation of the Updated Test Guideline 407: Repeat Dose 28-Day Oral Toxicity Study in Laboratory Rats (2006)*

No. 60,   *Report of the Initial Work Towards the Validation of the 21-Day Fish Screening Assay for the Detection of Endocrine Active Substances (Phase 1A) (2006)*

No. 61,   *Report of the Validation of the 21-Day Fish Screening Assay for the Detection of Endocrine Active Substances (Phase 1B) (2006)*

No. 62,   *Final OECD Report of the Initial Work Towards the Validation of the Rat Hershberger Assay: Phase-1, Androgenic Response to Testosterone Propionate, and Anti-Androgenic Effects of Flutamide (2006)*

No. 63,   *Guidance Document on the Definition of Residue (2006)*

No. 64,   *Guidance Document on Overview of Residue Chemistry Studies (2006)*

No. 65,   *OECD Report of the Initial Work Towards the Validation of the Rodent Utertrophic Assay - Phase 1 (2006)*

No. 66,   *OECD Report of the Validation of the Rodent Uterotrophic Bioassay: Phase 2. Testing of Potent and Weak Oestrogen Agonists by Multiple Laboratories (2006)*

No. 67,   *Additional data supporting the Test Guideline on the Uterotrophic Bioassay in rodents (2007)*

No. 68,   *Summary Report of the Uterotrophic Bioassay Peer Review Panel, including Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the follow up of this report (2006)*

No. 69,   *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models (2007)*

No. 70,   *Report on the Preparation of GHS Implementation by the OECD Countries (2007)*

No. 71,   *Guidance Document on the Uterotrophic Bioassay - Procedure to Test for Antioestrogenicity (2007)*

No. 72,   *Guidance Document on Pesticide Residue Analytical Methods (2007)*

No. 73,   *Report of the Validation of the Rat Hershberger Assay: Phase 3: Coded Testing of Androgen Agonists, Androgen Antagonists and Negative Reference Chemicals by Multiple Laboratories. Surgical Castrate Model Protocol (2007)*

No. 74,   *Detailed Review Paper for Avian Two-generation Toxicity Testing (2007)*

No. 75,   *Guidance Document on the Honey Bee (Apis Mellifera L.) Brood test Under Semi-field Conditions (2007)*

No. 76,   *Final Report of the Validation of the Amphibian Metamorphosis Assay for the Detection of Thyroid Active Substances: Phase 1 - Optimisation of the Test Protocol (2007)*

No. 77,   *Final Report of the Validation of the Amphibian Metamorphosis Assay: Phase 2 - Multi-chemical Interlaboratory Study (2007)*

No. 78,   *Final Report of the Validation of the 21-day Fish Screening Assay for the Detection of Endocrine Active Substances. Phase 2: Testing Negative Substances (2007)*

No. 79,   *Validation Report of the Full Life-cycle Test with the Harpacticoid Copepods Nitocra Spinipes and Amphiascus Tenuiremis and the Calanoid Copepod Acartia Tonsa - Phase 1 (2007)*

No. 80,   *Guidance on Grouping of Chemicals (2007)*

No. 81,   *Summary Report of the Validation Peer Review for the Updated Test Guideline 407, and Agreement of the Working Group of National Coordinators of the Test Guidelines Programme on the follow-up of this report (2007)*

No. 82,   *Guidance Document on Amphibian Thyroid Histology (2007)*

No. 83,   *Summary Report of the Peer Review Panel on the Stably Transfected Transcriptional Activation Assay for Detecting Estrogenic Activity of Chemicals, and Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the Follow-up of this Report (2007)*

No. 84,   *Report on the Workshop on the Application of the GHS Classification Criteria to HPV Chemicals, 5-6 July Bern Switzerland (2007)*

No. 85,   *Report of the Validation Peer Review for the Hershberger Bioassay, and Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the Follow-up of this Report (2007)*

No. 86,   *Report of the OECD Validation of the Rodent Hershberger Bioassay: Phase 2: Testing of Androgen Agonists, Androgen Antagonists and a 5 α-Reductase Inhibitor in Dose Response Studies by Multiple Laboratories (2008)*

No. 87,   *Report of the Ring Test and Statistical Analysis of Performance of the Guidance on Transformation/Dissolution of*

Metals and Metal Compounds in Aqueous Media (Transformation/ Dissolution Protocol) (2008)

No.88   *Workshop on Integrated Approaches to Testing and Assessment (2008)*

No.89   *Retrospective Performance Assessment of the Test Guideline 426 on Developmental Neurotoxicity (2008)*

No.90   *Background Review Document on the Rodent Hershberger Bioassay (2008)*

No.91   *Report of the Validation of the Amphibian Metamorphosis Assay (Phase 3) (2008)*

No.92   *Report of the Validation Peer Review for the Amphibian Metamorphosis Assay and Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the Follow-Up of this Report (2008)*

No.93   *Report of the Validation of an Enhancement of OECD TG 211: Daphnia Magna Reproduction Test (2008)*

No.94   *Report of the Validation Peer Review for the 21-Day Fish Endocrine Screening Assay and Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the Follow-up of this Report (2008)*

No.95   *Detailed Review Paper on Fish Life-Cycle Tests (2008)*

No.96   *Guidance Document on Magnitude of Pesticide Residues in Processed Commodities (2008)*

No.97   *Detailed Review Paper on the use of Metabolising Systems for In Vitro Testing of Endocrine Disruptors (2008)*

No. 98   *Considerations Regarding Applicability of the Guidance on Transformation/Dissolution of Metals Compounds in Aqueous Media (Transformation/Dissolution Protocol) (2008)*

No. 99   *Comparison between OECD Test Guidelines and ISO Standards in the Areas of Ecotoxicology and Health Effects (2008)*

No.100   *Report of the Second Survey on Available Omics Tools (2009)*

No.101   *Report on the Workshop on Structural Alerts for the OECD (Q)SAR Application Toolbox (2009)*

No.102   *Guidance Document for using the OECD (Q)SAR Application Toolbox to Develop Chemical Categories According to the OECD Guidance on Grouping of Chemicals (2009)*

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 developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organizations.

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 document provides guidance on how to use the OECD (Q)SAR Application Toolbox to build chemical categories according to the OECD Guidance on Grouping of Chemicals. It is part of an overall effort to provide guidance on the use of the (Q)SAR Application Toolbox.

This document is published on the responsibility of the Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology .

ENV/JM/MONO(2009)5

TABLE OF CONTENTS

INTRODUCTION

The first version of the OECD (Q)SAR Application Toolbox has been publicly released in March 2008. It can be downloaded from the following public internet site: www.oecd.org/env/existingchemicals/qsar. The main goal of the Toolbox is to allow the user to use (Q)SAR methodologies to group chemicals into categories and to fill data gaps by read-across, trend analysis and/or external (Q)SARs.

A guidance document for grouping of chemicals was published in September 2007 in the Series on Testing and Assessment of the OECD Environment, Health and Safety Publications (OECD, 2007b).

The 41st Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology agreed that an addendum to the Guidance on Grouping of Substances should be developed which provides guidance on how the (Q)SAR Application Toolbox can be used to develop chemical categories according to the guidance document.

This document is not a reference manual for the use of the OECD (Q)SAR Application Toolbox. It is meant to give guidance to the user on how the Toolbox can be used as a tool for filling data gaps and building chemical categories according to the OECD Guidance on Grouping of Chemicals. The Reference Manual for the Toolbox is downloaded together with the toolbox from the public OECD internet site. The same site also contains a short manual for getting started as well as other training material. It is recommended that users familiarize themselves with the Toolbox before using the present guidance document.

This document is written in a modular form, i.e. the different chapters can be used as stand-alone guidance. This results in some sections being repeated over several chapters. Furthermore, the screenshots in this document reflect version 1.0 of the Toolbox. Some discrepancies with subsequent versions of the Toolbox are possible but should not affect the guidance provided in this document.

CHAPTER 1. THE WORKFLOW OF THE (Q)SAR APPLICATION TOOLBOX

The Toolbox has six work modules which are used in a sequential work flow. Summary background information on the six modules is provided in Table 1.

Table 1. Sequential work flow of the Toolbox

Module	Summary background information
Chemical Input	This module provides the user with several means of entering the chemical(s) of interest or target chemical(s). Since all subsequent functions are based on chemical structure, the goal here is to make sure the molecular structure assigned is the correct one.
Profiling	"Profiling" refers to the electronic process of retrieving relevant information on the target compound(s), other than results for regulatory endpoints.
Endpoints	This module provides the user with an electronic process of retrieving results for regulatory endpoints (e.g. data on environmental fate, ecotoxicity or mammalian toxicity) which are stored in the Toolbox. This data gathering can be executed in a global fashion (i.e., collecting all data of all endpoints) or on a more narrowly defined basis (e.g., collecting data for a single or limited number of endpoints).
Category Definition	This module provides the user with several means of grouping chemicals into a (eco)toxicologically meaningful category that includes the target molecule(s). This is the critical step in the workflow and several options are available in the Toolbox to assist the user in refining the category definition via subcategorization.
Filling Data Gaps	This module provides the user with three options for making an endpoint-specific prediction for the untested chemical(s); in this case the target molecule(s). These options, in increasing order of complexity, are by read-across, by trend analysis, and through the use of QSAR models.
Report	The final module provides the user with a downloadable written audit trail of what functions the user performed using the Toolbox to arrive at the prediction.

To build a category or to perform a simple analogue approach, the user will go through these modules sequentially in the order shown above. When starting the toolbox, the user will be asked to choose between three tracks (or workflows):

- **(Q)SAR Track**: this workflow guides the user through the functionalities of the Toolbox with the aim to fill a data gap by using a library of (Q)SAR models.

- **Category Track**: this workflow guides the user through the functionalities of the Toolbox with the aim to group chemicals into categories and fill data gaps by read-across and/or trend analysis.

- **Flexible Track**: this option gives the user full flexibility of using all the functionalities of the Toolbox in any given order.

For a first use of the Toolbox for building categories, it is recommended to use the **Category Track**. As the user becomes more familiar with the functionalities of the Toolbox, it is recommended to use the **Flexible Track**.
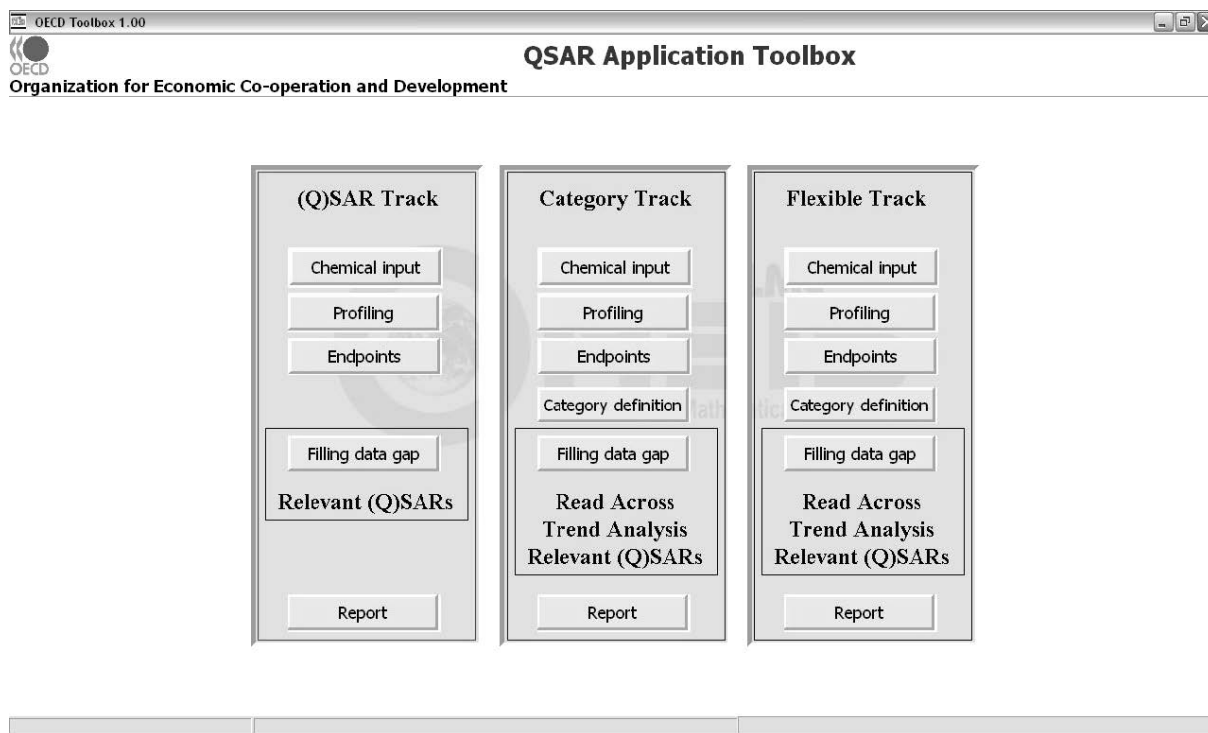
When choosing the **Flexible Track**, the top part of the screen of the Toolbox indicates which module is in use as shown in the screenshot below.
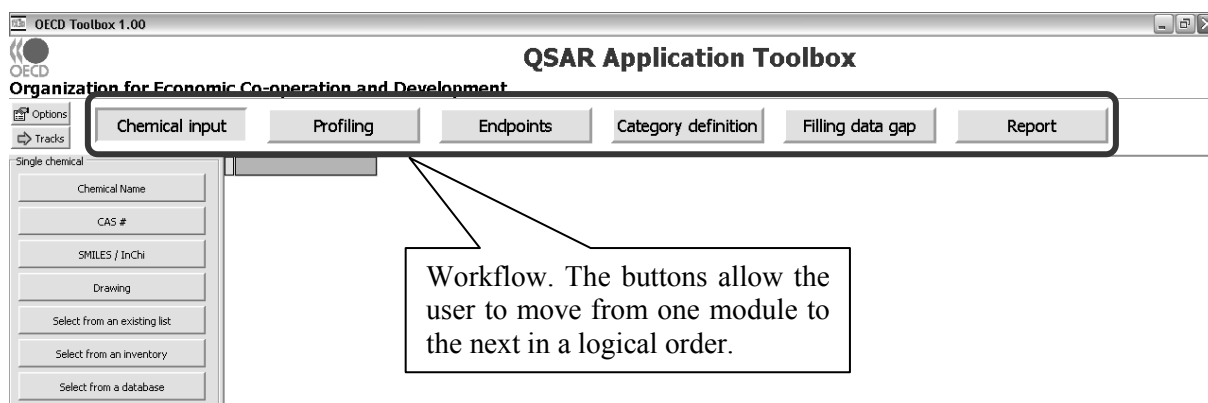
CHAPTER 2. USING THE TOOLBOX AS A SIMPLE DATA RETRIEVAL TOOL

To use the Toolbox as a simple data retrieval tool, the user has to (1) define the target chemical(s) in the module **Chemical Input** and to (2) interrogate the databases in the module **Endpoints**. This is illustrated with a simple example below. When starting the application, the user should choose the **Flexible Track**.

2.1. Defining the target chemical(s)

Defining the target chemical(s) is done in the module **Chemical Input**. The identity of the target chemical(s) can be entered by name, CAS number, SMILES or InChI Code, by drawing the molecule(s) or by selecting the chemical(s) from a list (see screenshot below).
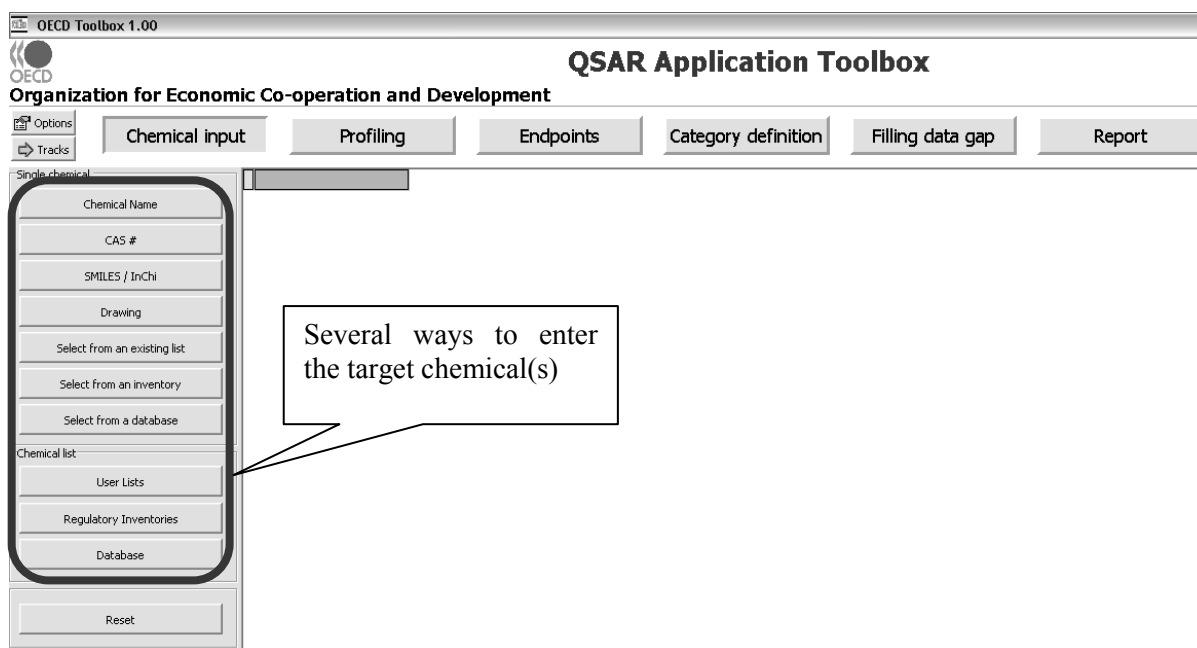
A few examples are presented below.

Defining the target chemical(s) by chemical name

The Toolbox contains a database of chemical names and synonyms. To define a chemical by entering the name, the user can press the button **Chemical Name (1)** and then type the name or a partial name **(2)**, followed by pressing the [🔍] button **(3)**.
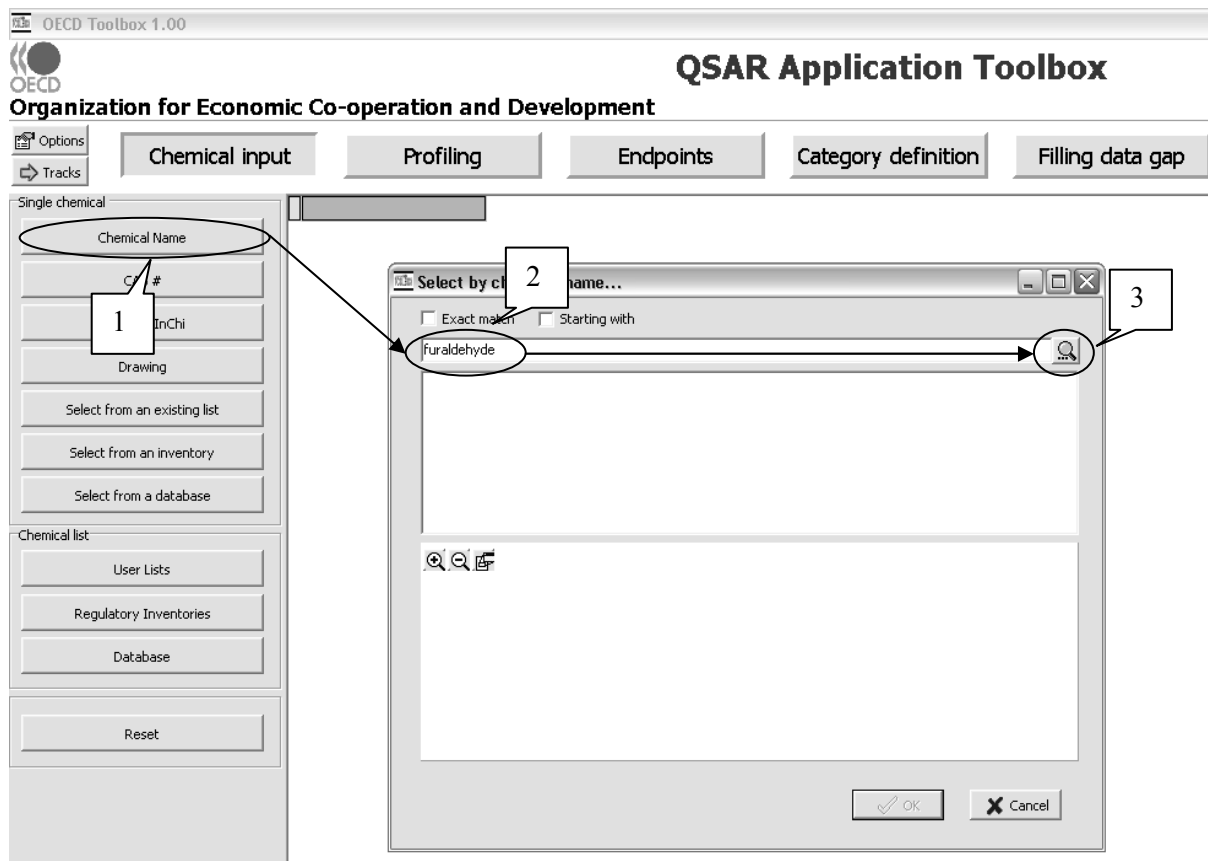
The Toolbox will then find all the chemicals in the database containing the search string in their name and display their structure.
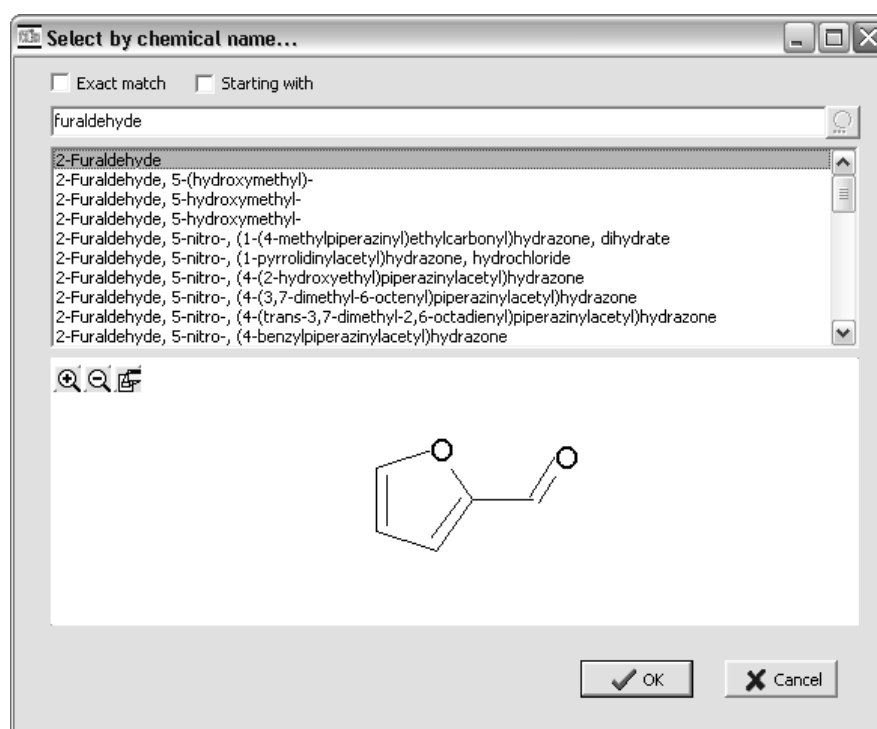
By highlighting the chemical of interest in the list and pressing the **OK** button, the chemical identity (2D-structure, CAS number, chemical name and structural formula in SMILES notation is inserted into the data matrix.

Defining the target chemical(s) by drawing the structure or entering the SMILES or InChi notation

The target chemical(s) can also be entered by defining the structure, either by drawing the structure or by entering the SMILES or InChi notation. By pressing either the button **SMILES/InChi** or **Drawing**, a 2D editor is called up, allowing the user to define the structure of the target chemical(s).
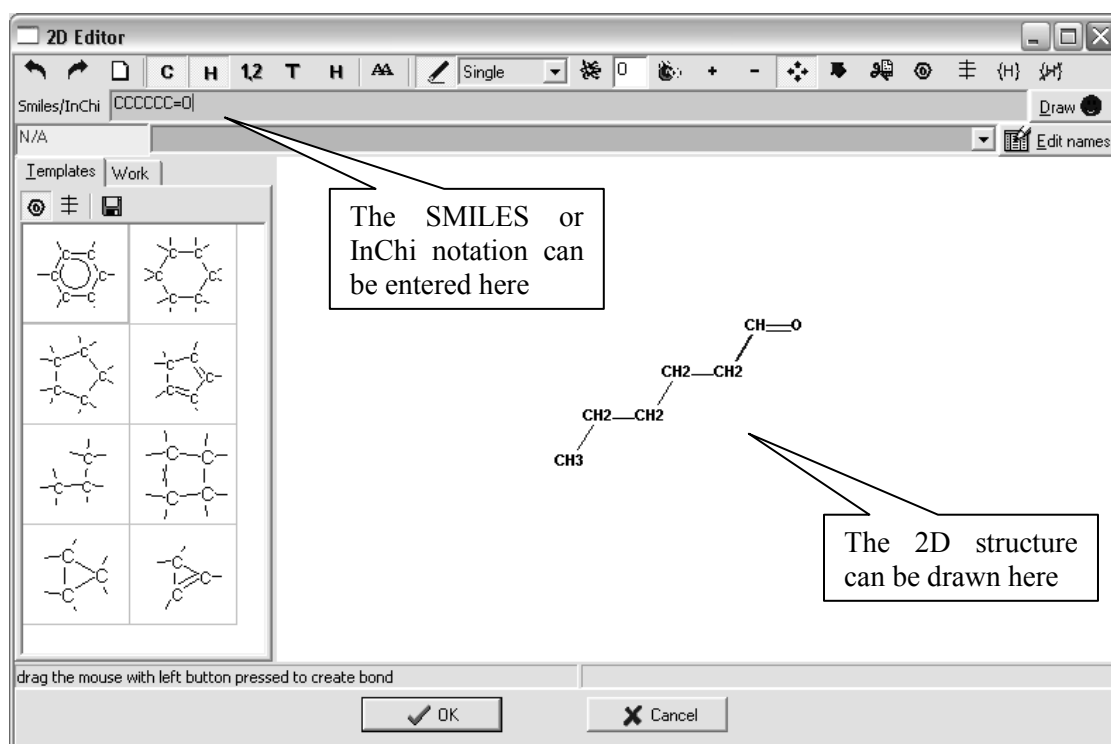
Adding additional target chemicals to the data matrix

It is possible to add additional target chemicals to the data matrix by **(1)** right-clicking on top of the structure of the already defined chemical, selecting **Add Target** in the drop down menu and then **(2)** selecting one of the four proposed possibilities to define the additional chemical (see screenshot below).

Defining lists of target chemicals

The user can also choose to work on user-defined lists of chemicals or whole inventories or complete lists of chemicals figuring in one of the resident databases. For example the user can choose to load the list of OECD HPV Chemicals. To do so, the user can simply press the **Regulatory Inventories** button **(1)** and select **OECD HPVC Inventory (2)** and then press the **OK** button **(3)** (see screenshot below).
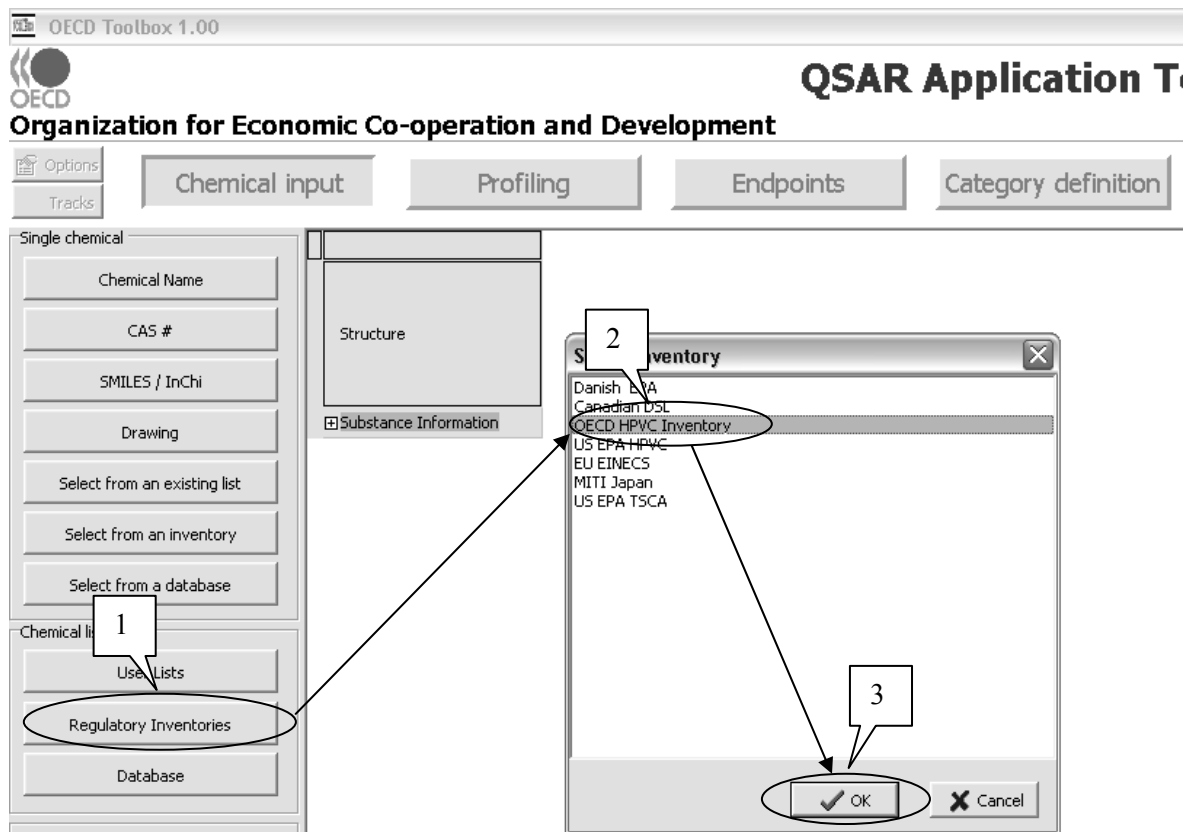
User defined lists can also be loaded by pressing the button **User Lists** and loading a file containing a list with chemical identities. The file has to have the extension "smi" and its format should be tab-delineated ASCII. The file should at a minimum contain the CAS numbers of the chemicals, but it can also contain chemical names and SMILES notations.

An example of the format of a smi-file is illustrated below:

```
CAS      NAME   SMILES
5351-04-2  Propanenitrile, 3-(diethylamino)-C(#N)CCN(CC)CC
78-85-3 2-Propenal, 2-methyl-   C(=C)(C)C=O
536-59-4   1-Cyclohexene-1-methanol, 4-(1-methylethenyl)-   C(=C)(C)C1CC=C(CO)CC1
106-95-6   1-Propene, 3-bromo- C(=C)CBr
598-98-1   Propanoic acid, 2,2-dimethyl-, methyl ester   C(=O)(C(C)(C)C)OC
5837-78-5  2-Butenoic acid, 2-methyl-, ethyl ester, (E)- C(=O)(C(C)=CC)OCC
97-62-1 ethyl_isobutyrate C(=O)(C(C)C)OCC
868-57-5   Butanoic acid, 2-methyl-, methyl ester C(=O)(C(C)CC)OC
868-57-5   Butanoic acid, 2-methyl-, methyl ester C(=O)(C(C)CC)OC
3121-61-7  2-Propenoic acid, 2-methoxyethyl ester C(=O)(C=C)OCCOC
103-41-3   Benzyl_Cinnamate C(=O)(C=Cc1ccccc1)OCc1ccccc1
111-79-5   2-Nonenoic_acid,_methyl_ester C(=O)(C=CCCCCCC)OC
117-81-7   Diethylhexyl phthalate  C(=O)(c1c(C(=O)OCC(CCCC)CC)cccc1)OCC(CCCC)CC
```

2.2. Interrogating the resident databases and evaluating the available results

The interrogation of the resident databases can be done in the module Endpoints .
The list of databases appears on the left-hand part of the screen (see screenshot below).

There are two types of databases. Those containing estimated properties based on QSAR models and those containing experimental results. Currently only the Danish EPA database contains estimated properties. All other resident databases contain experimental results. They are further described in **Table 2** and **Table 3** below. Additional information can be found in part III of the Reference Manual of the Toolbox.

Table 2: Resident databases of version 1.0 of the Toolbox and description of their content

Database	Content	Developed and donated to the Toolbox by	Further information available in/at
Danish EPA	Estimations for numerous properties and effects based on QSAR models.	Danish EPA	http://ecb.jrc.it/qsar/qsar-tools/index.php?c=DDB http://qsar.food.dtu.dk
ECETOC	Experimental results for aquatic toxicity.	ECETOC	ECETOC 2003. Aquatic Hazard Assessment II, Technical Report No. 91. Issn-0773-8072-91 http://www.ecetoc.org/content/Default.asp?PageID=22
ECOTOX	Experimental results for aquatic and terrestrial toxicity as well as bioconcentration. This database integrates three previously independent databases - AQUIRE, PHYTOTOX, and TERRETOX.	US-EPA	http://cfpub.epa.gov/ecotox/
ISSCAN	Experimental results for genotoxicity and carcinogenicity.	Instituto Superiore de Sanità, Italy	http://www.epa.gov/NCCT/dsstox/sdf_isscan_external.html http://www.iss.it/ampp/dati/cont.php?id=233&lang=1&tipo=7
OASIS Aquatic	Experimental results for aquatic toxicity, specifically for five aquatic species Pimephales promelas, Tetrahymena pyriformis, Oryzias latipes, Poecilia reticulate and Daphnia magna.	Laboratory of Mathematical Chemistry, Bulgaria US-EPA University of Knoxville, Tenessee METI Japan	http://cfpub.epa.gov/ecotox/ http://www.safe.nite.go.jp/english/kizon/KIZON_start_hazkizon.html
OASIS Bioaccumulation	Experimental results on bioaccumulation	Laboratory of Mathematical Chemistry, Bulgaria, METI Japan,	http://www.safe.nite.go.jp/english/kizon/KIZON_start_hazkizon.html

		Exxon Mobile	
OASIS Biodegradation	Experimental results on biodegradation	Laboratory of Mathematical Chemistry, Bulgaria, METI Japan	http://www.safe.nite.go.jp/english/kizon/KIZON_start_hazkizon.html
OASIS ERBA	Estrogen Receptor Binding Affinity, expressed as relative binding affinities in comparison with the estradiol affinity.	Laboratory of Mathematical Chemistry, Bulgaria CERI Japan US FDA US EPA University of Illinois University of Kentucky	This database contains results provided by US-EPA, University of Illinois, University of Kentucky as well as from the following sources: Akahori et al., SAR QSAR Environ Res. 16(4), 323-337, 2005.). Fang et al., Chem. Res. Toxicol. 14(3), 280–294, 2001.
OASIS Genetox	Experimental results for genotoxicity.	Laboratory of Mathematical Chemistry, Bulgaria US-NIH US-NTP P&G Danish EPA	This database contains results gathered by The US-HIH, the US-NTP, P&G and the Danish EPA as well as from the following sources: Kazius et al., J. Med. Chem., 48, 312 – 320, 2005 Kirkland et al., Mutation Research, 584, 1–256, 2005 Kier et al., Mutation Research, 168, 69-240, 1986)
OASIS Skin sensitisation	Experimental results for skin sensitisation	Laboratory of Mathematical Chemistry, Bulgaria, Unilever, Exxon Mobile	This database contains results provided by Unilever and ExxonMobile as well as from the following sources: Cronin & Basketter. SAR QSAR Environ. Res. 2(3): 159-179, 1994. Kayser, D., E. Schlede. (eds.) Chemikalien und Kontagtalergie – eine bewertende Zusammen-stellung. Medizin&Wissen Verlagges, Munchen. ISBN 3-86094-163-1, 2001.

Version 1.1 of the Toolbox contains additional databases, as described in**Table 3**.

Table 3: Additional resident databases of version 1.1 of the Toolbox and description of their content

Database	Content	Developed and donated to the Toolbox by	Further information available in/at
CANADA Bioaccumulation	Experimental results on bioaccumulation in aquatic organisms	Environment Canada	
CEFIC – LRI BCF	Experimental results on bioaccumulation in fish.	CEFIC - LRI	http://www.euras.be/eng/project.asp?ProjectId=92
ECETOC Eye Irritation	Experimental results on rabbit eye irritation	ECETOC	ECETOC 1992. Eye Irritation Reference Chemicals Data Bank, Technical Report No. 48. ISSN-0773-8072-48(2) http://staging.idweaver.com/ECETOC/Documents/TR%20048.pdf
ECETOC Skin Sensitisation	Experimental results on skin and respiratory sensitization	ECETOC	ECETOC 1999. Skin and Respiratory Sensitisers: Reference Chemicals Data Bank. Technical report No. 77. ISSN-0773-8072-77 http://staging.idweaver.com/ECETOC/Documents/TR%20077.pdf
EPISUITE_OBS_DATA	Experimental results on physical-chemical properties, as accessed from EPISUITE. This is an abridged version of the PHYSPROP database maintained at Syracuse Research Corporation.	US-EPA & Syracuse Research Corporation	For the full database and references see http://www.syrres.com/eSc/physprop.htm
Japan Aquatic	Experimental results on aquatic toxicity based on tests performed within the Japanese Existing Chemicals Programme.	Japanese Ministry of Environment	http://www.safe.nite.go.jp/english/db.html
Japan EXCHEM	Experimental results from single dose toxicity test and mutagenicity test results performed under Japan's Existing Chemicals Programme.	Japanese Ministry of Health, Labour and Welfare.	http://dra4.nihs.go.jp/mhlw_data/jsp/SearchPageENG.jsp
RIVM skin Irritation	Primary Skin Irritation Indices from skin irritation	RIVM, the Netherlands	The database contains results from the following sources:

	tests		Barratt, ECVAM Workshop 6: Corrosivity
			ECETOC 1995. Skin Irritation and Corrosion: Reference Chemicals Data Bank. Technical Report No.66. [http://staging.idweaver.com/ECETOC/Documents/TR%20066.pdf]
			Kodithala et al. (2002) Toxicol.Sci. 66(336-346)

Additional databases are made available on the public OECD web site and can be downloaded and installed into the Toolbox [www.oecd.org/env/existingchemicals/qsar].

As updated versions of the Toolbox become available, the tables above will become outdated. Background information on any database in the Toolbox can be obtained by highlighting a database **(1),** then right-clicking and selecting **About** in the pop-up menu **(2).**
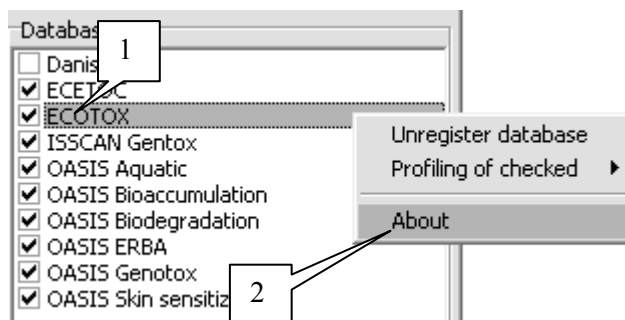
NOTE: The resident databases of the Toolbox have been implemented into the Toolbox as they have been donated. They have not undergone any modification and they have not undergone any quality assurance by the OECD. Users of the Toolbox have therefore to decide on their own whether the data is adequate for their purposes.

The user can select the databases to interrogate by checking the tick-boxes next to their name in the **Databases** box **(1)**, then selecting whether the Toolbox should retrieve Tested (experimental) results or Estimated results or Both in the **Data Summaries** box **(2)** and then pressing the **Gather data** button **(3)**.

The Toolbox will then interrogate the selected databases and retrieve the available results. As some of the databases may contain overlapping data, or because some results are published in different publications, some of the results may be retrieved more than once. The Toolbox identifies those multiple entries and lets the user choose which results to import into the data matrix.
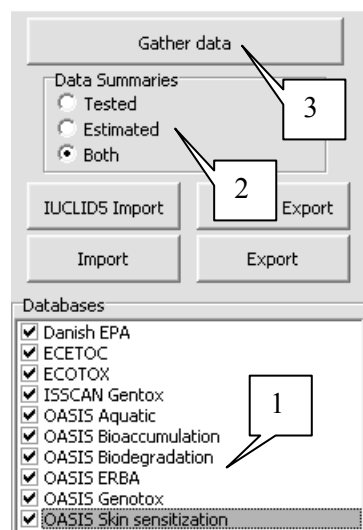
This is shown in the screenshots below. In this example, identical results with different references are retrieved for acute toxicity to fish for n-hexanal. The user can then either manually select those results s/he wants to insert into the data matrix by unchecking the redundant information **(1)** or by pressing the **Select single** button **(2)**, which will automatically deselect any redundant entries.
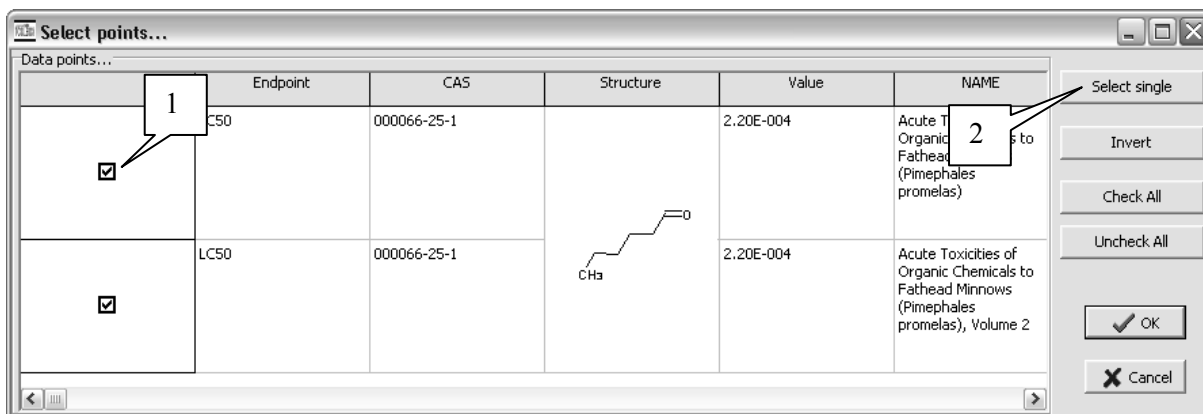
After pressing the **OK** the button, the results will be inserted into the data matrix (see screenshot below). The user can navigate through the data tree by opening or closing the tree-nodes by double-clicking upon them **(1)**. Experimental results are differentiated from estimated results by different letters in front of the numerical values (T = tested; E = Estimated) **(2)**.

Details on the results can be obtained by double-clicking the corresponding cell in the data matrix **(1)** (see screenshot below). This will open a window displaying additional information on the result, such as test conditions and reference **(2)**.

The user can thereby decide whether the result is adequate for the regulatory purpose used.

CHAPTER 3. ANALOGUE APPROACH: HOW TO FILL DATA GAPS USING THE SIMPLIFIED ANALOGUE APPROACH

The OECD Guidance on Grouping of Chemicals recommends a stepwise procedure for read-across using the analogue approach (see Figure 1). The following sections indicate how the Toolbox can be used to perform the different steps. Simple examples are used to illustrate the different steps.

Before performing the operations outlined below, the user has to enter the chemical identity of the chemical (see Chapter 2).

Figure 1. Stepwise approach to an analogue approach
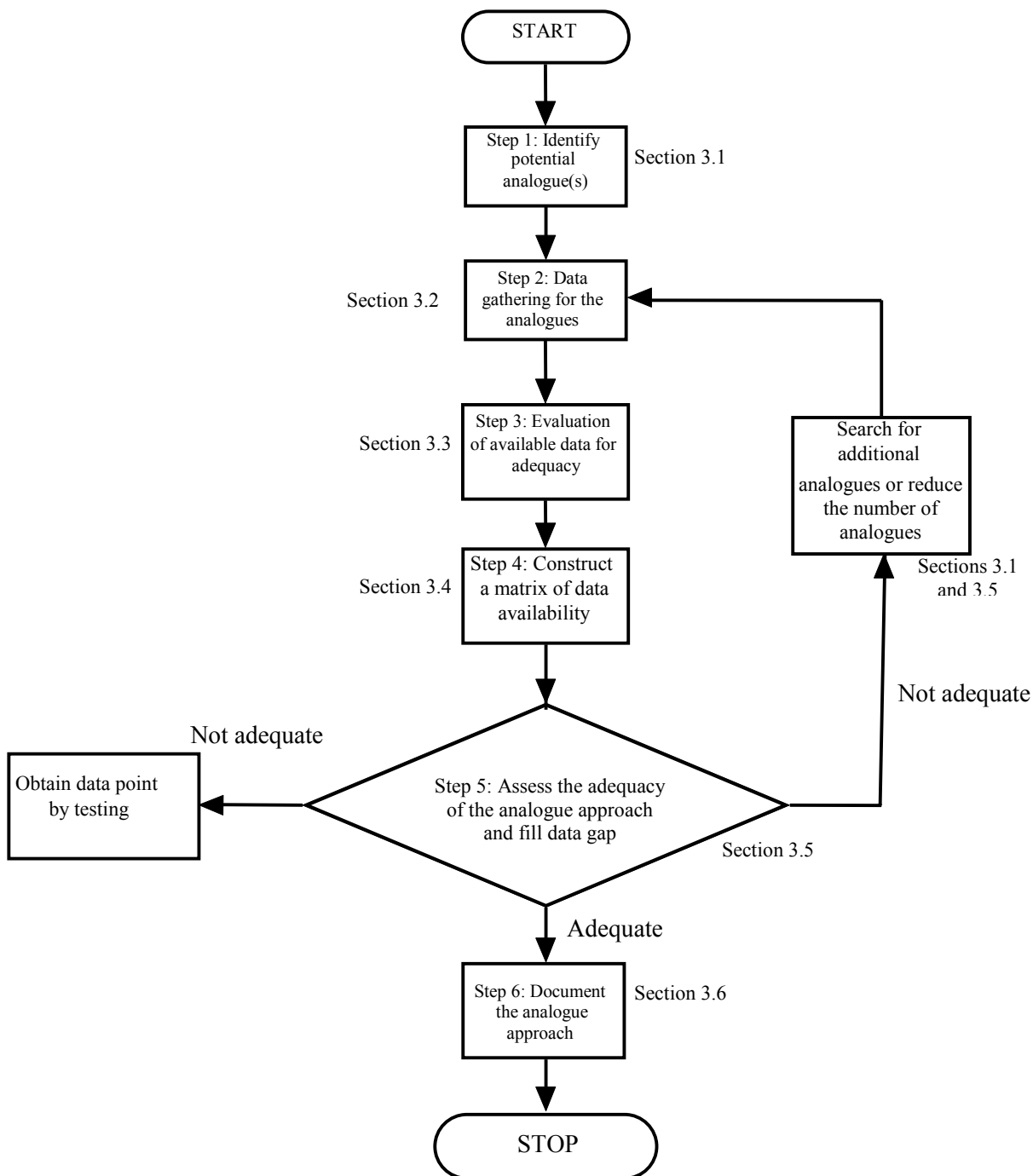
3.1. Identification of potential analogues

The Toolbox provides many tools to identify potential analogues, e.g., structural analogues, i.e. chemicals that have a similar structure compared to the target chemical(s) or mechanistic analogues, i.e. chemicals that have similar mechanisms or modes of action compared to the target chemical(s). In the Toolbox, the identification of potential analogues is a three-step process.
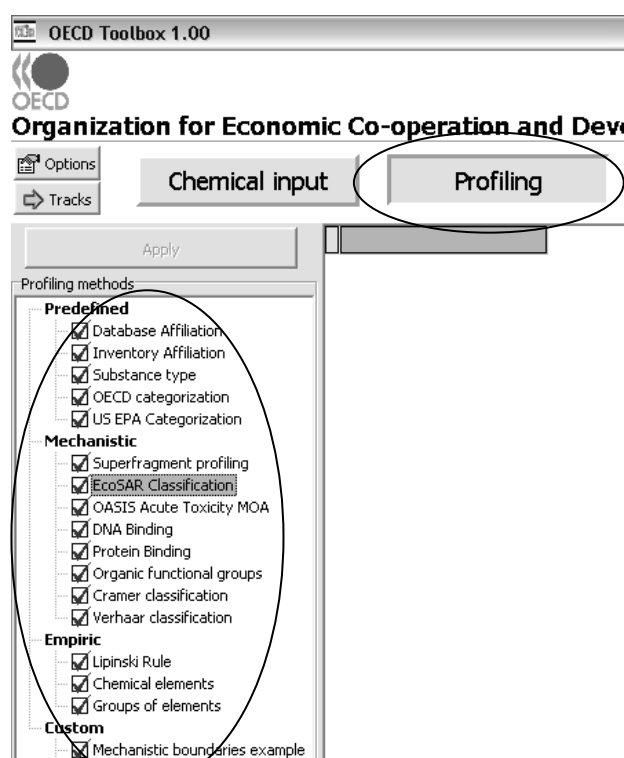
The first step is to profile the target chemical(s) with the module **Profiling** . Version 1.0 of the Toolbox contains eighteen profilers, which form the basis of identifying potential analogue(s). These grouping methods are divided into four types, Predefined, Mechanistic, Empiric, and Custom. The Predefined profilers include; (1) database or (2) inventory affiliation, (3) substance type, and (4) OECD or (5) USEPA categorization. The Mechanistic profilers include; (1) superfragment, (2) ECOSAR classification, (3) OASIS acute toxicity mode of action, (4) DNA binding, (5) Protein binding, (6) Organic functional groups, and (7) Cramer and (8) Vehaar classification. Empiric profilers include; (1) Lipinski rules, (2) Chemical elements and (3) Groups of elements, while the Custom profilers are grouping methods developed by the user. One example is part of the default installation. Additional profilers have been implemented in version 1.1 of the Toolbox. These profilers or grouping methods vary in their design, complexity, and selectivity. The net result of this is a preferred pattern of use, which is endpoint-dependent.

NOTE: It is difficult to give precise guidance on what constitutes a "good" analogue. Indeed, dramatic changes in toxic potency are often linked to subtle changes in structure. The practice in chemical review programmes like the OECD HPV Chemicals Programme or the US HPV Challenge Program is to identify similar chemical structures and then to gather information to support or refute the robustness of the use of those analogues. Experience with using the Toolbox shows that best results are obtained by first identifying chemicals with similar mechanisms or modes of action and then identifying among those the chemicals which have the closest structural similarity to the target chemical(s). This philosophy is therefore implemented in this guidance document. Nevertheless, as the examples provided below demonstrate, either approach can be used successfully.

It should be highlighted that not all mechanistic profiling tools are relevant for all regulatory endpoints. Based on our current understanding of the chemical and biological events leading to toxicity, different profilers are amenable to different endpoints (see Table 4 and Table 5). For example it is generally accepted that sensitization is linked to protein binding and DNA binding is one mechanism well linked to carcinogenicity and genotoxicity. So these profilers are logical choices in modelling these two endpoints, respectively. Furthermore, some profilers might be redundant.

Table 4: Summary information for the mechanistic profilers included in version 1.0

Profiler	Summary background information
Protein binding	The protein binding categorization scheme includes 38 categories such as haloalkanes, isocyanates, isothiocyanates, diketones, aldehydes, acyl halides, alkyl sulfates, sulfonates, etc. Each category is presented by defined 2D structural alerts that are responsible for the eliciting of effects as a result of protein binding, such as skin sensitization. The associated mechanisms are in accordance with the existing knowledge on electrophilic interaction mechanisms of various structural functionalities (Dimitrov et al., 2005). This classification scheme is particularly relevant for skin and respiratory sensitization and acute aquatic toxicity, but also for chromosomal aberration and acute inhalation toxicity. It is built on conventional organic chemical mechanisms and as such is qualitative in character.
Verhaar Classification	Utilizing an acute toxicity data collection for guppies, Verhaar et al. (1992) delineated the classes of inert, less inert, reactive, and specifically-acting chemicals, and provided the chemical rules for discrimination of the first three. This classification scheme is specifically relevant for acute aquatic toxicity endpoints.
OASIS Acute Toxicity MOA	This profiler also classifies chemicals for their acute aquatic toxicity mode of action. This is a behavioral mode of action. It was originally developed by the US-EPA and is based on a set of chemicals tested with fathead minnow (Russom, 1997). It was further elaborated by LMC. This classification scheme is specifically relevant for acute aquatic toxicity endpoints.
ECOSAR Classification	ECOSAR is a program developed by the US-EPA containing structure-activity relationships (SARs) used to predict the aquatic toxicity of chemicals based on their similarity of structure to chemicals for which the aquatic toxicity has been previously measured [http://www.epa.gov/oppt/newchems/tools/21ecosar.htm]. SARs are developed for chemical classes based on measured test results. To date, over 150 SARs have been developed for more than 50 chemical classes. The ECOSAR Classification profiler in the Toolbox retrieves the class that a chemical belongs to. This classification scheme is specifically relevant for aquatic toxicity endpoints.
DNA binding	DNA binding categorization scheme is based on the model of Ames mutagenicity developed by LMC (Mekenyan et al., 2004). The scheme includes 19 categories. Each category is defined by 2D structural alerts that are a necessary condition for a chemical to covalently interact with DNA and elicit mutagenicity. Definition of these alerts was justified by their interaction mechanisms with DNA, found in the literature (Serafimova et al., 2007). This classification scheme is particularly relevant for Ames mutagenicity.
Cramer Classification	The Cramer classification scheme (tree) is probably the best known approach for structuring chemicals in order to make an estimation of a Threshold of Toxicological Concern (TTC), which is a concept that aims to establish a level of exposure for all chemicals below which there would be no appreciable risk to

	human health. The tree relies primarily on chemical structures and estimates of total human intake to establish priorities for testing. The procedure uses recognized pathways for metabolic deactivation and activation, toxicity data and the presence of a substance as a component of traditional foods or as an endogenous metabolite.
	Substances are classified into one of three classes.
	Class 1 contains substances of simple chemical structure with known metabolic pathways and innocuous end products which suggest a low order of oral toxicity.
	Class 2 contains substances that are intermediate. They possess structures that are less innocuous than those in Class 1 but they do not contain structural features that are suggestive of toxicity like those in Class 3.
	Class 3 contains substances with a chemical structure that permit no strong initial impression of safety and may even suggest a significant toxicity.
Organic functional groups	This profiler simply identifies the organic functional groups in the molecule.
Superfragment profiling	The extended (super) fragment identification is based on the algorithm developed by BioByte Corporation. Most current methods of structure evaluation break up the structure according to conventional chemistry-based schemes. The method used in CLOGP is different in that it is almost entirely 'rule based'; fragments as defined by the CLOGP algorithm are not simply those chosen by chemical intuition. Instead, an analysis of the molecule is performed that identifies 'isolating carbons', locations in the compound that act as high-threshold barriers to movement of electrons. All remaining connected group of atoms are then designated as 'polar fragments', i.e. simple fragments.
	An extended fragment therefore consists of a combination of simple fragments that are in such close proximity that their solvation behavior (evidenced by octanol/water log P) is markedly affected. The simplest definition of an extended or super fragment is the "largest electronically-connected substructure". The CLOGP software defines nearly 1,000 fragments, and defines a large number of ways in which fragment proximities are analyzed. Presently, the software is restricting the candidates for extended fragments to those separated by one or two isolating carbons, designating these Y-C-Y and Y-C-C-Y respectively.

Additional profilers have been included in version 1.1 of the Toolbox, as outlined in Table 5

Table 5: Summary information for the additional mechanistic profilers included in version 1.1

Profiler	Summary background information
Benigni/Bossa rulebase	The Benigni/Bossa rulebase for mutagenicity and carcinogencity was developed as a module (plug-in) to the Toxtree software (EC, 2008). The structural alerts (SAs) from the rulebase (Benigni et al., 2008) have been included as a profiler in the Toolbox.
	The list of SAs refers mainly to the knowledge on the action mechanisms of genotoxic carcinogenicity (thus they apply also to the mutagenic activity in bacteria), but includes also a number of SAs flagging potential nongenotoxic carcinogens.
	Please note that only the structural alerts have been implemented as a profiler. No

	final classifications on mutagenicity or carcinogenicity are provided as a result of using the profiler.
BFR rulebases for skin and eye irritation/ corrosion	The rulebase for skin and eye irritation and corrosion developed by the German Federal Institute for Risk Assessment (BfR) and collaborators have been implemented into the Toolbox as a set of structural alerts and physical-chemical exclusion criteria (Gerner et al., 2005; Walker et al., 2005)
	Please note that only the structural alerts and physical-chemical rules have been implemented as a profiler. No final classifications on irritation or corrosion are provided as a result of using the profiler.
BIOWIN MITI fragments	The BIOWIN MITI fragments categorisation scheme is based on the structural alerts used by the MITI Biodegradation Probability Model developed for the US Environmental Protection Agency. Details on the model and structural alerts can be found in Tunkel et al. (2000)
	Please note that only the structural alerts have been implemented as a profiler. No predictions on biodegradability are provided as a result of using the profiler.
	This categorisation scheme is particularly relevant for ready biodegradability.
ER-binding	Estrogen receptor (ER) binding is a molecular initiating event much like protein binding. It is an endpoint where several comprehensive databases exist, which has lead to the development of several approaches for using (Q)SARs to predict ER-binding and possible subsequent endocrine disruption. The ER-binding profiler in the Toolbox is based on the "four phase" assessment that includes Comparative Molecular Field Analysis (CoMFA) and the Common Reactivity Pattern Approach (COREPA).
	Background information on these approaches can be found in Schultz et al. (2006); Cronin & Worth (2008); Tong et al. (2004); Schmieder et al. (2003); Serafimova (2007b).

The second step is to determine the scope of the search by selecting the databases or inventories to which a search should be applied. This is done in the module **Endpoints**. It is usually recommended to focus on databases which contain results for the endpoint of interest; however, information on other endpoints can sometime provide valuable insights (e.g. chemical reactivity responsible for sensitization or genotoxicity often also result in differentiating classifications or categorization for aquatic toxicity. Where such consistencies can be found, such information taken together strengthens a category evaluation). The content of the different databases is described in Chapter 2.

The third step is to launch a query in the module **Category definition**. The outcome of the profiling determines the most appropriate way to search for analogues. The following recommendations can be made:

- If specific mechanisms or modes of action are identified for a target chemical(s), which are relevant for the investigated endpoint, then it is recommended to search for chemicals which have

the same mechanisms or modes of action. The search results can then be refined by identifying among those the chemicals which have the closest structural similarity to the target chemical.

- If <u>no</u> specific mechanisms or modes of action are identified for a target chemical(s), which are relevant for the investigated endpoint, then it is recommended to search for chemicals which are structurally similar to the target chemical. The search results can then be refined by eliminating those chemicals which have specific mechanisms or modes of action.

> NOTE: This recommendation is based on experience gained by using the Toolbox and should not be regarded as restricting any other approach for identifying category members.

As indicated above, the strategy for searching analogues is also endpoint-dependant. Some guidance can be provided for a number of endpoints.

Sensitization

Sensitization is modelable from a qualitative perspective as the molecular initiating event for sensitization is universally considered to be the covalent binding of sensitizers to proteins. This means the **Protein binding** profiler provides a relevant grouping method for qualitatively indentifying chemicals that are potential sensitizers. The iterative use of the **Organic functional groups** profiler is a useful tool for conducting a secondary grouping to indentify structurally similar compounds within a mechanism of protein binding with which to conduct quantitative read-across.

Since in selected cases biotransformation of a non-protein-binding parent compound can lead to a protein-binding metabolite, the metabolism profilers are often included either as a primary or secondary profiler for sensitization (see section 3.7 for specific guidance on the metabolism profilers).

Carcinogenicity and Genotoxicity

There are genotoxicity as well as non-genotoxicity mechanisms of carcinogenicity. The best known genotoxicity mechanisms are those of DNA-binding. This means the **DNA binding** profiler provides a logical grouping method for qualitatively indentifying chemicals that are potential carcinogens and quantitatively indentifying chemicals that are potential mutagens. Furthermore the **Benigni/Bossa rulebase** profiler contains alerts for both genotoxicity as well as non-genotoxicity mechanisms. Since in many cases biotransformation of a non-DNA-binding parent compound can lead to a DNA-binding metabolite the metabolism profiler is often included either as a primary or secondary profiler for carcinogenesis and genotoxicity (see section 3.7 for specific guidance on the metabolism profilers).

Relevant profilers for chromosomal aberration are reflective of the fact that proteins are a major structural aspect of chromosomes. Thus, chemicals, which bind to proteins, have the potential to cause chromosomal aberrations, especially *in vitro*. Therefore when searching for analogues to read-across for chromosomal aberration, the **DNA binding** profiler and the **Protein binding** profiler should be used concurrently.

Skin Irritation

The relevant profiler to search for analogues to read-across for skin irritation is the **BfR rulebase for skin irritation**.

Eye Irritation

The relevant profiler to search for analogues to read-across for skin irritation is the **BfR rulebase for eye irritation**. It should be noted though that the databases for eye irritation in the Toolbox (up to version 1.1) are quite limited and it is often difficult to identify adequate analogues.

Aquatic Toxicity

Acute aquatic effects are currently the most data-rich endpoints in the Toolbox. The depth and breadth of these data plus the long history of modelling such acute effects brings a greater sophistication to profiling. This is aided by several facts including (1) acute aquatic toxicity has a water solubility-related baseline, (2) the majority of industrial organic chemicals are modelled as baseline toxicants, and (3) several historical profilers including **ECOSAR classification**, **OASIS acute toxicity mode of action** and **Vehaar classification** have been developed based on these measured data. Each of these latter three profilers has their advantages and disadvantages and they are best used in parallel. It should be noted though that the **ECOSAR classification** scheme is the most recent, most comprehensive and most discriminating of the three profilers.

An examination of structural alerts most often associated with excess acute aquatic toxicity reveals them to be remarkably similar to those found in the **Protein binding** profiler. Thus, the **Protein binding** profiler should be included along with the three above noted profilers as an initial battery of grouping methods.

As shown in the two screenshots below, when multiple profilers are used, common patterns of responses are often observed.

As with sensitisation, in selected cases of aquatic toxicity biotransformation of a non-protein-binding parent compound can lead to a protein-binding metabolite thus, the metabolism profiler is often included either as a primary or secondary profiler for acute aquatic effects (see section 3.7 for specific guidance on the metabolism profilers).

The iterative use of the **Organic functional groups** profilers is a useful tool for conducting a secondary grouping to identify structurally similar compounds with which to conduct read-across.

Bioaccumulation:

Bioaccumulation is another endpoint for which the Toolbox contains extensive data. Since bioaccumulation is related to lipophilicity, size, and shape, and is independent of the toxicity mechanism, the **Organic functional groups** profiler is a useful tool to indentify structurally similar compounds with which to perform read-across.

Biodegradation:

Only results from ready biodegradability tests are included in the Toolbox (version 1.0 and 1.1). The **BIOWIN MITI fragment** profiler having been developed based on results from tests on ready biodegradability, this profiler can be used to group chemicals for read-across of ready biodegradability.

The three steps described above (Profiling, Endpoints, and Category definition) for identifying potential analogues can be illustrated with the two examples below.

Example 1: Using a mechanistic profiling result to identify analogues

The example is illustrated by filling a data gap for skin sensitization for 4-nitrobenzoyl chloride (CAS No 122-04-3). For guidance on how to enter the chemical identity, see Chapter 2. Of the profilers available in the Toolbox, that which is most relevant for skin sensitization is protein binding. After selecting the **Protein Binding** profiler **(1)** and pressing the **Apply** button **(2)**, the Toolbox inserts the result into the data matrix (see screenshot below). According to the profiler, the target chemical has the potential to react with proteins by **nucleophilic substitution of acyl halides (3)**. Background information on the profiling result can be obtained by double-clicking in the profiling result in the matrix **(4) (5)**.

Identification of a specific mechanism for the target chemical(s) means that it should be possible to identify other chemicals which have the same mechanistic profile, for which experimental data is available.

The next step is therefore to select the databases that contain experimental data on sensitisation under

Endpoints

(1) (see screenshot below). Among the resident databases, currently only the database **OASIS Skin Sensitisation (2)** contains experimental data on skin sensitisation. The radio button **Tested (3)** should also be selected in case the query is to be restricted to experimental results.

The actual query can then be launched under **Category definition** **(1)** (see screenshot below) by selecting the **Protein Binding** as the grouping method of the target chemical **(2)** and pushing the button **Defining Category (3)** (see screenshot below).

After confirming the mechanism that should be used in the query **Nucleophilic substitution of acyl halide (4)**, the Toolbox will now retrieve those chemicals that have the same protein binding mechanism as the target compound and for which data is available in the previously selected database. In this example, 6 additional structures were retrieved (see screenshot below).

Example 2: Searching for structural analogues

As indicated above, analogues can also be identified based on structural similarity. This is illustrated with this example where a data gap for Ames mutagenicity is filled for n-hexanal (CAS No 66-25-1). For guidance on how to enter the chemical identity, see Chapter 2. The profiler which is most relevant for mutagenicity is DNA binding. After selecting the **DNA binding** profiler **(1)** and pressing the **Apply** button **(2)**, the Toolbox inserts the result into the data matrix (see screenshot below). According to the profiler, **No Binding** mechanism is identified **(3)**.

No Binding is the default profiling result when the Toolbox does not identify a known binding mechanism which is programmed in the Toolbox. The utility of using the Toolbox for identifying other chemicals with the same (absent) interaction mechanism is tenuous because the profiling result **No Binding** does not necessarily mean that the compound cannot interact with DNA, it only means that the Toolbox (based on the knowledge currently programmed into it) has not identified an interaction mechanism. The identification of potential analogues that can be used for read-across has therefore to be based on structural similarity.

The next step is to select the databases that contain experimental data on mutagenicity under

Endpoints

(1) (see screenshot below). Among the resident databases in version 1.0 the databases **ISSCAN Gentox** and **OASIS Genotox (2)** contain experimental data on mutagenicity. The radio button **Tested (3)** should also be selected in case the query is to be restricted to experimental results.

Category definition

The actual query can then be launched in the module [Category definition]. Among the different grouping methods available in the Toolbox, there are two tools which are specifically aimed at grouping chemicals according to structural similarity. They are listed as **Organic functional groups** (under Mechanistic) and **Structure similarity** (under Empiric). Details on those two tools can be viewed by selecting the profiler in the list and pushing the button **Show Category Boundaries**.

The grouping method **Organic functional groups** allows the user to query for chemicals which have the same organic functional groups as the target compound. Currently about 180 functional groups are defined in this tool (see screenshot below for the functional group Aldehyde).

The grouping method **Structure similarity** refers to a number of mathematical models to estimate the structural similarity between two chemicals. Currently there are 14 different mathematical models that the user can apply, defining a threshold above which s/he considers the two chemicals to be sufficiently similar, expressed in % (see screenshot below). Additional information on these mathematical models can be found in Chapter B of the Reference Manual for the Toolbox.
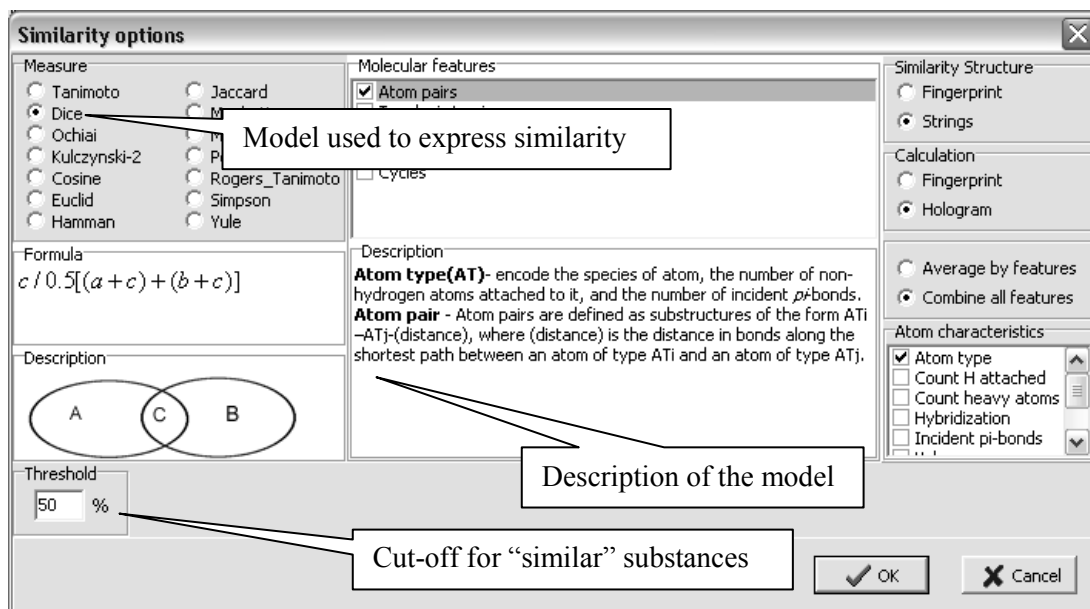
Since there is no preferred way of identifying structural similarity with mathematical models, it is recommended to always use the grouping method **Organic functional groups** as a first option. The query can be launched by selecting the grouping method **Organic functional groups (1)** and pushing the **Defining Category** button **(2)** (see screenshot below).

After confirming the identified functional groups in the target molecule **Aldehyde (3)** that should be used in the query, the Toolbox will upon pushing the **OK** button **(4)** retrieve those chemicals that contain the functional groups of the target chemical in their molecular structure and for which data is available in the previously selected database. In this example, 35 additional structures were retrieved.

> NOTE: Any of the grouping methods finds chemicals which comply with the structural characteristics of the target compound. But among the identified chemicals, many will have additional characteristics, e.g. additional organic functional groups. To define a more homogeneous group of chemicals, it is therefore good practice to subcategorise the group of identified chemicals and to prune all chemicals with additional characteristics.

This group of 35 chemicals can be further subcategorised by eliminating chemicals with additional organic functional groups. To do so, select the defined category of **[35] Aldehydes <AND>…(1)** [1], press the **Subcategorisation** button **(2)** and select **Organic functional groups (3)** as a subcategorisation grouping method. The Toolbox then identifies all chemicals with additional organic functional groups and the user can remove those chemicals from the category by pressing the button **Prune (4)** (see screenshot below).

---

[1] When selecting a defined category for the first time, the Toolbox asks the user to select the data to be retrieved from the previously selected databases. See section 3.2 for guidance on these steps.

This will result in a subcategory of 9 chemicals with very similar structures (see screenshot below).

3.2. Gathering data for the analogues

Gathering data from the databases in the Toolbox is not a separate step. By selecting a newly defined category in the **Defined categories** section, the Toolbox will automatically request the user to select the endpoint that should be retrieved. The user can either select a specific endpoint or by default choose to retrieve data on all endpoints.

NOTE: Data will only be retrieved from those databases which have been selected previously

under **Endpoints**

Example 1: Gathering data for "mechanistic" analogues

Continuing the example of a simple analogue approach of sensitisation for 4-nitrobenzoyl chloride, after selecting the defined category of chemicals acting by **nucleophilic substitution of acyl halides (1)**, the user can select to retrieve only results on **skin sensitisation (2)**.

In this case, the skin sensitisation results will be retrieved for the 6 identified analogues and a data-matrix will be built with the retrieved structures and the retrieved experimental results (see screenshot below).

Example 2: Gathering data for "structural" analogues

Continuing the example of a simple analogue approach of Ames mutagenicity for n-hexanal, after selecting the same defined category of chemicals **[35] Aldehydes <AND>…(1)**, the user can select to retrieve only results on **Ames_Mutagenicity (2)** (see screenshot below).

In this case, the Ames mutagenicity results will be retrieved for the initially defined 35 analogues. After subcategorization (see section 3.1 above) 9 identified analogues will remain and a data-matrix will be built with the retrieved structures and the retrieved experimental results (see screenshot below).

3.3. Evaluation of available data for adequacy

The evaluation of the available data for adequacy is a manual process. As described in Chapter 2, the resident databases of the Toolbox have been donated by several stakeholders. They have not undergone any modification and they have not undergone any quality assurance by the OECD. Users of the Toolbox have therefore to decide on their own whether the data used for read-across is appropriate and adequate.

The evaluation of available data for adequacy is nevertheless facilitated by providing background information on the results (see Chapter 2). Guidance for determining the quality of data can be found in section 3.1 of the OEDC Manual for Investigation of HPV Chemicals (OECD, 2008).

3.4. Construction of the matrix of data availability

The Toolbox automatically constructs the data matrix as illustrated in the two examples above (see section 3.2), with the data from the databases which were selected in the module **Endpoints**. No manual intervention is necessary.

3.5. Assess the adequacy of the analogue approach and fill the data gap

Assessing the adequacy of the analogue approach for the endpoint of interest

In the Toolbox the operations of filling the data gap and assessing the adequacy of the analogue approach for the endpoint of interest are performed in parallel. The module

 gives access to three different data gap filling tools:

- Read-across

- Trend analysis

- (Q)SAR models

Read-across and trend analysis use the available results of the analogues to fill the data gap. "(Q)SAR models" gives access to a library of external (Q)SAR models which have been integrated into the Toolbox. Depending on the situation, the most relevant data gap filling method should be chosen, taking into account the following considerations:

- Read-across is the data-gap filling method for "qualitative" endpoints like skin sensitisation or mutagenicity for which a limited number of results are possible (e.g. positive, negative, equivocal). Furthermore read-across is recommended for "quantitative endpoints" (e.g., 96h-LC50 for fish) if only a low number of analogues with experimental results are identified.

- Trend analysis can also be considered for filling data gaps for "quantitative endpoints" (e.g., 96h-LC50 for fish) if a high number of analogues with experimental results are identified.

- "(Q)SAR models" can be used to fill a data gap, independent of the analogue approach. They can also be used to provide additional support for a read-across or trend.

Before applying the read-across or trend-analysis option, the Toolbox allows the user to decide which results should be used for read-across in cases where more than one result is available for any analogue, e.g., all values, average values, minimum or maximum results.
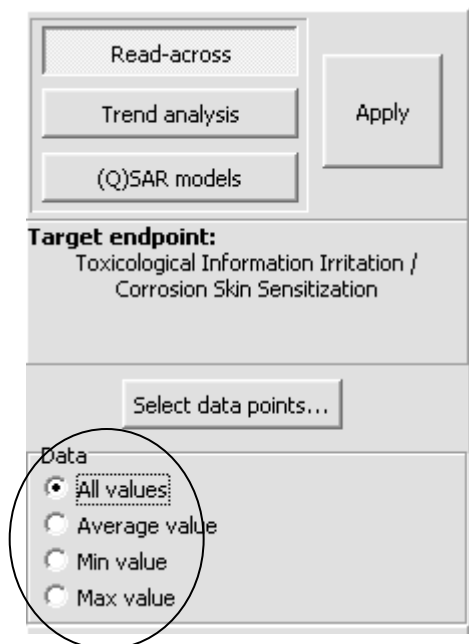
It should be noted that the **Average value** option is only functional for "quantitative endpoints".

The OECD Guidance on Grouping of Chemicals recommends that wherever possible, the adequacy of the analogue approach should be evaluated in the light of the known or suspected mechanism or mode of action. Here again it has to be distinguished between two situations:

1.   If a specific mechanism or mode of action relevant for the endpoint is identified for the target chemical(s), than all the analogues considered should have the same mechanism or mode of action.

2.   If <u>no</u> specific mechanism or mode of action relevant for the endpoint is identified for the target chemical(s), it is difficult to give precise guidance. It can be recommended that structural analogues for which a specific mechanism or mode of action is recognised should not be considered for read-across. Nevertheless this approach might not be sufficient as the remaining analogues could exert excess toxicity by specific mechanisms which are not yet recognised by the Toolbox. More weight should therefore be put on assessing the adequacy of the analogue approach by comparison of physical-chemical properties and other endpoints (see below).

The two cases are again illustrated with the two examples below.

The adequacy of the analogue approach can be assessed by verifying the consistency of the available experimental results for other endpoints for the target and source chemicals. This can be done by retrieving all available experimental results from the resident databases [see also Chapter 2 for further information on data retrieval] or by applying (Q)SAR models to the target and source chemicals. Version 1.1 contains a number of databases covering several endpoints. All available results can be included in the data matrix in

the module .

To add all available experimental results from the resident **Databases (1)** into the data matrix, it is sufficient to select all the relevant databases. You can then press the **Gather data** button **(3)** to retrieve the data.

The results will then be inserted into the data matrix and the user can then compare the available test results for other endpoints and evaluate the robustness of the analogue approach.

NOTE: The option of retrieving all available data on the analogues has to be chosen before moving to the module **Category definition** .

Example 1: Filling a data gap and assessing the adequacy of analogues, based on a specific mechanism or mode of action

If the identification of analogues was performed according to a specific mechanism or mode of action, the target and source chemicals will already have the same relevant mechanism and mode of action. The workflow illustration for the example chemical 4-nitrobenzoyl chloride (CAS No 122-04-3) is continued below.

As shown above, the Toolbox has identified six additional chemicals which have the same protein binding mechanism as the target chemical and for which experimental results are available. As sensitisation is a "qualitative" endpoint, **All values (1)** can be used for the evaluation and the data gap can be filled by **Read-across (2)** (see screenshot below).

The resulting graph plots the existing experimental results of all analogues (Y axis) against a descriptor (X axis). The default descriptor is log Kow. The dark red dots on the graph represent the experimental results available for the analogues and which are used for the read-across. The blue dots on the graph represent the experimental results available for the analogues and which are not used for the read-across, because they are furthest away from the target chemical, based on the X-axis descriptor. The red dot represents the estimated result for the target chemical based on the read-across from the analogues.

In this particular example, all results of the analogues are consistent. They all present a high sensitising potential. The same high sensitising potential is therefore also predicted for the target chemical. By default, the Toolbox averages the result of the 5 "nearest" analogues[2] (as defined by the X-axis descriptor) to estimate the result for the target chemical. The user can change those settings, as appropriate, e.g. choosing to base the read-across on the maximum or minimum value found for the analogues and/or by changing the number of analogues to be taken into account. It should be noted that averaging the results for the analogues is only functional for quantitative endpoints. For the present example, the read-across option should be set to either maximal or minimal.

---

[2] In this example only four dark red dots appear in the graph as two of the chemicals have the same estimated Kow and therefore appear as one dot on the graph.

The user can then verify the mechanistic robustness of the analogue approach. As indicated above, if the identification of analogues was performed according to a specific mechanism or mode of action, the target and source chemicals will already have the same relevant mechanisms and modes of action. Nevertheless, the source chemicals may also have additional mechanisms and modes of action due to additional functional groups of the molecule. This can be verified by "subcategorising" the list of identified analogues. When the user presses the **Subcategor.** button **(1)** and selects in the pop-up window the **Protein Binding** profiler **(2)** it is revealed that all six analogues have the same protein binding mechanism as the target molecule. This procedure indicates that no analogues need be deleted on a mechanistic basis as all 6 chemicals are **Nucleophilic substitution of acyl halides**. This increases the confidence that the analogues identified are appropriate.

Furthermore, if relevant, the user can then identify those chemicals within the list of analogues that have the <u>lowest structural similarity</u> to the target chemical, so that they can be removed from the list of

analogues if relevant. The tools for determining structural similarity are described in detail in part III of the Reference Manual for the Toolbox.

Following the same procedure as above, but using the profiler **Structural similarity (2)**, it is revealed that there is only one chemical that has a high structural similarity with the target compound. The user can decide whether to keep only this source chemical for the read-across and **Remove (3)** all others or whether to keep also the other analogues.

Once the user is satisfied that the analogue approach is sufficiently robust, both from a point of view of mechanism or mode of action and structural similarity, the prediction can be accepted and the result is inserted into the data matrix.

In the data matrix, the entry is preceded with an "S" for results estimated with the Toolbox compared to "T" for experimental results.

Example 2: Filling a data gap and assessing the adequacy of analogues based on structural similarity

The workflow illustration for the example chemical n-hexanal (CAS No 66-25-1) is continued below.

As shown above, the Toolbox has identified nine structurally similar chemicals based on organic functional groups. As Ames mutagenicity is a "qualitative" endpoint, **All values (1)** can be used for the evaluation and the data gap can be filled by **Read-across (2)** (see screenshot below).

The resulting graph plots the existing experimental results of all analogues (Y axis) against a descriptor (X axis). The default descriptor is log Kow. The dark red dots on the graph represent the experimental results available for the analogues and which are used for the read-across. The blue dots on the graph represent the experimental results available for the analogues and which are not used for the read-across, because they are furthest away from the target chemical, based on the X-axis descriptor. The red dot represents the estimated result for the target chemical based on the read-across from the analogues.

In this particular example, all results of the analogues are consistent. They all present a negative (-1) outcome in the Ames test. The same absence of mutagenicity in the Ames test is therefore also predicted for the target chemical. By default, the Toolbox averages the result of the 5 "nearest" analogues (as defined by the X-axis descriptor) to estimate the result for the target chemical. The user can change those settings, as appropriate, e.g. choosing to base the read-across on the maximum or minimum value found for the analogues and/or by changing the number of analogues to be taken into account. It should be noted that averaging the results for the analogues is only useful for quantitative endpoints. For the present example, the read-across option should be set to either maximal or minimal.

The user can then verify the mechanistic robustness of the analogue approach. As indicated above, if the identification of analogues was performed according to structural similarity, it is recommended to streamline the category by eliminating chemicals which have the potential to act by a specific mechanism or mode of action relevant for the endpoint of interest. This can be done by "subcategorising" the list of identified analogues. When the user presses the **Subcategor.** button **(1)** and the **DNA binding (2)** grouping method is selected in the pop-up window it is revealed that no specific DNA binding mechanism is identified for any of the analogues. This procedure indicates that no analogues need be deleted on a mechanistic basis and increases the confidence that the analogues identified are appropriate. The pop-up window can be closed **(3)** (see screenshot below).

Once the user is satisfied that the analogue approach is sufficiently robust, both from a point of view of mechanism or mode of action and structural similarity, the prediction can be accepted and the result is inserted into the data matrix.

In the data matrix, the entry is preceded with an "S" for results estimated with the Toolbox compared to "T" for experimental results.

3.6. Document the analogue approach

The first version of the Toolbox does not provide a function for drafting a report for the analogue approach according to the OECD Guidance on Grouping. However, in the module **Report (1)**, the Toolbox provides all the data and documentation needed to fill the Reporting Format as outlined in Chapter 7 of the OECD Guidance on Grouping. The study history of the work session is generated by pressing the button **Show history (2)**. This study history can be copied and appended to a report.

3.7. Using the metabolic profilers to refine an analogue search

The Toolbox provides many tools to refine an analogue search for read-across. One of those is metabolic profilers. The example is illustrated by filling a data gap for fish acute toxicity for 2-ethyl-4-methoxyphenol. In this example, the target chemical is entered using the **SMILES/InChi** writer **(1)**. At the same time the 2D structure is shown **(2)**; as the structure is correct the **OK** button is pressed **(3)** (see screenshot below).
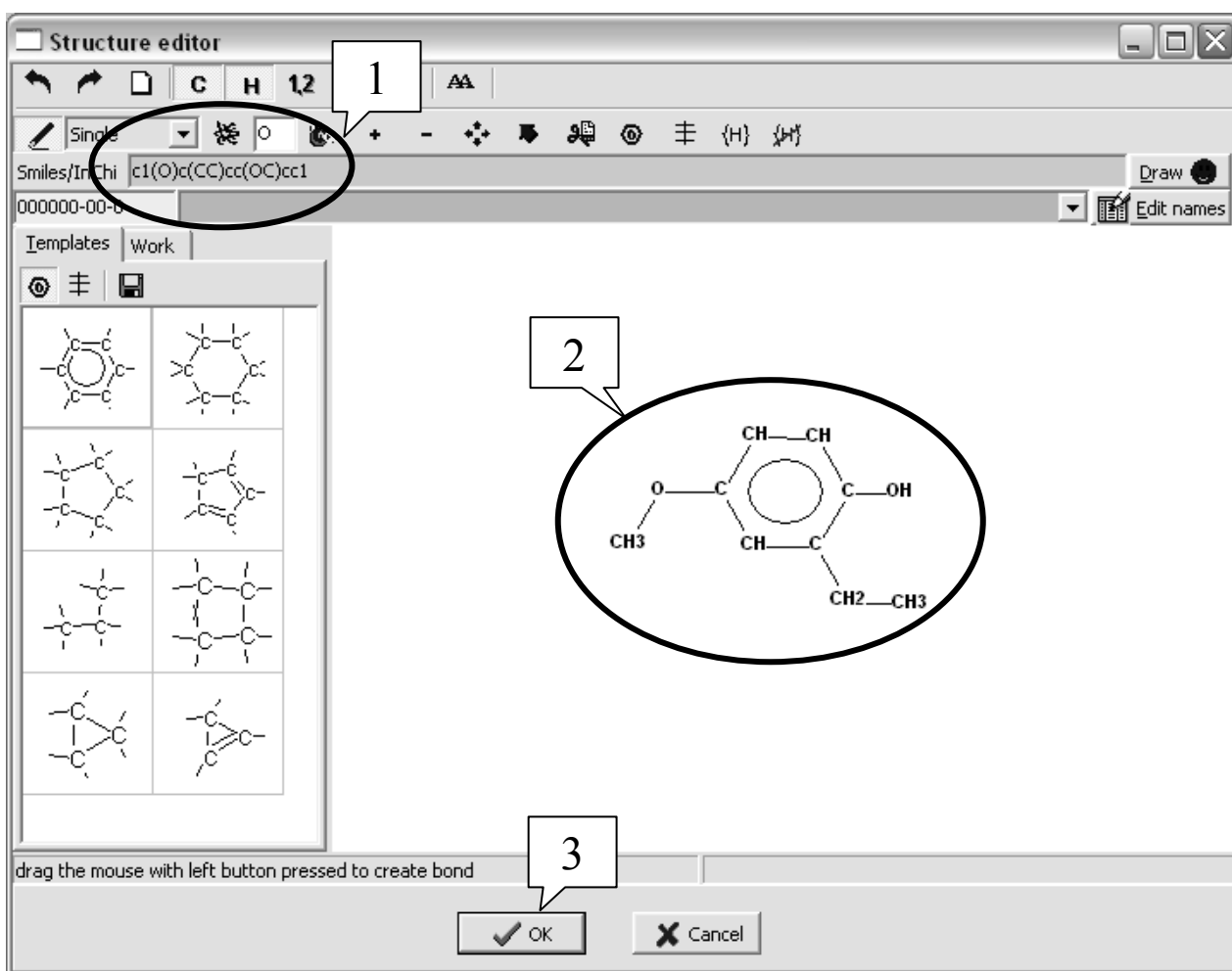
In this case there is very little substance information available (see screenshot below).

NOTE: Chemicals for which neither name nor CAS number is retrieved are not listed in any of the databases or inventories implemented in the Toolbox. This means therefore also that no experimental results on those chemicals will be retrieved via the Toolbox.

The next step is to **Profile (1)** the target chemical; there are a number of **Mechanistic** profilers that are relevant to aquatic toxicity **(2)**. With this particular target molecule, a phenol, we are also concerned about metabolism so the **Liver metabolism simulator (3)** is also selected before pressing the **Apply** button **(4)**, the Toolbox inserts the results into a data matrix **(5)** (see screen shot below).

The profiling results **(5)** (see above screenshot) reports that the target structure is a **Phenol**, which does not bind to either DNA or protein. However, the **Liver metabolism simulator** identifies reactive metabolites that are captured by the **EcoSAR Classification**, **OASIS Acute Toxicity MOA** and **Protein Binding** profilers. This profiling strongly suggests that metabolism needs to be considered in refining the analogue search.

The Toolbox provides further information on the metabolites. Clicking on the box under the target compound for **Protein Binding** of the metabolite **(1)** and **Profiling results** drops down **(2)**. By highlighting and clicking on a given metabolite **(3)** and second drop down box, **Profile Explainer**, appears **(4)**, which notes the particular mechanism of that metabolite, in this case **Michael-type nucleophilic addition (5)** (see screenshot below). While no data is available for the target chemical, it should be possible to identify other chemicals that have the same mechanism of this metabolite (Michael-type nucleophilic addition), for which experimental data is available.

The next step under **Endpoints (1)** is to select from the resident databases the three that contain data on fish acute toxicity **(2)** (see screenshot below).

The query can be launched under **Category definition (1)**. In this case it is done by selecting the **ECOSAR Classification (2)** and pressing the **Defining Category** button **(3)**. However, rather than selecting the ECOSAR Classification of the target compound (**Phenol**) **(4)** the ECOSAR Classification of the metabolite of the target compound (**Quinone/Hydroquinone**) is selected **(5)** and the **OK** button is pressed **(6)** (see screenshot below).

Note that you will be alerted in a **Warning** box that you have selected different from the target category (see screenshot below). Click the **Yes** button.
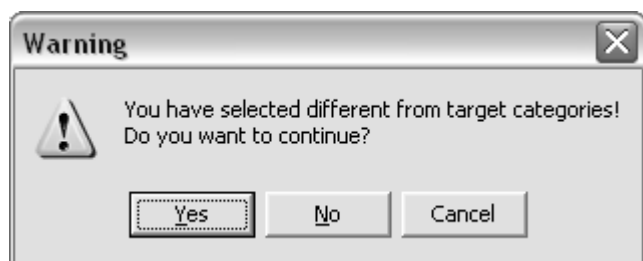
This brings into view the **Define category name** box (see screenshot below). Clicking the **OK** button defines the category.
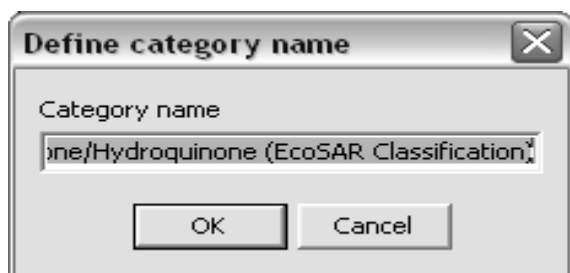
Highlighting **[50]Quinone/Hydroquinone (ECOSAR Classification) (1)** brings the dropdown box **Read data?** into view **(2)**. The radio button **Choose** should be selected **(3)** and under **Fish** the **Pimephales promelas**, L**C50** and **96hr** boxes checked before clicking the **OK** button **(4)**.
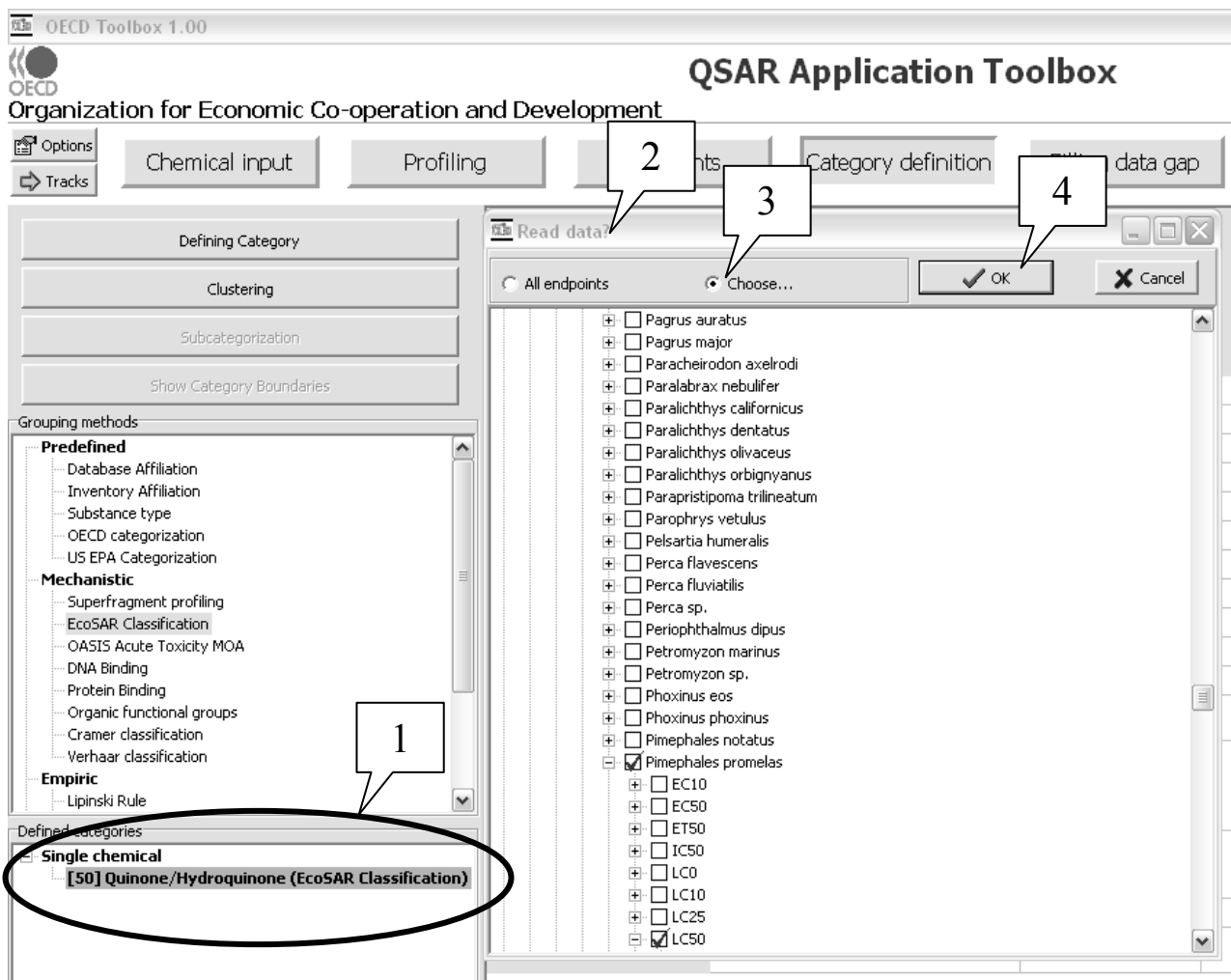
For the 50 chemicals **Quinone/Hydroquinone ECOSAR classification (1)** six chemicals have measured **96hr LC50 data for Pimephales promelas (2)**. This is enough data to read-across to predict the toxicity of the likely metabolite of the target compound (see screenshot below).

The next step is to **Fill data gap (1)** using **all values** (2) by **Read-across (3)** and clicking the **Apply** button **(4)**.

The results show five measured log LC50 values of just over 6.0 log (1 mole-1) **(1)** with one chemical with lower reported log LC50 values of just under 4.0 log (1 mole^{-1}) **(2)**.
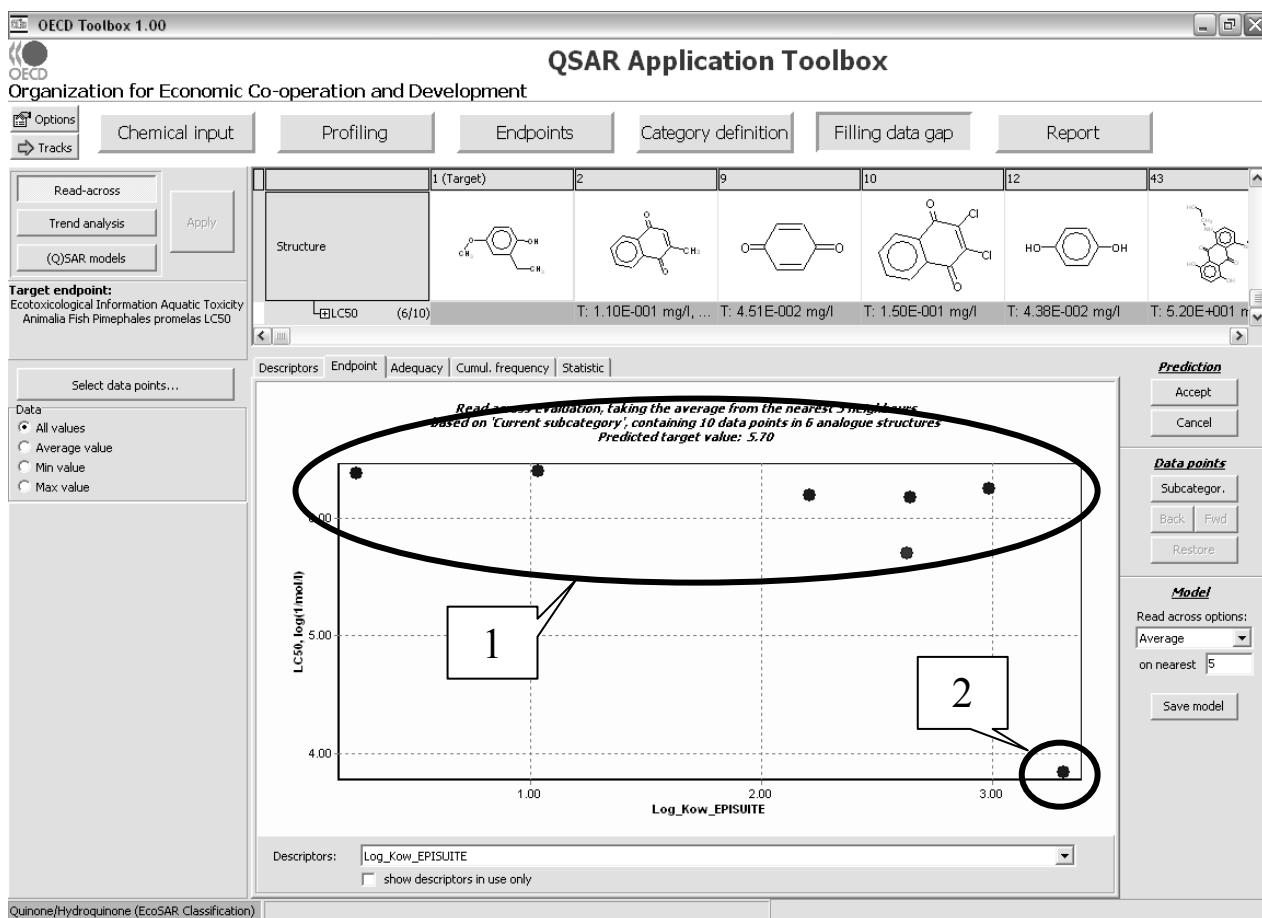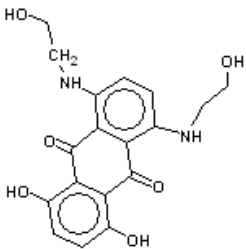
Double clicking on **(2)** in the above screenshot reveals further information on this chemical (see screenshot below). This substituted anthracenedione can be considered to be the least structurally similar of the six analogues to the target chemical, just by visual comparison. It also contains numerous ionizing groups that will impact bioavailability and thus aquatic toxicity. One would be justified in deleting this chemical from the evaluation resulting in a predicted target value of 6.2 log (1 mole^{-1}).
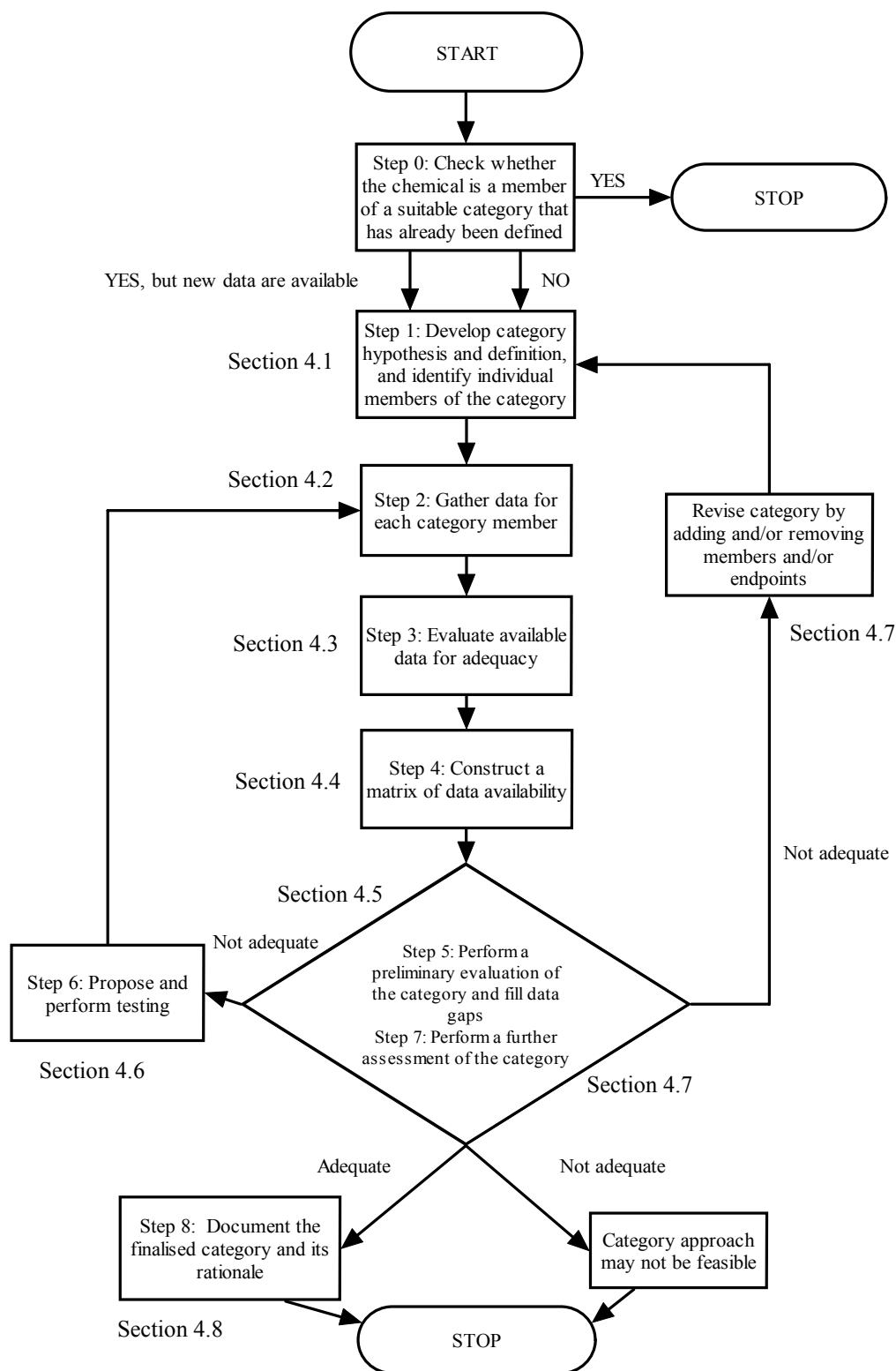
CHAPTER 4. BUILDING A TYPICAL CHEMICAL CATEGORY

The OECD Guidance on Grouping of Chemicals recommends a stepwise procedure to category development (see figure 2). The following sections indicate how the Toolbox can be used to perform the different steps. Simple examples are used to illustrate the different steps.

It is assumed that an assessor will start out with one or several closely related chemicals for which a category with many more members is to be developed. Before performing the operations outlined below, the user has to enter the chemical identity of the chemical(s) according to Chapter 2 of this guidance document.

Figure 2. Stepwise approach to category development

```
                              ┌──────────────┐
                              │    START     │
                              └──────────────┘
                                     │
                                     ▼
                        ┌─────────────────────────┐
                        │ Step 0: Check whether    │      YES    ┌──────────┐
                        │ the chemical is a member │ ─────────▶  │   STOP   │
                        │ of a suitable category   │            └──────────┘
                        │ that has already been    │
                        │ defined                  │
                        └─────────────────────────┘
         YES, but new data are available          NO
                        │                          │
                        ▼                          ▼
                        ┌─────────────────────────┐
 Section 4.1            │ Step 1: Develop category │ ◀──────┐
                        │ hypothesis and           │        │
                        │ definition, and identify │        │
                        │ individual members of    │        │
                        │ the category             │        │
                        └─────────────────────────┘        │
                                     │                      │
                                     ▼                      │
                        ┌─────────────────────────┐  ┌──────────────────┐
 Section 4.2            │ Step 2: Gather data for  │  │ Revise category  │
                        │ each category member     │  │ by adding and/or │
                        └─────────────────────────┘  │ removing members │
                                     │                │ and/or endpoints │
                                     ▼                └──────────────────┘
                        ┌─────────────────────────┐       Section 4.7
 Section 4.3            │ Step 3: Evaluate         │
                        │ available data for       │
                        │ adequacy                 │
                        └─────────────────────────┘
                                     │
                                     ▼
                        ┌─────────────────────────┐
 Section 4.4            │ Step 4: Construct a      │
                        │ matrix of data           │
                        │ availability             │
                        └─────────────────────────┘
                                     │
                                     ▼
 Section 4.5              ◆ Step 5: Perform a preliminary
   Not adequate           evaluation of the category
 ┌──────────────┐         and fill data gaps           Not adequate
 │ Step 6:      │ ◀──────  Step 7: Perform a further  ──────▶
 │ Propose and  │         assessment of the category
 │ perform      │              Section 4.7
 │ testing      │
 └──────────────┘
   Section 4.6
         Adequate              Not adequate
 ┌──────────────┐         ┌──────────────────┐
 │ Step 8:      │         │ Category approach│
 │ Document the │         │ may not be       │
 │ finalised    │         │ feasible         │
 │ category and │         └──────────────────┘
 │ its rationale│                 │
 └──────────────┘                 │
   Section 4.8  │         ┌──────────────┐
                └───────▶ │    STOP       │ ◀──────┘
                          └──────────────┘
```

The first step is to check whether the chemical is a member of a suitable category that has already been defined. The Toolbox contains the category definitions of a number of categories that have (or are in the process of being) assessed in the OECD HPV Chemicals Programme.

By using the module **Profiling** a chemical for "OECD Categorisation", a user can check whether a chemical could fit one of those definitions. For example, by profiling 1-nonanethiol (CAS No 1455-21-6, the user will find that the chemical could be assessed as part of the category of C8-C12 Mercaptons which is in the process of being assessed in the OECD HPV Chemicals Programme (see below).

Alternatively, the user can decide to develop his/her own category definition (see section 4.1).

The United States Environmental Protection Agency's Office of Pollution Prevention and Toxics has developed Chemical Categories to facilitate pre-marketing notifications (i.e. new chemicals) reviews. While all new chemical substances are evaluated on a case-by-case basis using the most appropriate structural analogue to support any concerns for health or environmental effects, the 45 categories included in the "US EPA Categorization" compilation represent groupings of chemicals for which sufficient assessment experience has been accumulated by the Agency so that hazard concerns and testing recommendations vary little for chemicals that fit the category definition. These EPA categories are defined by a series of structural alerts acquired through years of experience. These categories are not intended to be a comprehensive list nor do they necessarily represent the chemicals of greatest concern to the Agency.

4.1. Develop category hypothesis and definition and identify category members

The Toolbox provides many tools to set up a category definition, e.g., structural, i.e. defining a category based on close structural similarity between its members; mechanism or mode of action, i.e. defining a category based on similar mechanisms or modes of action of its members; metabolic pathway, i.e grouping chemicals which are degraded into the same metabolite by e.g. the liver.
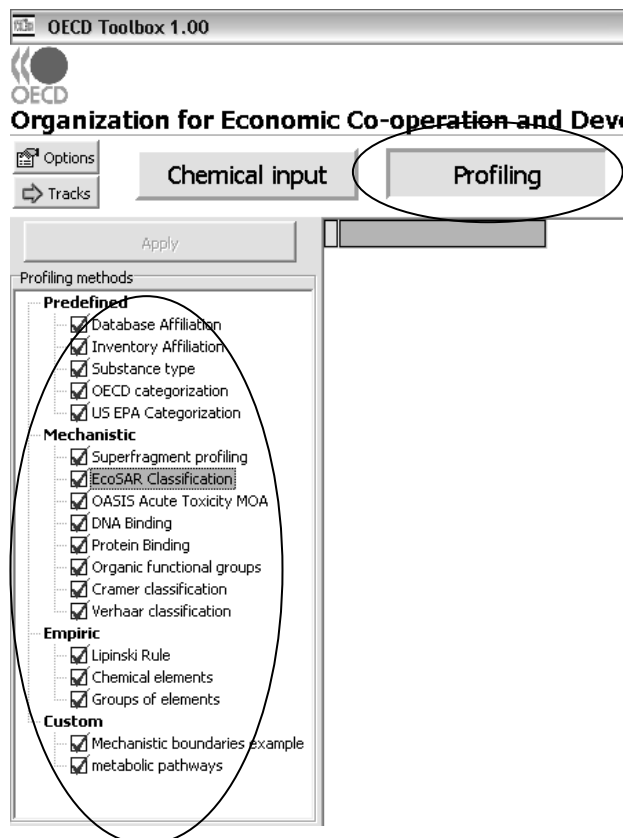
The first step is to profile the target chemical(s) with the module [Profiling]. The use of the profiler is described in more detail in section 3.1 of this guidance document.

Furthermore, as indicated in section 3.1 while all profilers provide valuable information on the target chemical(s), the mechanistic profilers, i.e. those retrieving information on the potential mechanisms or modes of action of the target chemical are often recommended for subsequently identifying potential analogues of the target chemical(s). Summary background information on these profilers is therefore outlined in Table 4 (see section 3.1).

NOTE: The outcome of the profiling determines the most appropriate way to set up the category definition. The following recommendations can be made:

- If specific mechanisms or modes of action are identified for a representative member of the category, then it is recommended to base the category definition on those mechanisms or modes of action. Members of the category can be found by searching for chemicals which have the same mechanisms or modes of action. The search results can then be refined by identifying among those the chemicals which have the closest structural similarity to the target chemical.

- If no specific mechanisms or modes of action are identified for a representative member of the category, then it is recommended to base the category definition on close structural similarity. Members of the category can be found by searching for chemicals which are structurally similar to the target chemical. The search results can then be refined by eliminating those chemicals which have specific mechanisms or modes of action.

This recommendation is based on experience gained by using the Toolbox and should not be regarded as restricting any other approach for identifying category members.

Nevertheless, it is not necessary that all chemicals in a category have the exact same mechanistic profile. It is indeed possible to build subcategories within a category. For example a category can have the same mechanistic behaviour relevant for one endpoint (and hence a consistent trend for that endpoint across the category) and one or two different mechanistic behaviours relevant for another endpoint (and hence different trends for that endpoint across the category). This can be illustrated with the example below.

Example: Using a mechanistic profiling result to develop a category definition

The example is illustrated by building a category around the chemical propanoic acid, 3-mercapto, butyl ester (CAS No 16215-21-7). After choosing the **Mechanistic** profilers **(1)** and pushing the **Apply** button **(2)**, the profiling results are inserted into the **Profile** matrix **(3)**. The most relevant profiling results for that chemical in terms of chemical reactivity and class/structure are that it can bind to protein by disulfide formation and that it belongs to two specific aquatic toxicity classes according to ECOSAR Classification ("Esters" and "Thiols(mercaptans)").

Based on this profile, the user can identify chemicals with the same profile. Before doing so, the user has to define the inventory they want to search. For example the user can choose to search for chemicals with the same profile in a national index like the US –TSCA inventory or EU EINECS or in more

restricted inventories like the OECD HPV list. This is done in the module **Endpoints** as outlined below.

NOTE: It is recommended that the user also select all databases containing experimental results. This implies that the Toolbox will also find chemicals with the same profile, for which experimental data is available, but which are not necessarily in the selected inventory.

These chemicals could be used as supporting chemicals. Furthermore, by selecting all the databases containing experimental results, these will be automatically retrieved and added to the data matrix, as requested in Steps 2 and 4.

The actual query can then be launched under **Category definition** .

> NOTE: Currently it is not possible to query directly by several profiling results in parallel. The user has first to query according to one profiler and then subcategorise the results step-by-step according to other profilers.

For this example, the user could first select the **Protein Binding** grouping method **(1)** of the target chemical and query for all the chemicals with the same mechanism **Disulfide formation (2)** in the selected inventory and databases (see screenshot below).

Upon pushing the **OK** button the Toolbox will now retrieve those chemicals that have the same protein binding mechanism as the target compound and which are either in the selected inventory or for which experimental data is available in the previously selected databases. In this example, 80 additional structures were retrieved **(1)** (see screenshot below).

The user can then conduct a further **Subcategorization (2)** the results according to the **ECOSAR Classification** grouping method **(3)**.

As the **Target** chemical belongs to two EcoSAR classifications **(4)**, the user can decide to eliminate all chemicals which do not belong to both these classes by selecting the **All categories** button **(5)**, in which case only 5 additional **Analogues** are identified **(6)** pushing the **Prune** button **(7)** remove the other analogues. In this case, the experimental results will be retrieved for all the 6 category members and a data-matrix will be built with the retrieved structures and the retrieved experimental results.

In this example, the retrieved chemicals have identical mechanistic profiles (except for Organic functional groups). The number of chemicals retrieved is therefore rather low. It could therefore be considered to expand the category by building subcategories. For example, the user could decide to build a category with the same protein binding mechanism but allowing chemicals belonging to either one of the two ECOSAR classes.

The same sequence of actions as previously described can be performed, except that the button **At least one category** is selected **(1)** (see next screenshot).

Pushing the **Prune** button **(2)** results in a subcategory of 31 chemicals (see next screenshot).

NOTE: After identifying category members according to a specific mechanism or mode of action it is always necessary to verify whether any of the selected chemicals do not have additional mechanisms or modes of action which would make it unsuitable for the category. This can be done by using the "Subcategorisation" procedures outlined below.

For example there could be chemicals that have specific DNA binding mechanisms, due to additional functional groups in the molecule. The same sequence of actions as previously described can be performed, but using the **DNA Binding** profiler **(1)** (see screenshot below).

In this particular example, no outliers in terms of mechanism or mode of action are identified and no additional chemicals have to be eliminated from the category.

The result is a group of chemicals that can bind to protein by the same mechanism (disulfide formation) and that belong to either the ECOSAR class(es) of "Thiols (mercaptans)" (1) or "Thiols (mercaptans)" and "Esters" (2) (see screenshot below). Chemicals with other specific mechanisms or modes of actions have been eliminated.

It is therefore expected that these chemicals have a very similar behaviour for many regulatory endpoints. For aquatic toxicity, it is expected that differences in trends could be observed between chemicals belonging to the EcoSAR categorization class(es) of "Thiols (mercaptans)" or "Thiols (mercaptans)" and "Esters" and therefore these should be considered as two subcategories.

4.2. Gather data for each category member

The Toolbox automatically retrieves the data available in the previously selected databases. No manual intervention is necessary. The data retrieval will occur with the first category that is defined. By selecting the category, the Toolbox will request the user to select the endpoint for which experimental results should be retrieved. The user can either select specific endpoints or by default choose to retrieve data on all endpoints. See also Chapter 2 on data retrieval.

NOTE: Data will only be retrieved from those databases which have been selected previously under **Endpoints**

As the category is further subcategorised, the data matrix will be pruned. No further data retrieval steps are necessary.

Example: Gather data for a "mechanistic" category

Continuing the example of the "mechanistic" category, the user can retrieve all the data from the previously selected databases by selecting the **[81] Disulfide formation (Protein Binding)** category **(1)** and choosing "**All endpoints**" in the "**Read data?**" window **(2)** (see screenshot below).
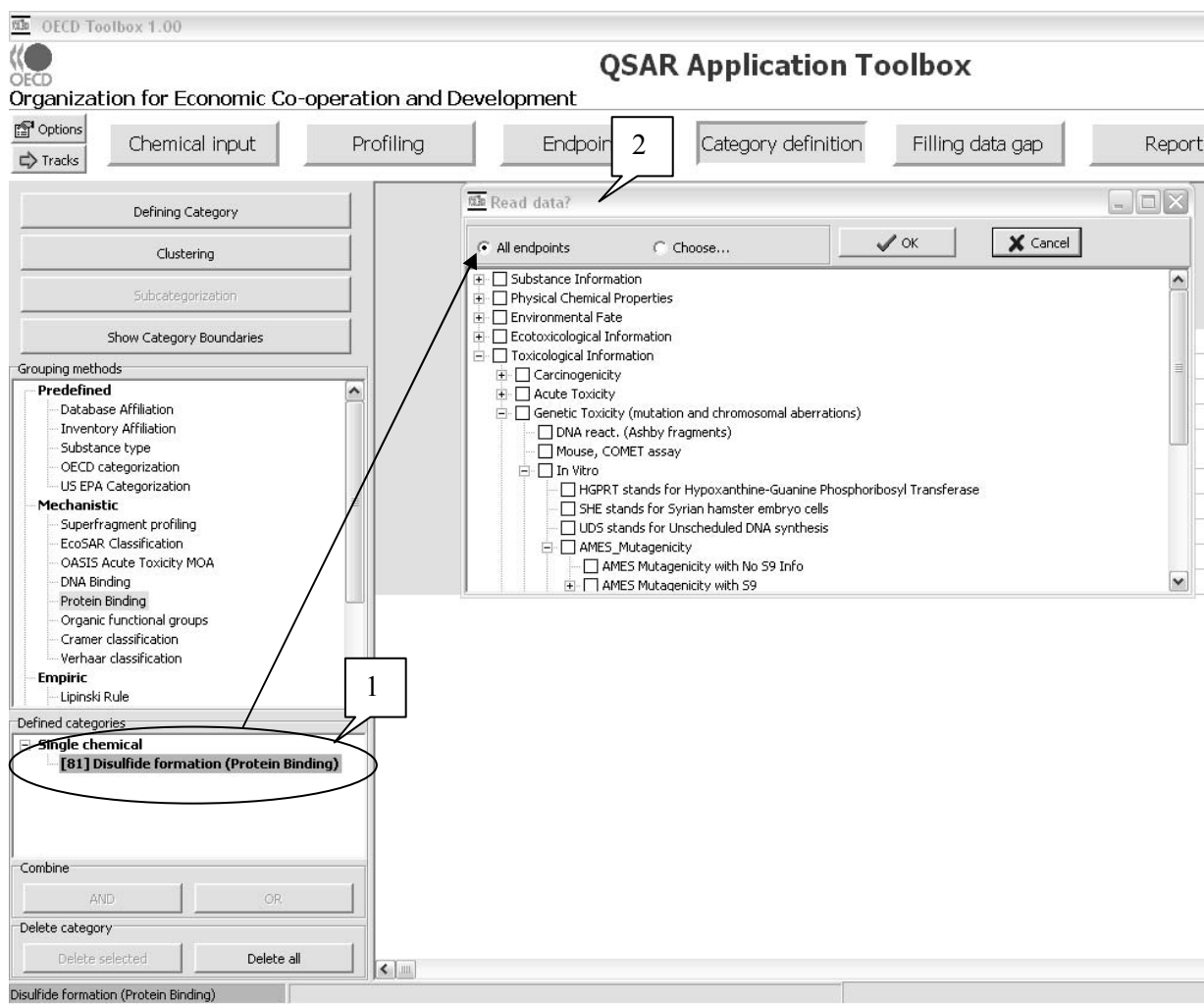
After deselecting multiple entries, the available results are entered into the data matrix (see screenshot below.

4.3. Evaluate available data for adequacy

The evaluation of the available data for adequacy is a manual process. As described in Chapter 2, the resident databases of the Toolbox have been donated by several stakeholders. They have not undergone any modification and they have not undergone any quality assurance by the OECD. Users of the Toolbox have therefore to decide on their own whether the data used for read-across is appropriate and adequate.

The evaluation of available data for adequacy is nevertheless facilitated by providing background information on the results (see Chapter 2). Guidance for determining the quality of data can be found in section 3.1 of the OEDC Manual for Investigation of HPV Chemicals (OECD, 2008).

4.4. Construct a matrix of data availability

The Toolbox automatically constructs the data matrix, as illustrated in the two examples above (see section 4.2), with the data from the databases which were selected in the module **Endpoints**. No manual intervention is necessary. No manual intervention is necessary.

4.5. Perform a preliminary evaluation of the category and fill the data gap

To evaluate a category, the OECD Guidance on Grouping of Chemicals recommends

1.   to determine whether the category does in fact exhibit one or more of the trends postulated in step 1 and,

2.   to determine whether the category contains sufficient relevant and reliable information on the category members.

In addition, the use of "external" (Q)SAR models may be used to provide additional support for a trend.

In the Toolbox the operations of filling the data gap and assessing the adequacy of the category are performed in parallel. The module gives access to three different data gap filling tools:

- Read-across

- Trend analysis

- (Q)SAR models

Read-across and trend analysis use the available results for the analogues to fill the data gap. "(Q)SAR models" gives access to a library of external (Q)SAR models which have been integrated into the Toolbox. Depending on the situation, the most relevant data gap filling method should be chosen, taking into account the following considerations:

- Read-across is the data-gap filling method for "qualitative" endpoints like skin sensitisation or mutagenicity for which a limited number of results are possible (e.g. positive, negative, equivocal). Furthermore read-across is recommended for "quantitative endpoints" (e.g., 96h-LC50 for fish) if there is only a low number of category members with experimental results.

- Trend analysis can also be considered for filling data gaps for "quantitative endpoints" (e.g., 96h-LC50 for fish) if there is a high number of category members with experimental results.

- "(Q)SAR models" can be used to fill a data gap of the category approach. As indicated above, they can also be used to provide additional support for a read-across or trend.

Before applying the read-across or trend-analysis option, the Toolbox allows the user to decide which results should be used for read-across in cases where more than one result is available for any category member, e.g., all values, average values, minimum or maximum results.

It should be noted that the **Average value** option is only functional for "quantitative endpoints".

How to assess the category is illustrated with the two examples below.

Example: Filling a data gap and assessing the adequacy of the category, based on a specific mechanism or mode of action

The workflow illustration for the example category around the chemical propanoic acid, 3-mercapto, butyl ester (CAS No 16215-21-7) is continued below.

As shown above, the Toolbox has identified 31 chemicals which belong to the category. Unfortunately the amount of available experimental results is very limited for this category. The statistics are indicated on the left of the data matrix (see screenshot below: 1 chemical for environmental fate, 6 chemicals for skin sensitisation etc.).

In this particular example, only very few experimental results are available across the data matrix.

Nevertheless, for some endpoints, the coherence and consistency of the available data can be assessed. For example, regarding Ames mutagenicity, the Toolbox has identified four chemicals for which results are available. The adequacy of the category for this endpoint can be tested by trying to use the available data for read-across to any of the chemicals for which no experimental data is available. As Ames mutagenicity is a "qualitative" endpoint, the data gap can be filled by read-across.

To do so, the user should highlight an empty cell in the line **Ames_mutagenicity** in the data matrix, select the radio button **All values (1)**, push the **Read-across** button **(2)** and then push the **Apply** button **(3)** (see screenshot below).

The resulting graph plots the existing experimental results of all analogues (Y axis) against a descriptor (X axis). The default descriptor is log Kow. The dark red dots on the graph represent the experimental results available for the category members and which are used for the read-across. The red dot represents the estimated result for the target chemical based on the read-across from the category members.

In this particular example, all four of the results of the category members are consistent. They are negative (-1) in the Ames test. The available results for Ames mutagenicity therefore confirm the adequacy of the category for this endpoint.

The same exercise can be performed for skin sensitization.
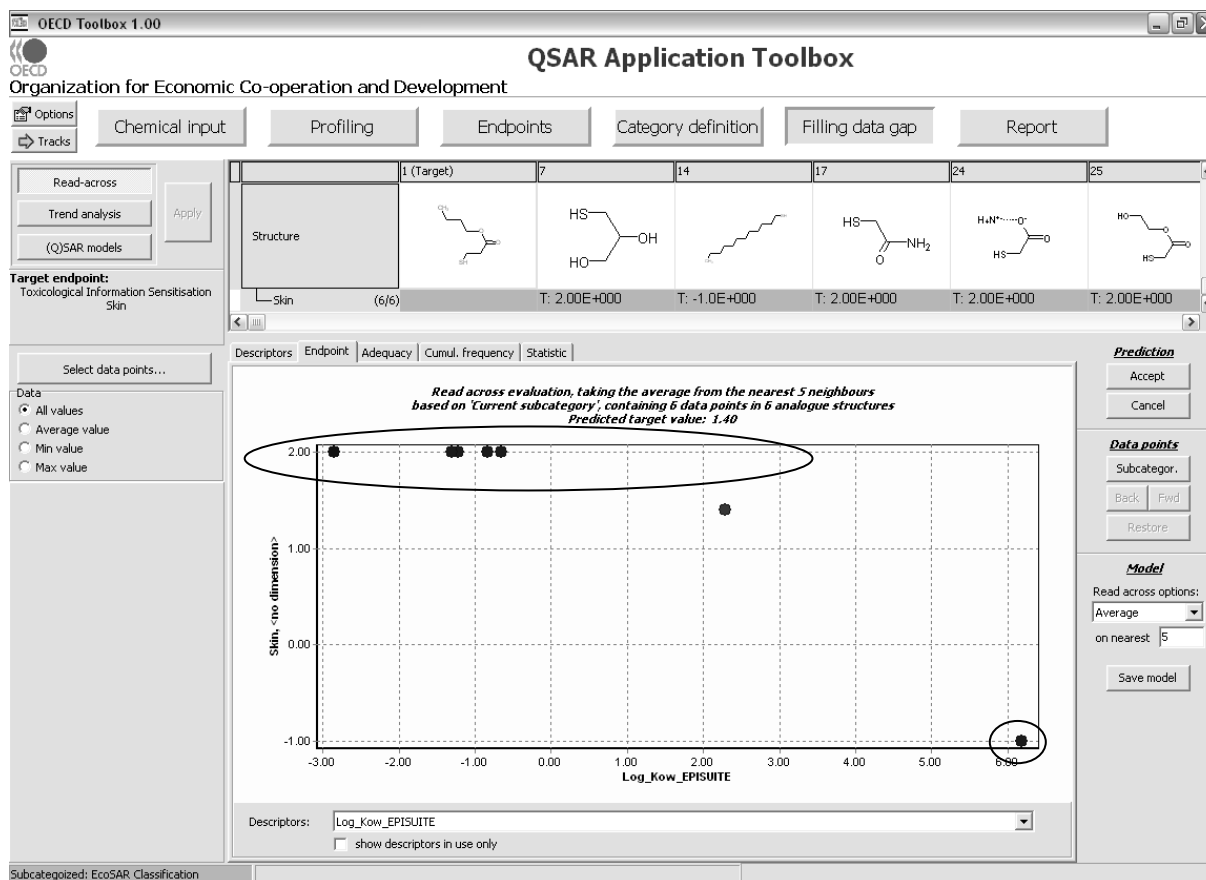
In this case all available skin sensitization results are positive, except for one. The negative result has been found for dodecyl mercaptan [$C_{12}H_{26}S$]. The possible explanation is that the bioavailability of this molecule is too low (Log Kow > 6) to induce sensitisation. Overall the limited available experimental results confirm the adequacy of the category for this endpoint.

The same exercise as outlined above can be done with any endpoint for which results are available for more than one member of the category.

As indicated above, the robustness of the category could be further evaluated with the help of external QSARs from the Toolbox library of QSARs.

To access the available models for a given endpoint, highlight a cell in the matrix for a given endpoint, which in this example is **Sensitisation skin (1)** and click on **(Q)SAR models (2)**. The list of available QSAR models will then appear in the box **QSAR models (3)** (see screenshot below).

Before applying a QSAR model it is recommended to consult its documentation. To access the background information of a model, push the button **Ranking (4)** (see screenshot above).

A window with summary information on the available models for that endpoint will appear (see screenshot below). Especially the QSAR model report according to the "QSAR Model Reporting Format" or QMRF will provide information on how the model was built an especially how it fulfils the OECD Principles for the Validation, for Regulatory Purposes, of (Q)SAR models (OECD 2007a).

Once the user is satisfied that the model can be used to evaluate the category, it can be applied to all the chemicals in the category and the results can be analysed. To apply the model simultaneously to all the chemicals in the category, select the model **(1)**, right-click upon it **(2)** and select **Predict Endpoint (3)** and **All chemicals in domain (4)** (see screenshot below).

The Toolbox then applies the model to all the chemicals in the category which are also in the applicability domain of the model. The estimations are inserted into the data matrix, preceded by the letter "S". Estimations are generated for 19 chemicals in the category (see screenshot below). Please note that the model only distinguishes between positive estimations (1) and negative estimations (-1), while the positive experimental results are separated into moderate sensitisers (1) and strong sensitisers (2).

Those chemicals which are in its applicability domain, are predicted to be skin sensitisers, with the exception of octadecanthiol (-1) and cyclohexanethiol (0 = ambiguous). Overall the model tends to confirm the evaluation of the category based on the available experimental data, namely that the chemicals in this category are probably skin sensitisers, except for some chemicals which are not bioavailable.

4.6. Propose and/or perform testing

The Toolbox does not provide any guidance as to how to set up a testing plan. For specific guidance see the OECD Guidance on Grouping of Chemicals (OECD, 2007b).

4.7. Perform a further assessment of the category / Revise category by adding and/or removing members and/or endpoints

If additional testing has been performed, the Toolbox can be used to further assess the category. New test results can be added to the database according to the guidance outlined in part IV of the Reference Manual of the Toolbox. The same approach as outlined in section 4.5 above can then be repeated.

4.8. Document the finalised category

The first version of the Toolbox does not provide a function drafting a report for the category according to the OECD Guidance on Grouping. Nevertheless, the Toolbox provides all the data and argumentation needed to fill the Reporting Format as outlined in Chapter 7 of the OECD Guidance on

Grouping. Furthermore the study history currently produced by the Toolbox can be copied and appended to a report.

CHAPTER 5. METABOLIC PATHWAY CATEGORIES

5.1. Introduction

According to the OECD Guidance on Grouping of Chemicals, the underlying hypothesis for a metabolic series is a sequential metabolism of a parent chemical to downstream metabolites that are chemicals of interest. Hazard identification studies with the parent compound could then be used to identify the hazards associated with the downstream primary and secondary metabolites and once quantified can be used in place of studies using direct exposure to primary and secondary metabolites themselves. In certain instances, the metabolism of the parent compound within barrier tissue (e.g. lung or gut tissue) occurs so rapidly that the initial primary metabolite is the predominant chemical found within the blood. Under these circumstances data from hazard identification studies conducted with that primary metabolite itself can be used to identify hazards for the parent compound.

The Toolbox contains a number of metabolism simulators as well as a metabolism database which allow the user to screen a list of chemicals and to identify all chemicals which potentially metabolise to a specific metabolite or which has been measured as metabolite.

> NOTE: The metabolism simulators in the Toolbox can only identify *potential* members of a metabolic pathway category. Indeed, they only identify potential metabolites, not stable metabolites.
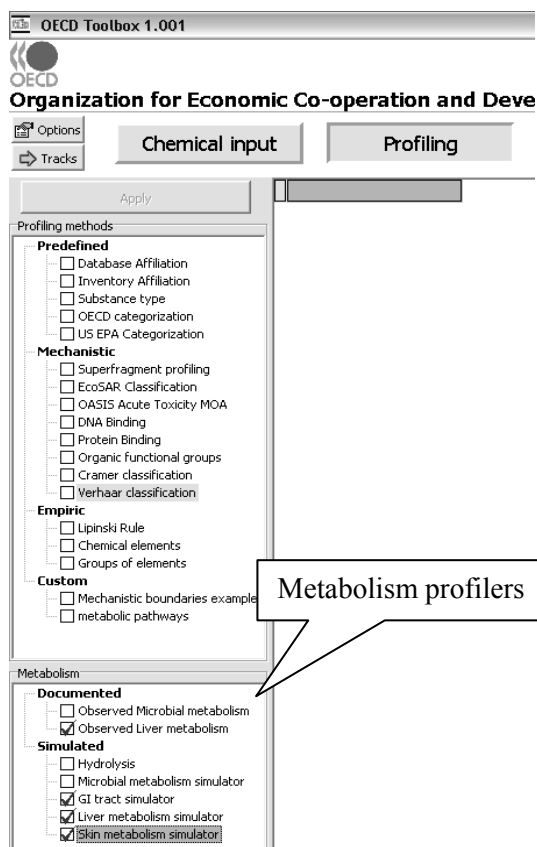
The following profilers are relevant for this workflow. These profilers are accessible via the module **Profiling** (see screenshot).

- Observed liver metabolism
  Metabolic pathways documented for 200 chemicals in different mammals are stored in a database format that allows easy computer access to metabolic information. The collection includes chemicals with variety of functionalities, aliphatic hydrocarbons, alicyclic rings, furans, halogenated hydrocarbons, aromatic hydrocarbons and haloaromatics, amines, nitro-derivatives, and multifunctional compounds. In vivo and in vitro (predominantly microsoms) studies were used to analyze the metabolic fate of chemicals. Different sources including monographs, scientific articles and public web sites were used to compile the database.

- GI tract simulator
  The simulator models chemical transformations that predominantly occur in the gastro-intestinal tract focusing on carbohydrate, lipid, amino

acid, and nucleotide metabolism. It includes 74 spontaneous and enzyme mediated reactions such as hydrolysis of alkaline salts, carbamates, acyl halides, oligophosphate, esters, amides, etc. Keto-enol tautomerism, ether hydrolytic cleavage, oxidative deamination, salvage reactions, and disulfide bond cleavage are also simulated.

- Liver metabolism simulator
  Multi-pathway modeling approach was used to simulate the metabolism in mammalian liver (Mekenyan et al., 2004). The scheme was conditioned by the fact that chemicals could be subject of variety of enzyme controlled reactions. Currently, 345 principal transformations are used to model metabolism in liver. They were separated into two major classes: non-rate determining and rate determining reactions. Transformations of highly reactive groups and intermediates such as acyl halide dehalogenation, geminal thiol halide dehalogenation, geminal halohydrine dehalogenation, N-nitrosoamine oxidative N-dealkylation, imide hydrolysis, sulfinic acid S-oxidation, etc. form the first class of molecular transformations. Various chemical equilibrium processes such as tautomerism are also included here. The second class includes oxidative, redox, reductive, hydrolytic and synthetic reactions. Initially, the target chemical is submitted to the list of hierarchically ordered transformations. All transformations meeting the associated substructures are implemented on the parent producing the first level of metabolites. Each of the obtained metabolites is further submitted to the same list of transformations, thus producing the second level of metabolites, etc.

- Skin metabolism simulator
  The core of the skin metabolism model is a set of 203 principal transformations defined on the basis of empirical and theoretical knowledge and hence peer-reviewed by human experts (Dimitrov et al., 2005). The principal transformations were separated into two major classes: non-rate determining and rate determining reactions. Non-rate determining included abiotic and biologically mediated transformations, which occur at very high rates. They also included nine non-rate determining reactions predominantly hydrolysis of salts. Eighty-three other transformations simulating the reactions of highly reactive groups and intermediates such as quinones, α,β-diketones, oxiranes, acyl halides, thiocarboxylic acids, hydroperoxides, nitrenes, geminal diols, etc., with cell proteins were also included.

To use the toolbox to identify potential members of a metabolic pathway category, the following steps can be identified:

1. Definition of the list of chemicals to screen.

2. Building a metabolic profiler.

3. Screening the list of chemicals for potential members of the category.

This sequence is elaborated below and illustrated with the example of identifying chemicals from the OECD HPV Chemicals List those which can potentially degrade to 2-methoxyethanol (CAS No 109-86-4) via liver metabolism.

5.2. Definition of the list of chemicals to screen

How to define target chemicals is outlined in Chapter 2 (section 2.1). The guidance from Chapter 2 is not repeated here. Simply the example of loading the OECD HPV Chemicals List is illustrated below.

To load the list of OECD HPV Chemicals in the module **Chemical Input**, the user can (1) simply press the button **Regulatory Inventories** and (2) select **OECD HPVC Inventory** and then (3) press **OK** (see screenshot below).
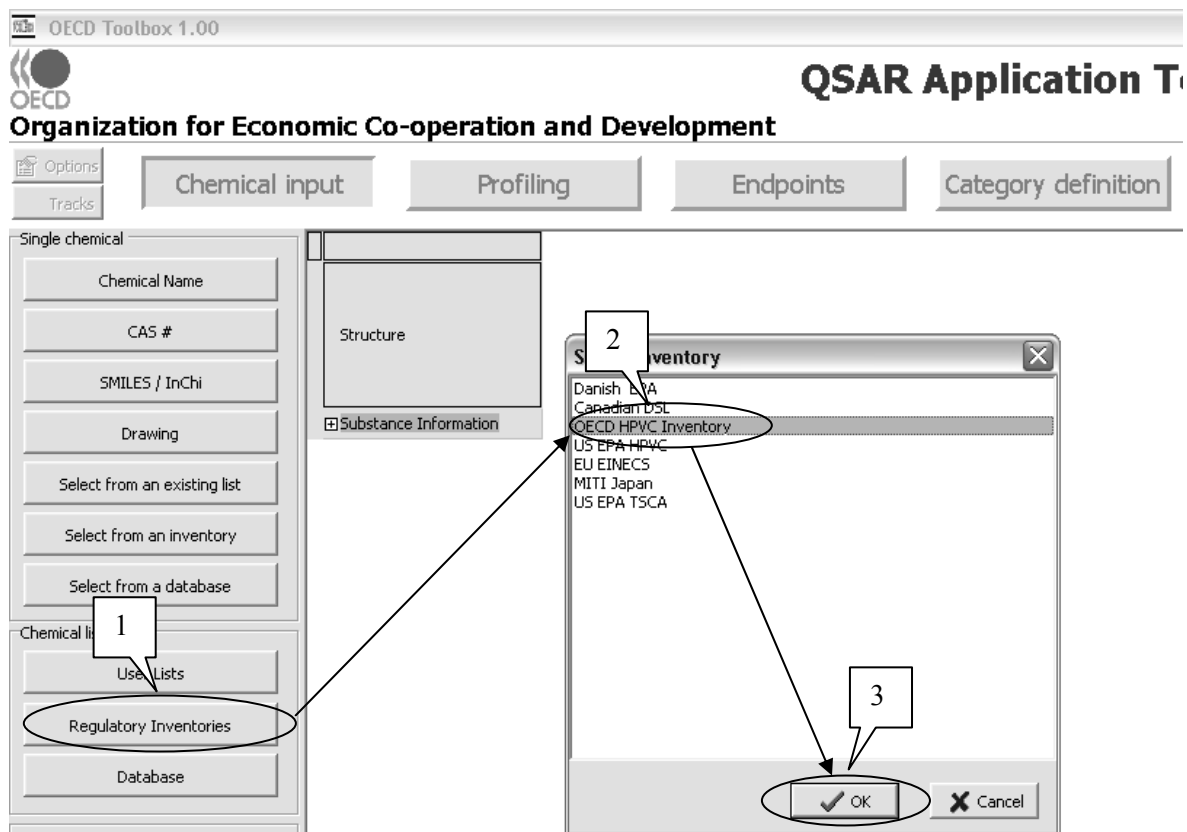
The 4843 chemicals from the OECD HPV Chemicals List will then be inserted into the data matrix.

5.3. Building a metabolic profiler

The target list of compounds can be screened by any of the available profilers of the module **Profiling**. The user can also add user-defined or **custom** profilers to the list of pre-existing profilers. Guidance is provided below on how to build a profiler that identifies chemicals which can potentially metabolise to a specific chemical. It is illustrated with building a profiler that identifies chemicals that can potentially degrade to 2-methoxyethanol (CAS No 109-86-4) via liver metabolism.

To build a new profiler in the module **Profiling (1),** the user should press the **Create new Profiler** button **(2)** and enter a name for the new profiler (see screenshot below). In this example the new profiler could be named "metabolic pathways".

The Toolbox then opens the profiling editor, in which the user can define the profiling criteria.

The first step in entering the profiling criteria is to add a new category by pressing the ⬚ button **(1)** and entering a category name **(2)**, e.g. methoxyethanol in this case.
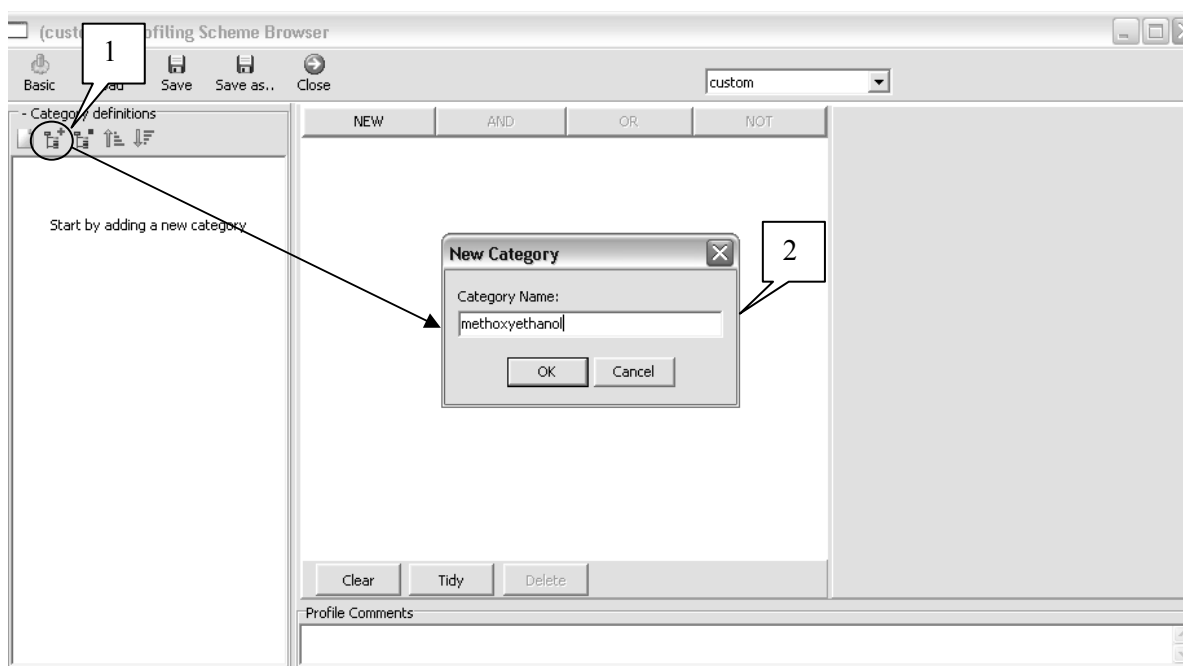
The user can then start to enter the category boundaries, by pressing the **New** button **(1)** and selecting **structural boundary (2)**, as in the current example the first step is to define the structure of the metabolite i.e. 2-methoxyethanol (see screenshot below).
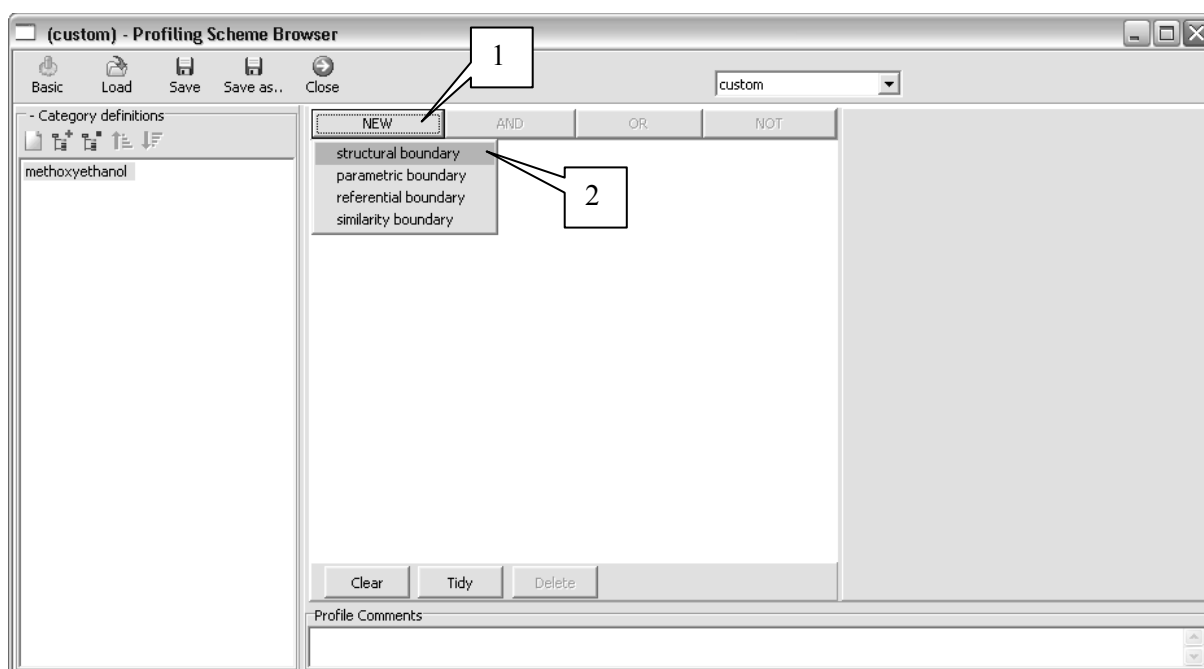
The pictogram **(1)** is then inserted into the middle window, representing the first set of boundaries and a new sub-window is opened on the right side where the user can define the boundaries. To define the structure of 2-methoxyethanol, the user has to press the **Edit** button **(2)** (see screenshot below). The user than selects **structure** to access the 2D structure editor.
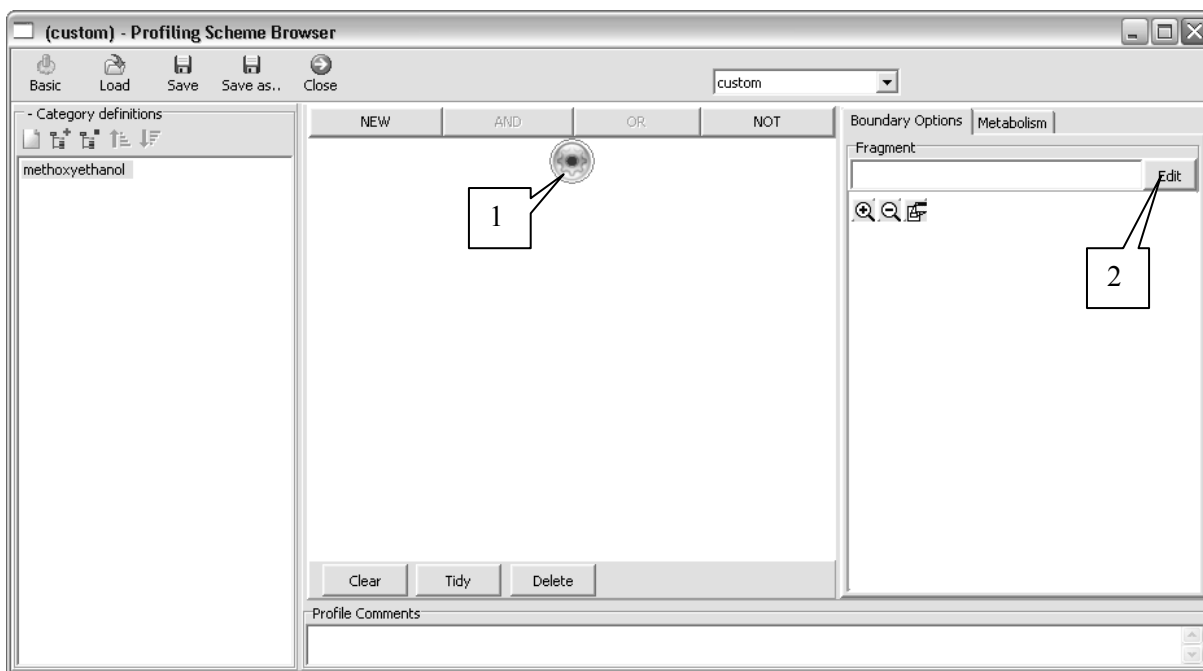
The Structure of 2-methoxyethanol can be defined by its SMILES notation i.e. C{H2}(O{H})C{H2}OC{H3} (see screenshot below). Note that all the hydrogens in the structure have to be defined as explicit hydrogens {H}, so that specifically 2-methoxyethanol is identified and not just any chemical containing the substructure C-O-C-C-O.
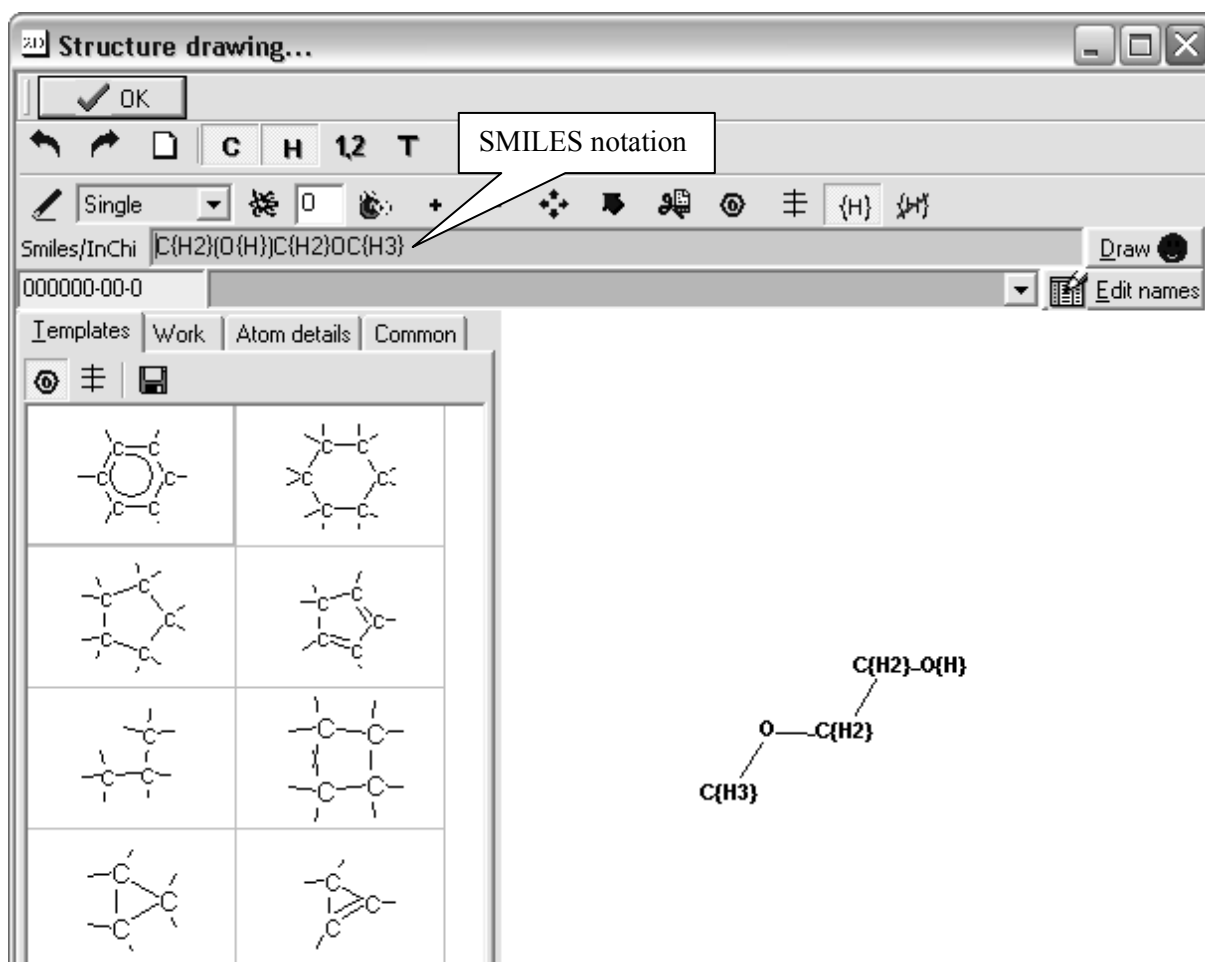
After pressing the **OK** button in the screenshot above, the structural boundaries just defined are entered into the right-hand window of the **Profiling Scheme Browser (1)** (see screenshot below). Having defined the structure of the metabolite, the user has to define the metabolic pathway in which he/she wants to identify the metabolite. For this he has to open the tab **Metabolism (2)** on the right hand window.
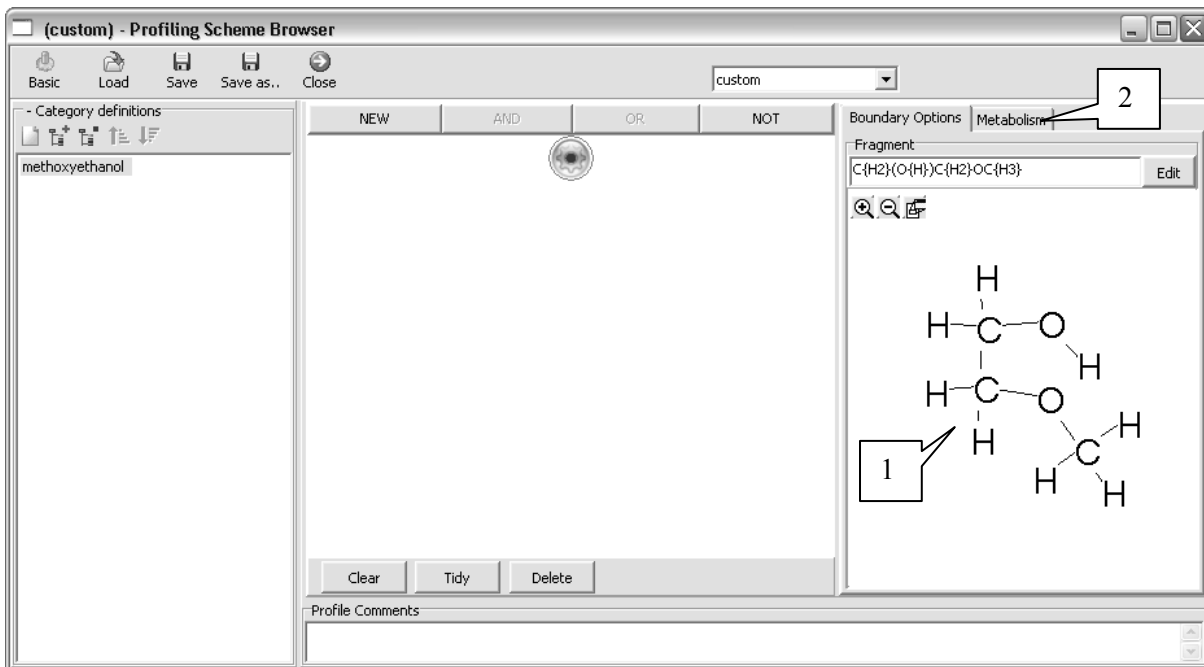
In the window for the tab **Metabolism**, the user has then to choose the simulator that he/she wants to apply to the list of chemicals to screen. In this example, the **Liver metabolism simulator (1)** is chosen (see screenshot below). Furthermore, the user should select the radio button **Metabolites (2)**, to indicate that the profiler is to identify 2-methoxyethanol solely among the metabolites of the compounds and not also among the "parents" or the starting list of compounds.
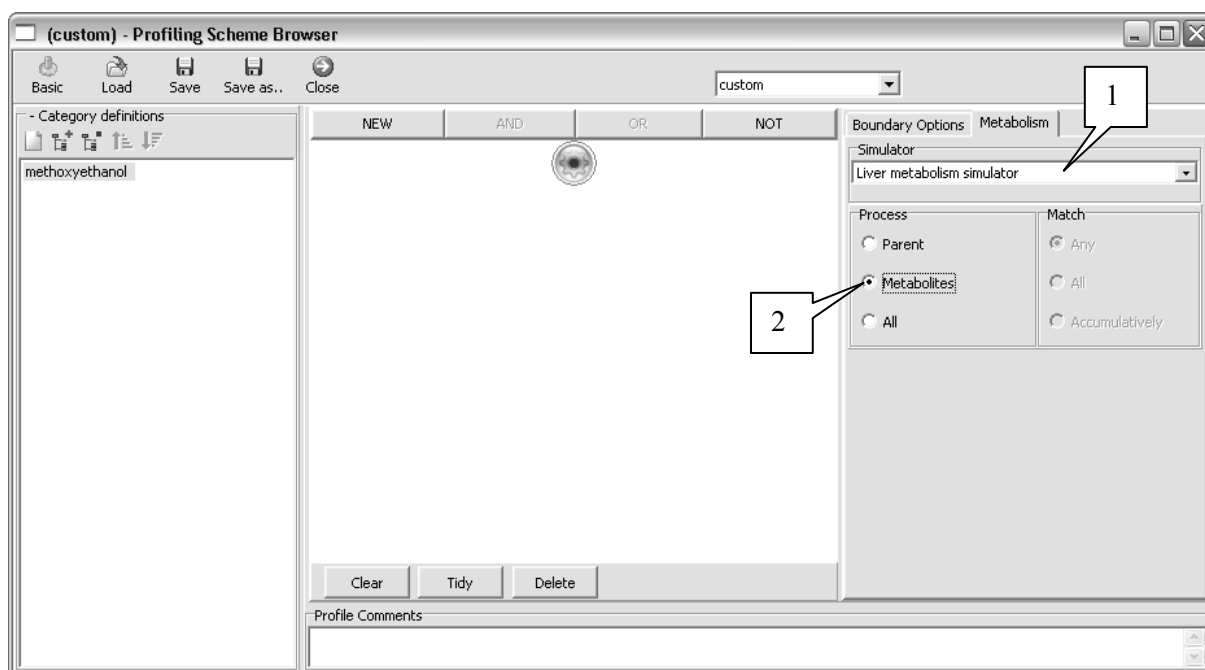
The user has now built a profiler which, for any substance can determine whether 2-methoxyethanol figures among its potential metabolites via liver metabolism.

As the toolbox also contains a database with actually observed liver metabolism pathways, it would be relevant to also identify those chemicals for which 2-methoxyethanol was actually identified as a liver metabolite. To do so, the user can add a profile boundary, by pressing the **NEW** button **(1)** and repeating the same steps outlined above, but selecting the **Observed Liver metabolism** (2) among the Simulator options (see screenshot below).
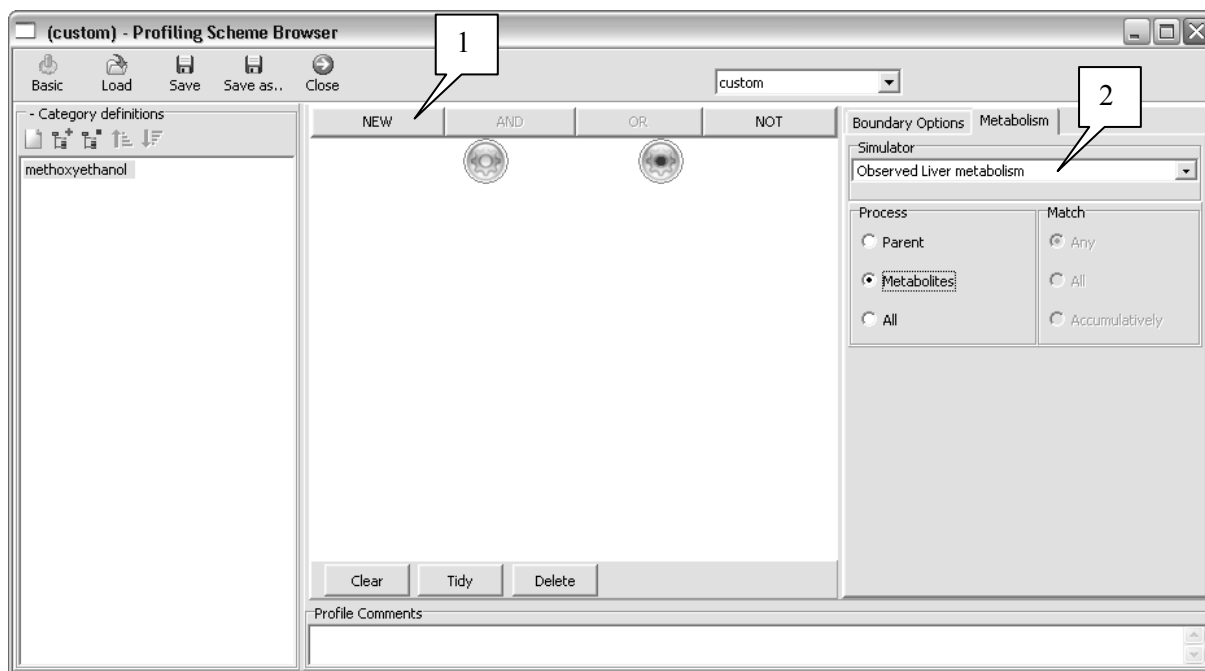
As a final step, the two defined profile boundaries need to be logically connected, so that the new profiler can identify chemicals which metabolize to 2-methoxyethanol either based on the liver metabolism simulator OR based on the liver metabolism pathway database.

To do so the user has to select the two pictograms , representing each a set of boundaries, by clicking on them while holding down the Shift key and then linking them by pressing the **OR** button (see screenshot below). The user can the **save (1)** the new profiler and **close (2)** the window.

The two sets of boundaries are now logically connected
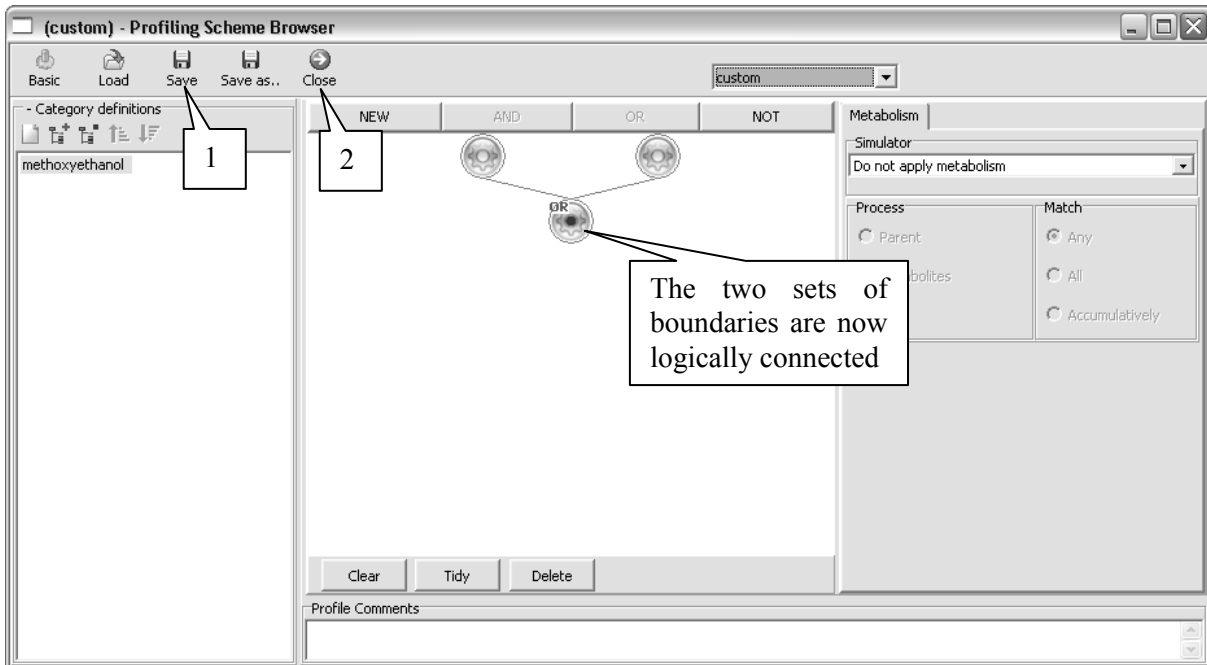
5.4. Screening the list of target chemicals

Once the new profiler has been defined it appears among the list of **Custom** profiling methods **(1)** and can simply be applied to the list of target chemicals by pressing the **Apply** button **(2)** (see screenshot below). The profiling will take a few minutes, depending on the speed of the computer. Once the Toolbox has profiled all target chemicals, the result can be visualized by right-clicking on **metabolic pathways** in the Profile and selecting **Profile statistics (3)** (see screenshot below).

The profile statistics shows that only four chemicals among the list of target chemicals are estimated to be able to metabolised to 2-methoxyethanol (see screenshot below). The four structures can be saved to a **sml file (1)** for further assessment with the Toolbox.
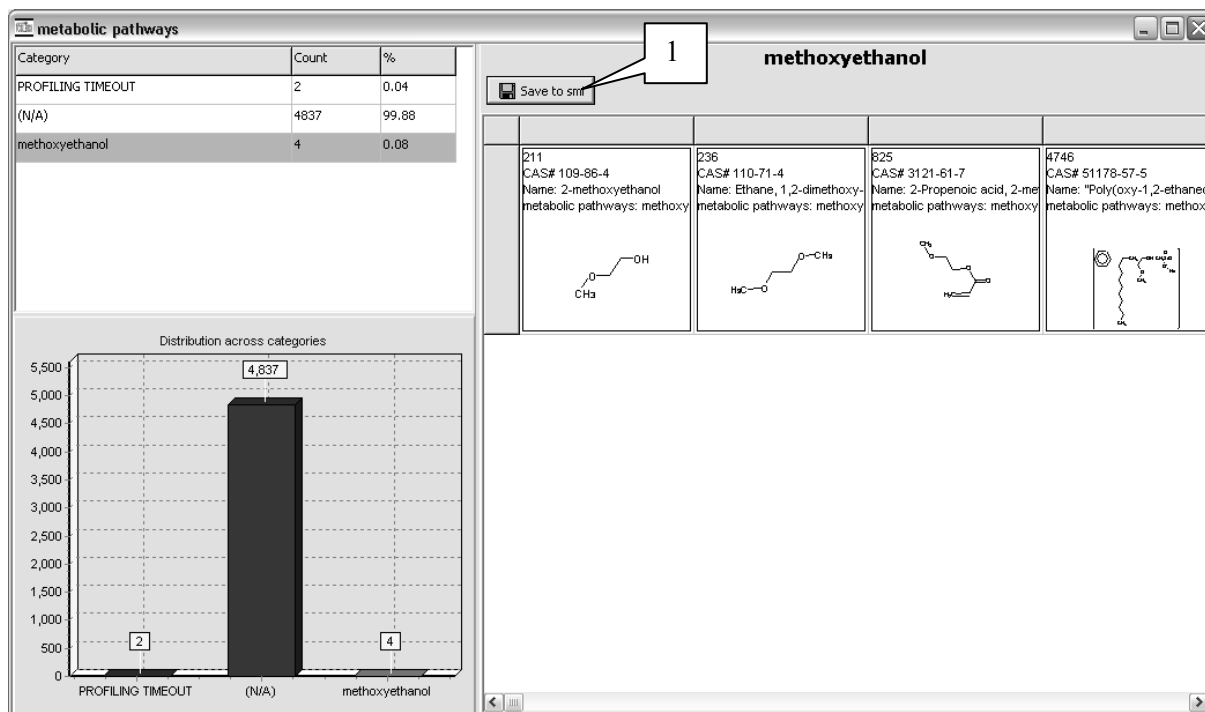
To conclude, the Toolbox can help identify chemicals that can potentially be grouped into a metabolic pathways category.

GLOSSARY

This Glossary provides an explanation for terms employed in guiding a user on how the Toolbox can be used.

Analogue Approach

The technique where endpoint information for one chemical is used to predict (read-across) the same endpoint for another chemical, which is considered to be "similar" in some way. This approach is used when the category is based on a very limited number of analogues are available and trends in properties are not apparent (see Figure 1 page 20).

Chemical Category

A group of chemicals whose physicochemical and human health and/or environmental toxicological properties and/or environmental fate properties are likely to be similar or follow a regular pattern as a result of structural similarity. A category of chemicals will often show the presence, absence or modulation of a particular effect for all members of the category, based on the presumption of a common mechanism of action.

Category Definition

The module of the workflow of the Toolbox, which allows the user to group chemicals into toxicologically meaningful categories, which include the target molecule.

Category Development

The methodology by which the user can use the Toolbox to construct a novel category for data filling (see Figure 2 page 63).

Category Track

The workflow designed to fill a data gap by read-across and/or trend analysis.

Chemical Input

The module in the workflow of the Toolbox which allows the user to define the target molecule or substance of interest by the use of various chemical identification properties (e.g., name, CAS number etc).

Data Matrix

A set of properties and/or effects data for chemical category members.

Endpoints

The module in the workflow of the Toolbox, which allows the user to interrogate the databases, which reside in the Toolbox and retrieve fate and toxicity data on the target molecule.

Filling Data Gaps

The module of the workflow of the Toolbox, which allows the user to make an endpoint specific prediction of the target molecule by read-across, trend-analysis or (Q)SAR models.

Flexible Track

The workflow designed to allow the open use of all functionalities in filling a data gap.

Grouping Chemicals

The general approach to assessing more than one chemical at the same time. It includes formation of a chemical category and identification of chemical analogues for read-across or trend analysis.

Mechanism

Synonym with mechanism of action and chemical mechanism; the chemical reaction between the target chemical or its metabolite and biological molecules, which leads to the toxicological response or endpoint.

Module

One of the six elements, which make up the workflow of the Toolbox.

Profiling

The module in the workflow of the Toolbox, which allows the user to retrieve relevant information on the target molecule other than fate and toxicity data (e.g., mechanism of chemical reactivity, organic functional groups within the chemical).

Pruning

The method by which the Toolbox reduces a data matrix for chemical category to a data matrix of a defined subcategory.

(Q)SAR

Quantitative Structure-Activity Relationship is a quantitative (mathematical) relationship between a numerical measure of chemical structure, and/or a physicochemical property, and an effect/activity. QSARs often take the form of regression equations, and can make predictions of effects/activities that are either on a continuous scale or on a categorical scale. Thus, in the term "QSAR", the qualifier "quantitative" refers to the nature of the relationship, not the nature of the endpoint being predicted.

(Q)SAR Application Toolbox

A sequence of six modules designed to build a chemical category and fill data gaps for fate and toxicity of industrial organic compounds.

(Q)SAR Track

The workflow designed to fill a data gap by using a library of (Q)SAR models.

Read-Across

The technique for filling data gaps, where endpoint information for an untested chemical is predicted by using data for the same endpoint from a tested chemical, which is considered to be "similar" in some way (e.g., activity, property or structure).

Report

The module of the work flow of the Toolbox, which provides the user with a downloadable audit trail describing each step in the process by which the Toolbox arrive at the prediction.

Subcategorization

A means within the Toolbox, which allows the user to refine the category definition.

Target Chemical

The compound being evaluated in the workflow of the Toolbox.

Track

Synonym with workflow.

Trend Analysis

The techniques used to predict endpoint information for an untested chemical by using a trend (increasing, decreasing or constant) in the experimental data for the same endpoint from several other chemicals, which have been tested and are considered to be "similar" in some way (e.g., activity, property or structure).

Workflow

The workflow or track is the sequence of modules, which guide the user through the Toolbox functions.

REFERENCES

Benigni, R, Bossa, C, Jeliazkova, N, Netzeva, T, Worth, A. 2008. The Benigni / Bossa rulebase for mutagenicity and carcinogenicity – a module of Toxtree. European Commission EUR 23241 EN - 2008.

Cronin, M.T.D. and Worth A.P. 2008. (Q)SARs for predicting effects relating to reproductive toxicity. *QSAR & Combinational Science* 27: 91-100.

Dimitrov, SD, Low, LK, Patlewicz, GY, Kern, PS, Dimitrova, GD, Comber, MH, Phillips, RD, Niemela, J, Bailey, PT, Mekenyan, OG 2005. Skin sensitization: modeling based on skin metabolism simulation and formation of protein conjugates. *Int. J. Toxicol.* 24, 189-204.

European Commission. 2008. Toxtree version 1.51. http://ecb.jrc.ec.europa.eu/qsar/qsar-tools/index.php?c=TOXTREE

Gerner, I, Liebsch, M, Spielmann, H. 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. *ATLA* 33, 215-237.

Mekenyan, O, Dimitrov, S, Serafimova, R, Thompson, E, Kotov, S, Dimitrova, N, and Walker, J. 2004. Identification of the structural requirements for mutagenicity by incorporating molecular flexibility and metabolic activation of chemicals I: TA100 model. *Chem. Res. Toxicol.* 17, 753-766.

Organisation for Economic Cooperation and Development. 2007a. Guidance Document on the Validation of (Quantitative) Structure –Activity Relationships. OECD Environment Health and Safety Publications. Series on Testing and Assessment. No 69. ENV/JM/MONO(2007)2. [http://appli1.oecd.org/olis/2007doc.nsf/linkto/env-jm-mono(2007)2].

Organisation for Economic Cooperation and Development. 2007b. Guidance on Grouping of Chemicals. OECD Environment, Health and Safety Publications. Series on Testing and Assessment. No 80. ENV/JM/MONO(2007)28: [http://appli1.oecd.org/olis/2007doc.nsf/linkto/env-jm-mono(2007)28].

Organisation for Economic Cooperation and Development. 2008. Manual for Investigation of HPV Chemicals. [http://www.oecd.org/env/existinjgchemcials].

Russom, CL, Bradbury, SP, Broderius, SJ, Drummond, RA, and Hammermeister, DE 1997. Predicting modes of toxic action from chemical structure: acute toxicity in Fathead minnow (*Pimephales Promelas*). Part I - development of a rule-based expert system. *Environ. Toxicol. Chem.* 16, 948-967

Serafimova, R, Todorov, M, Pavlov, T, Kotov, S, Jacob, E, Aptula, N, Mekenyan, O. 2007a. Identification of the structural requirements for mutagencitiy, by incorporating molecular flexibility and metabolic activation of chemicals II. General Ames mutagenicity model. *Chem. Res. Toxicol.* 20 (4), 662 - 676

R. Serafimova, M. Todorov, D. Nedelcheva, T. Pavlov, Y. Akahori, M. Nakai and O. Mekenyan, 2007b. QSAR and mechanistic interpretation of estrogen receptor binding, *SAR and QSAR in Environ. Research*, 18:1-33.

Schmieder, P.K., Ankley, G., Mekenyan, O., Walker, J.D., and Bradbury S. 2003. *Environmental Toxicology and Chemistry* 22: 1844-1854.

Schultz, T.W., Carlson. R.E., Cronin, M.T.D., Hermens, J.L.M., Johnson, R., O'Brien, P.J., Roberts, D.W., Siraki, A., Wallace, K.D. and Veith, G.D. 2006. A conceptual framework for predicting toxicity of reactive chemicals: Models for soft electrophilicity. *SAR QSAR in Environmental Research* 17: 413-428.

Tong, W., Fang, W.D., Hong, H., Xie, Q., Perkins, R. and Sheehan, D.M. 2004. Receptor-mediated toxicity: QSARs for estrogen receptor binding and priority setting of potential estrogenic endocrine disruptors. In: Cronin, M.T.D. and Livingstone D.J. (Eds) Predicting Chemical Toxicity and Fate. CRC Press Boca Raton FL pp.285-314.

Tunkel, J., Howard, P.H., Boethling, R.S., Stiteler, W. and Loonen, H. 2000. Predicting Ready Biodegradability in the Japanese Ministry of International Trade and Industry Test. *Environ. Toxicol. Chem.* 19: 2478-2485.

Verhaar, HJM, van Leeuwen, CJ, and Hermens, JLM 1992. Classifying environmental pollutants. 1. Structure-activity relationships for prediction of aquatic toxicity. *Chemosphere* 25, 471-491.

Walker, JD, Gerner, I, Hulzebos, E , Schlegel, K. 2005. The Skin Irritation Corrosion Rules Estimation Tool (SICRET). *QSAR Comb. Sci.* 24, 378-384