

OECD QSAR Toolbox v.3.4

Predicting developmental and reproductive toxicity of Diuron (CAS 330-54-1) based on DART categorization tool and DART SAR model

Outlook

- **Background**
- Objectives
- The exercise
- Workflow

Background

- This is a step-by-step presentation designed to take the user through the workflow for filling data gap for reproductive and developmental toxicity by read-across based on an analogue approach.

Outlook

- Background
- **Objectives**
- The exercise
- Workflow

Objectives

This presentation demonstrates a number of functionalities of the Toolbox:

- Identify analogues of target chemical by applying DART scheme via two different ways:
 - For primary categorization
 - For subcategorization
- Retrieve experimental results available for those analogues.
- Fill data gaps by read across.
- Support read-across prediction by DART model.

Outlook

- Background
- Objectives
- **The exercise**
- Workflow

The exercise

- In this exercise we will predict the developmental and reproductive (DART) toxicity of **3-(3,4-dichlorophenyl)-1,1-dimethylurea CAS 330-54-1 (Diuron)**.
- Two scenarios for defining the initial category of similar analogues will be applied:
 - Initial category identified by endpoint specific DART scheme
 - Initial category identified by empiric Organic functional group (OFG) with followed by subcategorization by DART scheme
- Gather available experimental data for the target chemical and identified analogues.
- Apply read across prediction based on analogue approach.
- Apply external DART model.

Outlook

- Background
- Objectives
- The exercise
- **Workflow**

Workflow

- **The Toolbox has six modules which are used in a sequential workflow:**
 - Chemical Input
 - Profiling
 - Endpoints
 - Category Definition
 - Filling Data Gaps
 - Report

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - **Input**

Chemical Input Overview

- This module provides the user with several means of entering the chemical of interest or the target chemical.
- Since all subsequent functions are based on chemical structure, the goal here is to make sure the molecular structure assigned to the target chemical is the correct one.

Chemical Input

Ways of Entering a Chemical

User Alternatives for Chemical ID:

A. Single target chemical

- Chemical Name
- Chemical Abstract Services (CAS) number (#)
- SMILES (simplified molecular information line entry system) notation/InChi
- Drawing chemical structure
- Select from User List/Inventory/Databases
- Chemical IDs such as EC number, EINECS number
- Query Tool

B. Group of chemicals

- User List/Inventory
- Specialized Databases

Getting Started

- Open the Toolbox.
- The six modules in the workflow are seen listed next to "QSAR TOOLBOX".
- **Click** on "Input" (see next screen shot).

Chemical Input Screen

Input target chemical by CAS#

0 Document

1/0/0

1. Click on "CAS#"

Chemical Input Screen

Enter CAS# 330-54-1

Search by CAS #

330-54-1 Tautomeric sets

Select All Clear All Selection Selected 1 of 1

Selected	CAS	Smiles	Depiction	Names	CAS/Name	2D/Name	CAS/2D
1. Yes	330-54-1	CN(C)C(=O)Nc1ccc(Cl)c(Cl)c1			1:: High 1:: Av 2:: Av 3:: Bz 4:: Bi 5:: D 6:: E 7:: E 8:: H 9:: M 10:: F 11:: f 12:: t	1:: High 1:: U 2:: Av 3:: T 4:: E 5:: k 6:: Av 7:: R 8:: Bi 9:: M 10:: F 11:: f 12:: f	

1. **Enter** the CAS# in the blank field; 2. **Click** "Search" button; 3. **Press** "OK"

Chemical Input

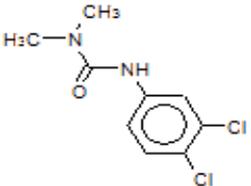
Target chemical identity

The Toolbox now searches the databases to find out if the CAS# you entered is linked to a molecular structure stored in the Toolbox. It is displayed as a 2-dimensional depiction.

Search by CAS #

330-54-1 Tautomeric sets

Select All Clear All Invert Selection Selected 1 of 1

Selected	CAS	Smiles	Depiction	Names	CAS/Name	2D/Name	CAS/2D
1. Yes	330-54-1	CN(C)C(=O)Nc1ccc(Cl)c(Cl)c1			1:: High	1:: High	: High
					1:: A	1:: U	A
				1:	2:: A	2:: A	
				2:	3:: B	3:: T	
				3:	4:: B	4:: E	
				4:	5:: D	5:: k	
				5:	6:: E	6:: A	
				6:	7:: E	7:: R	
				7:	8:: H	8:: B	
				8:	9:: M	9:: M	A
				9:	10:: F	10:: F	
				10:	11:: F	11:: F	
				11:	12:: T	12:: E	
				12:	13:: F	13:: F	

Chemical Input

Target chemical identity

- **Double click** “Substance Identity” displays the chemical identification information.
- The user should note that existing names of the target chemical are presented in different colours. This indicates the reliability of relation CAS-Name-SMILES for the target chemical (see next screen shots).
- The workflow on the first module is now complete, and the user can proceed to the next module.

Chemical Input

Target chemical identity

The CAS number is colored green* because this chemical with this CAS number and 2D structure belongs to high quality inventories. Double click over the cell with CAS number to see the sources of chemical ID

Explain QA Form

CAS/2D	Names	CAS/Name	2D/Name	CAS/2D	Status
CN(C)C(=O)Nc1ccc(Cl)c(Cl)c1		1:: High Quality	1:: High Quality	1:: High Quality	
CAS: 330541		1: Aquatic ECETOC	1: USER DEFINED	1: Aquatic ECETOC	
		2: Aquatic OASIS	2: Aquatic ECETOC	2: Aquatic OASIS	
		3: Bacterial mutagenic	3: ToxRefDB US-EPA	3: Bacterial mutag	
		4: Bioaccumulation fis	4: ECHA PR	4: Bioaccumulation	
		5: Developmental & R	5: km database Enviro	5: Bioaccumulation	
		6: ECHA CHEM	6: Aquatic OASIS	6: Bioconcentration	
		7: ECHA PR	7: REACH ECB	7: Biodegradation	
		8: Hydrolysis rate con	8: Bioaccumulation fis	8: Biodegradation	
		9: MUNRO non-cancer	9: MUNRO non-cancer	9: COSING	
		10: Phys-chem EPISUJ	10: Phys-chem EPISUJ	10: Canada DSL	
		11: REACH ECB	11: Hydrolysis rate coi	11: Carcinogenic P	
		12: ToxRefDB US-EPA	12: ECHA CHEM	12: Carcinogenicity	
		13: USER DEFINED	13: Bacterial mutagen	13: DSSTOX	
		14: km database Envir	14: Developmental & I	14: Developmental	
		2:: High Quality	2:: High Quality	15: ECHA CHEM	
		1: Bioconcentration NI	1: TSCA	16: ECHA PR	
		2: Biodegradation NIT	2: NICNAS	17: ECOTOX	
		3: ECHA CHEM	3: Bioconcentration NI	18: EINECS	
		4: HPV OCED	4: US HPV Challenge f	19: Genotoxicity O	

*More details about color legend are provided on next slide

Chemical Input

Chemical identity

The colour code indicates the reliability of the chemical identifier:

- **Green:** There is a high reliability between the identifier and the structure. This colour is applied if the identifier is the same in several quality assured databases.
- **Yellow:** There is only a moderate reliability between the identifier and the structure. The colour is applied if the identifier is the same in several databases for which the quality assurance could not be established.
- **Red:** There is a poor reliability between the identifier and the structure. The colour is applied if the identifier is allocated to different structures in different databases.

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Input
 - **Profiling**

Profiling Overview

- “Profiling” refers to the electronic process of retrieving relevant information on the target compound, other than environmental fate, ecotoxicity and toxicity data, which are stored in the Toolbox database.
- Available information includes likely mechanism(s) of action, as well as observed or simulated metabolites.

Profiling

Side-Bar to Profiling

- For most of the profilers, background information can be retrieved by highlighting one of the profilers and clicking on “View”.
- Detailed information for DART v. 1.0 scheme (Developmental and Reproductive toxicity) is provided on next slide.

Developmental and Reproductive Toxicity (DART)

Background

- DART scheme is an adaptation of a framework for identifying chemicals with structural features associated with the potential to act as reproductive or developmental toxicants outlined in the journal of Wu S et al [1].
- It is implemented as a pilot endpoint specific scheme, developed on the basis of the combination of known modes of action (MOA) and associated structural features.
- DART scheme include 25 categories and 125 sub-categories organized as a decision scheme. Definition of the categories are based on a detailed review of 716 chemicals that have been evaluated for their DART potential effect. Mechanistic interpretation and reliability is provided for each category.
- DART scheme is implemented as a profiling/categorization tool and as a SAR model.
- It can be used both as a component of a screening system to identify chemicals of potential concern, and as part of weight of evidence decisions based on structure-activity relationships (SAR), to fill data gaps without generating additional test data.
- Output of the scheme indicate that the chemical of interest is associated with chemical structures:
 - known to have DART - *Known precedent reproductive and developmental toxic potential*
 - not known to have DART - *Not known precedent reproductive and developmental toxic potential*
 - which have structural features outside the domain of the DART decision tree - *Not cover by the decision tree.*

1. Wu S, Fisher J, Naciff J, Laufersweiler M, Lester C, Daston G, Blackburn K. Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants. Chem Res Toxicol. 2013 Dec 16;26(12):1840-61.

Profiling

Overview of DART scheme

1. Highlight the profiler

2. Click "View"

Nodes of the tree

QSAR TOOLBOX

Profiling

Profiling methods

Select All Unselect All Invert About

- Protein binding by OECD
- Protein binding potency
- Superfragments
- Toxic hazard classification by Cramer (e)
- Toxic hazard classification by Cramer (c)
- Ultimate biodeg
- Endpoint Specific**
 - Acute aquatic toxicity classification by V
 - Acute aquatic toxicity MOA by OASIS
 - Aquatic toxicity classification by ECOSA
 - Bioaccumulation - metabolism alerts
 - Bioaccumulation - metabolism half-lives
 - Biodegradation fragments (BioWIN MIT)
 - Carcinogenicity (genotox and nongenotox)
 - DART scheme v.1.0
 - DNA alerts for Ames, MN and CA by OASIS
 - Corrosion Exclusion rules based on corrosion
 - Corrosion Inclusion rules based on corrosion
 - Mutagenicity (Ames test) alerts
 - Mutagenicity (Micronucleus) alerts
 - Cell gene expression
 - Oncologic Primary Classification
 - Protein binding alerts for Chromosomal
 - Protein binding alerts for skin sensitization
- Metabolism/Transformations**
 - Documented
 - Observed Mammalian metabolism
 - Observed Microbial metabolism

Filter endpoint tree...

Structure

- Substance Identity
- Physical Chemical Properties
- Environmental Fate and Transport
- Ecotoxicological Information
- Human Health Hazards

DART scheme v.1.0 (Endpoint Specific) - Profiling Scheme Browser

Save Options Close

Profiler

Inorganic and organometallic

Receptor mediated

Endocrine toxicity

Non-cyclic hydrocarbons

Non-aromatic (hetero)cyclic compounds

Specific (hetero)aromatics

Properties Reference Target/Metabolites Metabolism Packet Tr

Query Tree

Caption

Is the chemical ER/AR binder?

Group

Node library path (separated with _)

YES Category Name

Profiling

Overview of DART scheme

Mechanistic interpretation

Category 3: Retinoic acid receptor (RAR), aryl hydrocarbon receptor (AhR) binders and prostaglandin receptor agonists

3b. AhR binders

3b-4. Indole-related chemicals

Based on the effects of indole-3-carbinol (CAS#700-06-1) which causes reproductive effects in both male and female offspring linked to activation of AhR¹, the indole-related chemicals (3b-4) in Figure S11 are hypothesized to have DART potential despite the lack of direct test data. The substituents such as R and R₁ can't be clearly defined based on the available data. However, the other as yet untested chemicals in this group and di-, poly-indole that have the potential to bind AhR may also share similar DART effects.



R=OH, CH₂OH,
methylpyrrolidinyl methyl
R₁=H, hydrogen

3b-4

Fig. S11. The general of structural features indole-related derivatives

Original reference:

1. **Select** a node 2. **Click** "View scheme" 3. Mechanistic (textual) description associated with category 4. **Click** "Advanced" to see structural boundaries coding the rule

Profiling

Overview of DART scheme

1. Structural boundary

2. Definition of structural boundary

3. Members of the common fragment

Profiling Overview of DART scheme

1

2

3

4

Local training set chemicals associated with category

Developmental and reproductive effects for the selected training set chemical

1. Click on "Training set panel"

2. Select one of the chemicals

Profiling

Application of endpoint specific profiling schemes

- In this example profiling by DART and OFG is applied in order to analyse the potential to cause DART toxicity and to identify the general structural fragments available in the molecule, which further could be used for categorization.
- Follow the steps:
 - **Select** DART v1.0
 - **Select** two Organic functional group profilers
 - **Click** Apply

Profiling

Apply related profiling schemes

The screenshot shows the QSAR Toolbox Profiling interface. The top menu bar includes 'Profiling', 'Input', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Profiling' menu is open, showing 'Apply', 'New', 'View', and 'Delete' options. The main window is divided into two panes. The left pane, titled 'Profiling methods', contains a list of methods with checkboxes. Under the 'Empiric' section, 'Organic functional groups (US EPA)' is selected. The right pane, titled 'Filter endpoint tree...', shows a tree structure with 'Structure' selected. A chemical structure is displayed in the top right of the right pane. Three callout boxes with numbers 1, 2, and 3 point to the 'DART scheme v.1.0' checkbox, the 'Organic functional groups (US EPA)' checkbox, and the 'Apply' button, respectively.

1. **Select** the "DART v1.0 profiling scheme"
2. **Select** two "Organic functional groups" profilers
3. **Click** "Apply"

Profiling

Outcome of profiling results

Outcome of DART profiling results appears in the box under the chemical structure organized in several categories marked in red color distributed in a few rows. The first row shows general category "Known...." and indicate if the chemical has the potential to cause reproductive and developmental toxic effect. The rest three entries indicates the specific DART category (e.g Polyhalogenated benzene derivatives 8c) of the chemical. Some of the DART category are hierarchically organized and presented in more than one entry. This is the case with sub-category "N-aryl substituted urea..." part of more general "Non-steroid nucleus..." category. More details about DART category is given on next slide (follow the steps given the box below)

1

2

3

4

1. **Select** the cell with profiling result
2. **Right Click** and **select** "Explain"
3. **Select** category
4. **Click** "Details"

Profiling

Explain of profiling results

The screenshot displays the QSAR Toolbox software interface. On the left, a large decision tree is visible, categorized into sections like 'Inorganic and organometallic', 'Receptor mediated', 'Endocrine toxicity', 'Non-cyclic hydrocarbons', 'Non-aromatic (hetero)cyclic compounds', and 'Specific (hetero)aromatics'. A red callout '1' points to a 'Referential Node' in the tree. Below the tree, a 'View Scheme' button is highlighted with a red callout '2'. On the right, a detailed window titled 'Category 8: Aromatic compounds with alkyl, multi-halogen and nitro groups' is open. This window shows a 'Polyhalogenated benzene derivatives (8c)' rule. A red callout '3' points to a structural diagram of a benzene ring with multiple halogen substituents. A red callout '4' points to the 'Fragment' section, which shows a chemical structure with a blue box around a carbon atom and a red box around a chlorine atom, representing the rule's coding. A red callout '5' points to the 'Profile Description' section, which provides a textual explanation of the rule: 'Category 8: Aromatic compounds with alkyl, multi-halogen and nitro groups. Chemicals within this category include the following five sub-categories: 8a. toluene and small alkyl toluene derivatives; 8b. NO2-alkyl/NO2-benzene derivatives; 8c. polyhalogenated benzene derivatives; 8d. polyhalogenated-, NO2/halogenated-oxydibenzene; 8e. dihalogen-, dinitro-phenol and their ester derivatives. The

1. **Click** on "Referential Node" panel; 2. **Click** "View Scheme" 3. **Click** on first structural boundary 4. Structural fragment coding the rule 5. Textual description of the rule

Profiling

Interpretation of profiling results

- Profiling result shows that the target chemical is classified as:
 - “Known precedent for DART effect” based on classification into two DART sub-categories: “Polyhalogenated benzenes” and “N-aryl substituted ureas”. Will be further investigated.
 - “Aryl, Urea derivatives and Aryl halide” by OFG, which will be used further for identifying analogues.

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - **Endpoint**

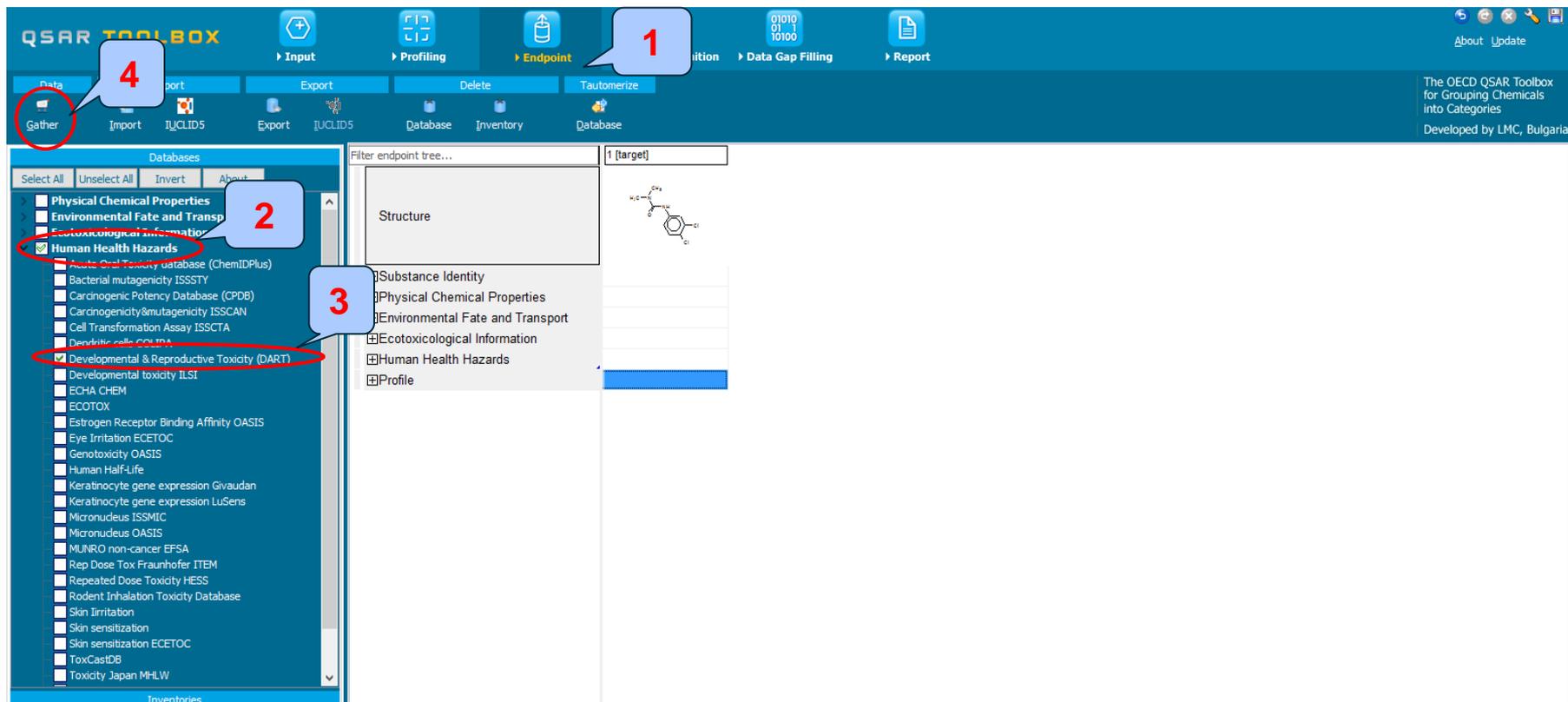
Endpoint Overview

- “Endpoint” refers to the electronic process of retrieving the environmental fate, ecotoxicity and toxicity data that are stored in the Toolbox.
- Data gathering can be executed in a global fashion (i.e., collecting all data for all endpoints) or on a more narrowly defined basis (e.g., collecting data for a single or limited number of endpoints).

Endpoint Case study

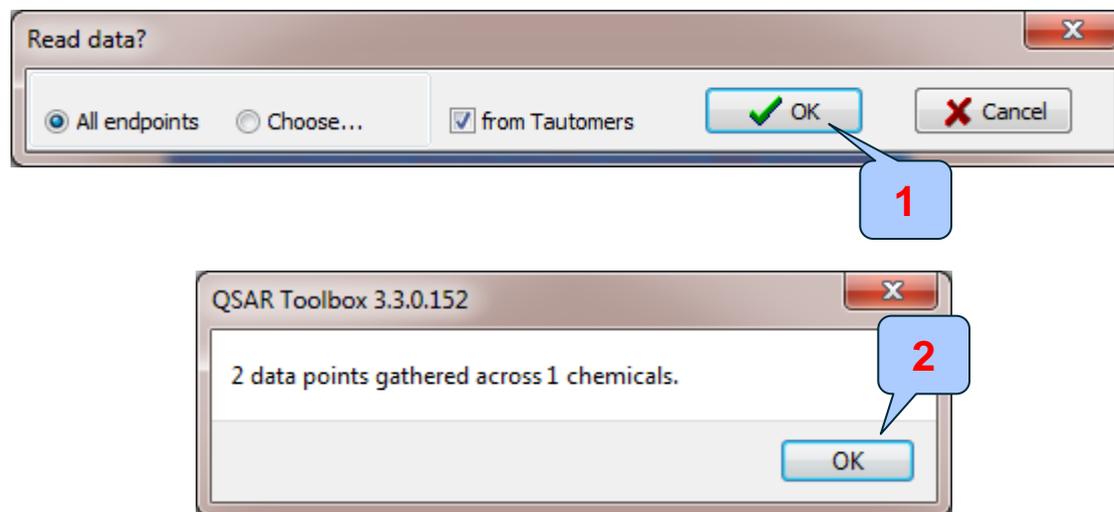
- In this example, we limit our data gathering to two toxicity endpoints: developmental and reproductive toxicity.
- DART database has been implemented into the Toolbox 3.3
- Developmental and Reproductive Toxicity (DART) database – 716 chemicals with 1430 data points separated as follows:
 - Developmental toxicity (716 data points)
 - Reproductive toxicity (714 data points)
- In this example, we collect data from the DART database containing experimental results for developmental and reproductive toxicity(DART).
 - **Click** on “Endpoint” in the Toolbox workflow.
 - **Expand the** “Human Health Hazards” section
 - **Click** on the box to select that database.
 - **Click** on “Gather data” (see next screen shot).

Endpoint Gather data



1. **Go** to "Endpoint"
2. **Expand** the "Human Health Hazards" section
3. **Select** database related to the target endpoint: "Developmental & Reproductive Toxicity (DART)"
4. **Click** "Gather"

Endpoint Gather data



1. **Click** "OK" to extract data from database
2. The message informs you that 2 data points are gathered for the target chemical. **Click** "OK"

Endpoint Gather data

The screenshot displays the QSAR Toolbox interface. The top navigation bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Endpoint' tab is active, showing a 'Filter endpoint tree...' search bar with the text '[target]'. On the left, the 'Databases' panel is expanded to 'Human Health Hazards', with 'Developmental & Reproductive Toxicity (DART)' selected. The 'Structure' panel shows a chemical structure. The 'Endpoint tree' on the right lists various categories, with 'Developmental Toxicity' and 'Reproductive Toxicity' highlighted. A red box highlights the data entries: '(1/1) M: Known developmental potential' and '(1/1) M: Not known reproductive potential'. A red arrow points from a text box to these entries.

Measured data for the target appeared on data matrix. There are positive and negative data for the target chemical. We will try to reproduce the measured data by read-across

Recap

- In the first module, you have entered the target chemical being sure of the correctness of the structure.
- In the second module, you have profiled the target chemical and found that the target could cause DART effect. It is categorized as known precedent for developmental and reproductive toxicity. This is due to the chemical pertaining to two chemical classes associated with DART toxicity.
- In the third module, you have found that there are two experimental data for the target structure: positive developmental and negative reproductive. We will try to reproduce them using read across analysis.
- Before proceeding with the “Data Gap Filling” module, the user should define a category with similar analogues. Two scenarios are played for identifying analogues:
 - DART scheme used as a categorization tool (used for phase I, see next slides)
 - DART scheme used in subcategorization procedure (used for phase II)
- **Click** on “Category Definition” to move to the next module.

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - Endpoint
- **Category definition**
 - **Overview**
 - Scenario 1
 - Scenario 2

Category Definition

Grouping methods

- The forthcoming 4 slides provide basic information about definition and procedure of “Category definition”.
- The different grouping methods allow the user to group chemicals into chemical categories according to different measures of “similarity” so that within a category data gaps can be filled by read-across.
- Detailed information about grouping chemical (Chapter 4) could be found in document “Manual for Getting started” published on OECD website:

<http://www.oecd.org/chemicalsafety/risk-assessment/theoecdqsartoolbox.htm>

Basic guidance for category formation and assessment

Suitable categorization phases:

1. Structure-related profilers.
2. Endpoint specific profilers (for sub-cat).
3. Additional structure-related profilers, if needed to eliminate dissimilar chemicals (to increase the consistency of category) (e.g. chemical elements).

Performing categorization:

1. Categorization phases should be applied successively.
2. The application order of the phases follows three general stages but variations within them are case specific.
3. More than one category can be used within one phase for forming one final category.
4. Some of the main phases could be skipped if consistency of category members is reached.

Graphical illustration of suitable categorization phases is shown on next slide.

Suitable Categorization/Assessment Phases

Phase I. Structure based

- US EPA Categorization
- OECD Categorization
- Organic functional group
- Structural similarity
- ECOSAR

**Broad grouping
Endpoint Non-specific**

Repeating Phase I due to Multifunctionality of chemicals

Phase II. Mechanism based

- DNA binding mechanism
- Protein binding mechanism
- Genotoxicity/carcinogenicity
- DART v1.0
- Cramer rules
- Verhaar rule
- Skin/eye irritation corrosion rules
- Repeated dose profiler (NITE)

**Subcategorization
Endpoint Specific**

Metabolism accounted for

Phase III. Eliminating dissimilar chemicals

**Apply Phase I – for structural dissimilarity
Filter by test conditions – for Biological dissimilarity**

**Subcategorization
Endpoint Specific**

Category Definition

Grouping methods – phase I

Suitable Categorization/Assessment Phases

Phase I. Structure based

- US EPA Categorization
- OECD Categorization
- Organic functional group
- Structural similarity
- ECOSAR

Broad grouping
Endpoint Non-specific

Phase I categorization in Toolbox

1 [target]

Structure

Substance Identity
 Physical Chemical Properties
 Environmental Fate and Transport
 Ecotoxicological Information
 Human Health Hazards
 Profile

Predefined

- OECD HPV Chemical Categories
- US-EPA New Chemical Categories

Endpoint Specific

- Aquatic toxicity classification by ECOSAR

Empiric

- Organic Functional groups
- Organic Functional groups (nested)

Not categorized
Neutral Organics

Substituted Ureas

Aryl
Aryl halide
Urea derivatives
Aryl halide
Overlapping groups
Urea derivatives

127 analogues are identified. It is not recommended to use "Neutral organic" as phase I*

9 analogues are identified**

8 analogues are identified

8 analogues are identified

3 analogue are identified

Structural similarity, Dice ACF, 50%

*Neutral organic category include chemicals having different functionalities as alcohols, ketones, ethers etc. In this respect the basic principle illustrated on slide 41 that structurally similar chemicals may elicit similar effects would not be preserved, because Neutral organic mixed many different functionalities.

**OFG is used for primary categorization, because the two basic functionalities available within the molecule: "Ureas" and "Aryl halides" will be preserved in the group of identified analogues, while the ECOSAR omits the "Aryl halide" functionality and identifies "Substituted ureas" only.

Category Definition

Grouping methods – phase II

Suitable Categorization/Assessment Phases

Phase II. Mechanism based

- DNA binding mechanism
- Protein binding mechanism
- DART v1.0
- Repeated dose profiler (NITE)

Subcategorization
Endpoint specific

Phase II categorization in Toolbox

The screenshot shows the 'Filter endpoint tree...' window in the QSAR Toolbox. On the left, a tree view shows the 'Endpoint Specific' category expanded, with sub-items like 'DART scheme v.1.0' and 'DNA alerts for AMES, MN and CA by OASIS v.1.3'. On the right, a search bar contains '1 [target]' and a chemical structure is displayed. Below the structure, a table lists the results of the search, including 'Known precedent reproductive and de...', 'Non-steroid nucleus derived estrogen ...', and 'Polyhalogenated benzene derivatives ...'. A red arrow points from a text box to the '5 analogues are identified' result.

5 analogues are identified

In this case it is not reasonable to use DNA or Protein binding profiler for categorization.

Category Definition

Grouping methods

- Based on these classifications and basic guidance for grouping chemicals explained on the previous slides, two scenarios for identifying similar analogues have been applied in further read-across analysis:
 - Scenario 1: DART v1.0 scheme used as a categorization tool (applied for primary categorization - phase I).
 - Scenario 2: OFG is used for primary categorization with forthcoming subcategorization by DART scheme (applied for subcategorization - phase II).
- Identifying analogues based on two scenarios mentioned above will be applied in further read-across analysis.

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - Endpoint
- **Category definition**
 - Overview
 - **Scenario 1**
 - Scenario 2

Category Definition

Grouping methods

- Based on these classifications and basic guidance for grouping chemicals explained on the previous slides, two scenarios for identifying similar analogues have been applied in further read-across analysis:
 - **Scenario 1: DART v1.0 scheme used as a categorization tool (applied for primary categorization - phase I).**
 - Scenario 2: OFG is used for primary categorization with forthcoming subcategorization by DART scheme (applied for subcategorization - phase II).

Category Definition

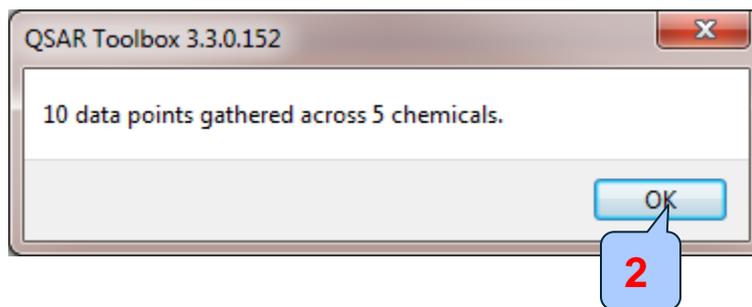
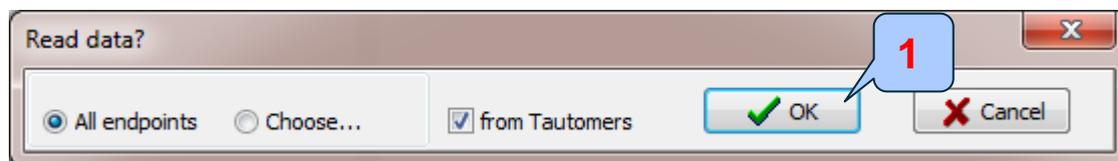
Scenario 1: DART scheme used as a categorization tool

The screenshot illustrates the 'Category Definition' process in the QSAR Toolbox. The 'Define' button (2) is used to select a grouping method from the 'Grouping methods' list (1). The 'DART scheme v.1.0' method is selected, and its target profiles are reviewed (3). The 'Define category name' dialog (4) shows the resulting category name and the number of chemicals identified (5).

1. **Highlight** the "DART scheme v1.0";
 2. **Click** "Define";
 3. **Click** "OK" to confirm the defined categories for the target chemical;
 4. Five analogues are identified. **Click** "OK"

Category Definition

Scenario 1: DART scheme used as a categorization tool



1. **Click** "OK" to extract data for the analogues from DART database
2. 10 data points are gathered for the identified 5 analogues. **Click** "OK"

Category Definition

Scenario 1: DART scheme used as a categorization tool

The experimental results for the analogues appeared on datamatrix

The screenshot displays the QSAR Toolbox interface with the 'Category Definition' workflow selected. The 'Filter endpoint tree...' panel is expanded to show 'Developmental and Reproductive Toxicity' (DART) endpoints. A red box highlights the 'Developmental and Reproductive Toxicity' row in the data matrix, with a callout bubble containing the number '1'. The data matrix shows values for five target compounds across various endpoints.

Endpoint	1 [target]	2	3	4	5
Developmental and Reproductive Toxicity	(5/5) M: Known develop...	M: Known develop...	M: Known develop...	M: Known develop...	M: Known develop...
Reproductive Toxicity	(5/5) M: Not known repr...	M: Known reprodu...	M: Known reprodu...	M: Known reprodu...	M: Known reprodu...

1. As mention on the previous slides we will try to reproduce the observed data.

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - Endpoint
- **Category definition**
 - Overview
 - **Scenario 1**
 - **Apply read-across**
 - Scenario 2

Read-across applied for developmental tox

Scenario 1: DART scheme used as a categorization tool

Investigated endpoint: Developmental toxicity

1. Click on "Data Gap filling"
2. Click on the cell corresponding to "Developmental Toxicity" endpoint
3. Select "Read-across"
4. Click "Apply"
5. A window indicating data inconsistency appears. DART data has been implemented into the system with original scale called "DART toxicity original". A less informative scale "DART toxicity" has been introduced. Also a scale conversion from original (more informative) to the less informative scale has been implemented. In our case we will use the less informative scale. Select "DART toxicity" scale.
6. Click "OK"

Read-across applied for developmental tox

Scenario 1: DART scheme used as a categorization tool

Investigated endpoint: Developmental toxicity

All 5 analogues are positive and structurally similar above 50%

All analogues are positive and similar with respect to DART endpoint. Let's check how similar are they with respect to "Structural similarity". Follow the steps:

1. **Open** "Select/filter data"; 2. **Click on** "Subcategorize" ; 3. **Select** "Structural similarity"; The following similarity options are used in the subcategorization. **Click** on "Adjust options" button (4) to see the options: Dice, Atom pairs and atom type as atom characteristics are selected only.

The analysis shows that all analogues are similar above 50 % with respect to the target. Continue the workflow with **accept** the prediction. Follow the steps: 5. **Click** "Accept prediction"; 6. **Click** "Return to matrix"

Read-across applied for reproductive tox

Scenario 1: DART scheme used as a categorization tool

Investigated endpoint: Reproductive toxicity

The screenshot shows the QSAR Toolbox software interface. The 'Data Gap Filling' step is active. A dialog box titled 'Possible data inconsistency' is open, showing a 'Scale/Unit' dialog with 'DART toxicity' selected. Numbered callouts (1-6) highlight key actions:

1. Selecting the cell corresponding to "Reproductive toxicity" endpoint in the endpoint tree.
2. Selecting the "DART toxicity" option in the dialog.
3. Selecting "Read-across" in the left sidebar.
4. Clicking "Apply" in the top left.
5. Clicking "DART toxicity" in the dialog.
6. Clicking "OK" in the dialog.

1. Positive prediction obtained for endpoint "Developmental toxicity" reproduces the positive observed data. Further read-across analysis continues with next endpoint "Reproductive toxicity". Follow the steps:
2. **Select** the cell corresponding to "Reproductive toxicity" endpoint;
3. **Select** "Read-across";
4. **Click** "Apply";
5. In our case less informative scale is used: "DART toxicity";
6. **Click** "OK"

Read-across applied for reproductive tox

Scenario 1: DART scheme used as a categorization tool

Investigated endpoint: Reproductive toxicity

The screenshot displays the 'Subcategorization' window in the OECD QSAR Toolbox. The 'Grouping methods' list on the left includes 'Empiric', 'Toxicological', and 'Experimental'. The 'Adjust options' panel shows 'Target' set to 'Similar 100%' and 'Differ from target by' set to 'All categories'. The 'Correlation' section shows 'Structural similarity' selected. The main window shows a 'Read across prediction' plot for 'Reproductive Toxicity'. The plot has a y-axis with 'Observed' (positive/negative) and 'Predicted' (positive/negative) and an x-axis with numerical values from 2.10 to 3.00. A target chemical (1) is at approximately x=2.65 with a negative observed data point. Four analogues (2-5) are at higher x-values (approx. 2.85-3.05) with positive predicted data points. A text box above the plot reads: 'Read across prediction of Reproductive toxicity, taking the highest mode from the nearest 5 neighbours, based on 4 values from 4 neighbour chemicals, Observed target value: 'Negative', Predicted target value: 'Positive''. The right-hand 'Accept prediction' panel has 'Subcategorize' selected.

1. Negative observed data is available for the target chemical
2. All analogues are positive. Lets check how similar are the analogues with respect to structural similarity
3. **Open** "Subcategorize"
4. **Select** "Structural similarity" (options used in the read-across prediction are displayed on slide 54, same as in the previous example). All 4 analogues are similar above 50% with respect to the target chemical.

Read-across applied for reproductive tox

Scenario 1: DART scheme used as a categorization tool

Investigated endpoint: Reproductive toxicity

QSAR Toolbox 3.3.0.152 [Document]

QSAR TOOLBOX

Input Profiling Endpoint Category Definition Data Gap Filling Report

Filing

Apply

Data Gap Filling Method

- Read-across
- Trend analysis
- (Q)SAR models

Target Endpoint

Human Health Hazards Developmental Toxicity / Teratogenicity Reproductive Toxicity

Structure

1 [target] 2 3 4 5

Reproductive Toxicity (5/5) M. Not known reproductive potential M. Known reprodu. M. Known reprodu. M. Known reprodu. M. Known reprodu.

Descriptors Prediction

Read across prediction of Reproductive toxicity, taking the highest mode from the nearest 5 neighbours, based on 4 values from 4 neighbour chemicals, Observed target value: 'Negative', Predicted target value: 'Positive'

Obtained positive read-across prediction could not explain the negative observed data.

Reproductive toxicity (obs.)

log Kow

Accept prediction

Return to matrix

Select/filter data

Subcategorize

Mark chemicals by descriptor

Filter points by test conditions

Mark focused chemical

Mark focused points

Remove marked chemicals/points

Clear existing marks

Selection navigation

Gap filling approach

Descriptors/data

Model/(Q)SAR

Calculation options

Visual options

Information

Miscellaneous

1. All analogues are positive;
2. However negative observed data is available for the target chemical;
3. **Click** "Accept prediction";
4. **Click** Return to matrix

Recap

- In this step of the workflow scenario 1 has been applied:
 - Scenario 1: DART scheme applied as a primary categorization
- Read-across results shows:
 - Positive prediction for developmental toxicity reproduces positive observed data.
 - Positive prediction for reproductive toxicity could not explain the negative observed data.
- The workflow continues with second categorization scenario.

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - Endpoint
- **Category definition**
 - Overview
 - Scenario 1
 - Apply read-across
 - **Scenario 2**

Category Definition

Grouping methods

- Based on these classifications and basic guidance for grouping chemicals explained on the previous slides, two scenarios for identifying similar analogues have been applied in further read-across analysis:
 - Scenario 1: DART v1.0 scheme used as a categorization tool (applied for primary categorization - phase I).
 - Scenario 2: OFG* is used for primary categorization with forthcoming subcategorization by DART scheme (applied for subcategorization - phase II).

*OFG is used for primary categorization in this case, because the two basic functionalities available within the molecule: "Ureas" and "Aryl halides" will be preserved in the group of identified analogues, while the ECOSAR categorization omits "Aryl halide" functionality and identifies "Substituted ureas" only (see slide 44).

Category Definition

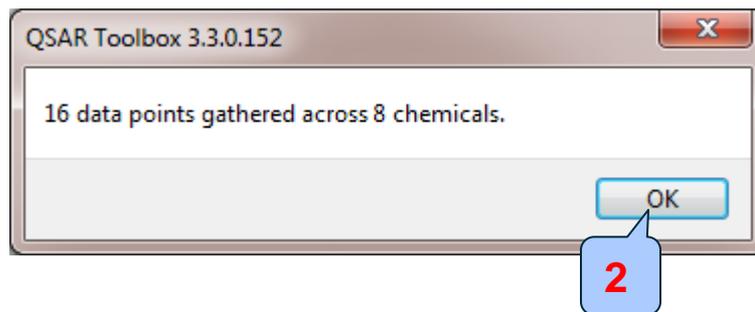
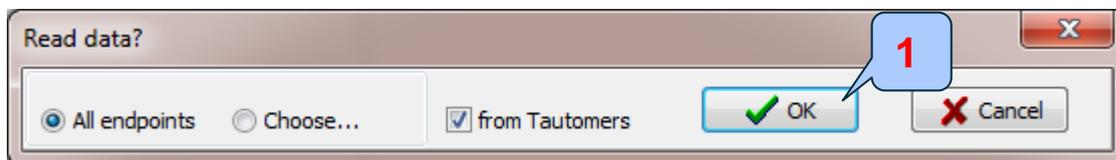
Scenario 2: OFG is used for primary categorization with forthcoming subcategorization by DART scheme

The screenshot displays the QSAR Toolbox software interface during the 'Category Definition' process. The top menu bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The toolbar below the menu contains buttons for 'Define', 'Define with metabolism', 'Subcategorize', 'Combine', 'Clustering', 'Delete', and 'Delete All'. The main workspace is divided into several sections: a 'Filter endpoint tree...' on the left, a 'Structure' view showing chemical structures, and a table of target profiles. A 'Define category name' dialog box is open, showing 'Category name (8 chemicals)' and 'Urea derivatives (Organic Functional groups)'. A 'Organic Functional groups' dialog box is also open, showing 'Target(s) profiles' and 'All profiles' lists. Red callouts with numbers 1 through 5 highlight key steps: 1. 'Category Definition' menu, 2. 'Organic Functional groups' selection, 3. 'Define' button, 4. 'OK' button in the 'Organic Functional groups' dialog, and 5. 'OK' button in the 'Define category name' dialog.

1. **Back** to "Category definition";
2. **Select** "OFG";
3. **Click** "Define"
4. **Click** "OK" to confirm the identified categories for the target chemical
5. Eight analogues are identified. **Click** "OK"

Category Definition

Scenario 2: OFG is used for primary categorization



1. **Click** "OK" to extract data for the analogues from DART database
2. 16 data points are gathered for the identified 8 analogues. **Click** "OK"

Category Definition

Scenario 2: OFG is used for primary categorization

The experimental results for the analogues appeared on datamatrix

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. Below the menu is a toolbar with 'Categorize' and 'Delete' options. The main area shows a 'Filter endpoint tree...' on the left and a data matrix on the right. The data matrix has 8 columns and multiple rows. A red box highlights a specific row with a blue callout bubble containing the number '1'.

Structure	1 [target]	2	3	4	5	6	7	8
Structure								
Substance Identity								
Physical Chemical Properties								
Environmental Fate and Transport								
Ecotoxicological Information								
Human Health Hazards								
Acute Toxicity								
Bioaccumulation								
Carcinogenicity								
Developmental Toxicity / Teratogenicity								
Developmental and Reproductive Toxicity								
Developmental Toxicity	(8/9) M: Known develop... R: Positive	M: Known develop...	M: Known develop...	M: Known develop...	M: Known develop...	M: Known develop...	M: Known develop...	M: Known develop...
Reproductive Toxicity	(8/8) M: Not known repr...	M: Undefined repro...	M: Not known repr...	M: Known reprodu...	M: Undefined repro...	M: Known reprodu...	M: Known reprodu...	M: Undefined repro...
Teratogenicity (FDA TERIS)								
Genetic Toxicity								
Immunotoxicity								
Irritation / Corrosion								
Neurotoxicity								
Photoinduced Toxicity								
Repeated Dose Toxicity								
Sensitisation								
ToxCast								

1. As mention on the previous slides there are observed data for the target, according to the two endpoints. We will try to reproduce it.

Read-across applied for developmental tox

Scenario 2: OFG is used for primary categorization

Investigated endpoint: Developmental toxicity

The screenshot shows the QSAR Toolbox 3.3.0.152 interface. The 'Data Gap Filling' menu is active, and the 'Read-across' method is selected in the left sidebar. The main table displays a list of endpoints, with 'Developmental Toxicity' selected. A dialog box titled 'Possible data inconsistency' is open, showing 'DART toxicity' selected as the gap filling scale/unit. The dialog also shows 'converted data' and 'Selected [8/8] points'. A 'Starting gap filling ...' progress bar is visible in the background.

1. **Click** on "Data Gap filling"
2. **Click** on the cell corresponding to "Developmental Toxicity" endpoint
3. **Select** "Read-across"
4. **Click** "Apply"
5. **Select** less informative scale "DART toxicity"
6. **Click** "OK"

Read-across applied for developmental tox

Scenario 2: OFG is used for primary categorization

Investigated endpoint: Developmental toxicity

The screenshot displays the QSAR Toolbox software interface. On the left, the 'Subcategorization' window is open, showing a list of grouping methods and endpoint-specific options. A red circle highlights the 'DART scheme v1.0' option, with a callout box '3' pointing to it. Below this, another red circle highlights the 'Correlation' button, with a callout box '4' pointing to it. The main workspace shows a grid of chemical structures, with the first one highlighted in blue. A callout box '5' points to the 'Accept prediction' button in the right sidebar. Below the grid, a plot titled 'Read across prediction of Developmental toxicity, taking the highest mode from the nearest 5 neighbours, based on 5 values from 5 neighbour chemicals, Observed target value: 'Positive', Predicted target value: 'Positive'' shows a scatter plot of 'Developmental toxicity (obs.)' vs 'log Kow'. A callout box '6' points to the 'Return to matrix' button in the right sidebar. A callout box '1' points to the 'Select/filter data' button in the right sidebar. A callout box '2' points to the 'Subcategorize' button in the right sidebar.

All analogues are positive. Let's check how similar are they with respect to endpoint specificity through applying DART profiling scheme. Follow the steps:

1. **Open** "Select/filter data";
2. **Select** "Subcategorize"
3. **Select** "DART scheme v1.0". There are 2 analogues having different DART toxicity that the target. They will be eliminated;
4. **Click** "Remove". Now all analogues are positive and consistent with respect to structure (OFG) and endpoint specificity (DART);
5. **Click** "Accept prediction";
6. **Click** "Return to matrix"

Read-across applied for developmental tox

Scenario 2: OFG is used for primary categorization

Investigated endpoint: Reproductive toxicity

The screenshot displays the QSAR Toolbox software interface during a 'Data Gap Filling' operation. The main window shows a chemical structure and a data table with columns for 'Structure', 'M', and 'R'. A dialog box titled 'Possible data inconsistency' is open, showing options for 'Scale/Unit' and 'Gap filling scale/unit'. The dialog box has a 'DART toxicity' option selected under 'Gap filling scale/unit'. The 'converted data' section shows '8 from scale DART toxicity original'. The 'Selected [8/8] points' section shows 'OK' and 'Cancel' buttons.

1. Positive prediction obtained for endpoint "Developmental toxicity" reproduces the positive observed data. Further read-across analysis continues with next endpoint "Reproductive toxicity". Follow the steps:
2. **Select** the cell corresponding to "Reproductive toxicity" endpoint;
3. **Select** "Read-across";
4. **Click** "Apply"
5. In our case less informative scale is used: "DART toxicity"
6. **Click** "OK"

Read-across applied for reproductive tox

Scenario 2: OFG is used for primary categorization

Investigated endpoint: Reproductive toxicity

The screenshot shows the QSAR Toolbox interface for a read-across prediction. The top navigation bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The left sidebar shows 'Data Gap Filling Method' with 'Read-across' selected. The main area displays a table of chemical structures and their predicted reproductive toxicity categories. Below the table is a scatter plot of 'Reproductive toxicity (obs.)' vs 'log Kow' with a 'Positive prediction' and 'Negative observed data' highlighted by red circles. A right sidebar contains 'Accept prediction' and 'Return to matrix' options.

Structure	1 (target)	2	3	4	5	6	7
Structure	<chem>CC1=CC=C(C=C1)C(=O)N</chem>	<chem>CC1=CC=C(C=C1)C(=O)N</chem>	<chem>CC1=CC=C(C=C1)C(=O)N</chem>	<chem>CC1=CC=C(C=C1)C(=O)N</chem>	<chem>CC1=CC=C(C=C1)C(=O)N</chem>	<chem>CC1=CC=C(C=C1)C(=O)N</chem>	<chem>CC1=CC=C(C=C1)C(=O)N</chem>
Reproductive Toxicity	(8/9) M: Not known reproductive potential R: Positive	M: Undefined repro.	M: Not known repr...	M: Known reprodu...	M: Undefined repro.	M: Known reprodu...	M: Know

Read across prediction of Reproductive toxicity, taking the highest mode from the nearest 5 neighbours, based on 5 values from 5 neighbour chemicals, Observed target value: 'Negative', Predicted target value: 'Positive'

Reproductive toxicity (obs.)

log Kow

Descriptor X: log Kow

Accept prediction

Return to matrix

- Select/filter data
- Selection navigation
- Gap filling approach
- Descriptors/data
- Model/(Q)SAR
- Calculation options
- Visual options
- Information
- Miscellaneous

1. The obtained read-across prediction could not reproduce the negative observed data of the target. It is not straightforward and not reliable due to variable endpoint data of the analogues. The purpose of the further workflow is to subcategorize and refine the initial category of analogues. Follow the steps illustrated on next slide.

Read-across applied for reproductive tox

Scenario 2: OFG is used for primary categorization

Investigated endpoint: Reproductive toxicity

The screenshot shows the 'Subcategorization' window in the OECD QSAR Toolbox. The left sidebar lists various methods, with 'DART scheme v.1.0' selected under 'Endpoint Specific'. The main area shows a table of chemical structures and their predicted reproductive toxicity categories. The table has 8 columns, each representing a different chemical structure. The predicted categories are: (8/8) M: Not known reproductive po., M: Undefined repro., M: Not known repr., M: Known reprodu., M: Undefined repro., M: Known reprodu., M: Known reprodu., M: U. Below the table is a scatter plot titled 'Read across prediction of Reproductive toxicity, taking the highest mode from the nearest 5 neighbours, based on 5 values from 5 neighbour chemicals, Observed target value: 'Negative', Predicted target value: 'Positive''. The y-axis is 'Reproductive toxicity (obs.)' with categories 'Positive', 'Undefined', and 'Negative'. The x-axis is 'log Kow' ranging from 2.00 to 4.20. The plot shows several data points, with a blue dot at approximately (4.1, Positive) and a green dot at approximately (3.1, Undefined). The right sidebar contains 'Accept prediction' and 'Return to matrix' options, with 'Subcategorize' selected. Numbered callouts (1-4) highlight key steps: 1. 'Accept prediction', 2. 'Subcategorize', 3. 'DART scheme', and 4. 'Remove'.

1. **Open** "Select filter data" to eliminate the analogues with different DART toxicity
2. **Select** "Subcategorize"
3. **Select** "DART scheme";
4. **Click** "Remove"

Read-across applied for reproductive tox

Scenario 2: OFG is used for primary categorization

Investigated endpoint: Reproductive toxicity

The screenshot displays the QSAR Toolbox interface for a read-across prediction. The top navigation bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The left sidebar shows the 'Data Gap Filling Method' set to 'Read-across' and the 'Target Endpoint' as 'Human Health Hazards Developmental Toxicity / Teratogenicity Reproductive Toxicity'. The central matrix shows 7 chemical structures with their predicted categories: 1 (target) is 'R. Positive', 2 is 'M. Undefined repro.', 3 is 'M. Not known repr.', 4 is 'M. Known reprodu.', 5 is 'M. Known reprodu.', 6 is 'M. Known reprodu.', and 7 is 'M. Known'. The bottom plot, titled 'Read across prediction of Reproductive toxicity, taking the highest mode from the nearest 5 neighbours, based on 5 values from 5 neighbour chemicals, Observed target value: 'Negative', Predicted target value: 'Positive'', shows a scatter plot of 'log Kow' (x-axis, 2.20 to 4.00) versus 'Reproductive toxicity (obs.)' (y-axis, Positive, Undefined, Negative). A callout box with the number '1' points to the 'Return to matrix' button in the bottom right panel.

In this case the obtained read-across is not reliable and we do not recommend to accept the prediction. The reason for that obtained read-across is not reliable enough is that the category members do not show similar test results, but note that it is not replicating the experimental data. 1. **Click** "Return to matrix"

Recap

- In this step of the workflow two scenarios for identifying analogues are played:
 - Scenario 1: DART scheme applied as a primary categorization.
 - Scenario 2: DART scheme applied in subcategorization procedure.
- Read-across results of scenario 2 shows:
 - The obtained positive prediction for developmental toxicity reproduces the positive observed data.
 - The obtained read-across prediction for reproductive toxicity is not reliable and could not be accepted. The reason for that obtained read-across is not reliable enough is that the category members do not show similar test results, but not that it is not replicating the experimental data.
- The further workflow continues with applying of external DART SAR model.

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - Endpoint
- **Category definition**
 - Overview
 - Scenario 1
 - Scenario 2
- **Support the prediction by DART model**

DART SAR model

- DART SAR model is based on profiling results obtained by DART profiling scheme v 1.0. The profiling scheme is implemented following Wu S. paper [1]. The SAR model follows the same organization as DART profiling scheme with an exception of prediction result. It is used for identifying chemicals with structural features associated with the potential to act as reproductive or developmental toxicants.
- The prediction outcome from the SAR model provides more general information for potential of chemical to cause DART toxicity, while the DART profiling scheme provides specific information about the DART category associated with the specific chemical class.
- The outcome from the model is “Known precedent reproductive and developmental toxic potential”, when the chemical meet the structural criteria of the model; “Not known precedent reproductive and developmental toxic potential”, if the toxic potential of the input chemical is not known and “Not covered by current version of the decision tree” if the identified structural features are not object of the DART tree.
- Prediction outcome from DART model is provided for repro/developmental toxicity as joint effect, while read-across analysis is performed for each of the DART toxicity independently (previous exercise).
- The corresponding category is displayed in the report generated for the obtained prediction.

1. Wu S, Fisher J, Naciff J, Laufersweiler M, Lester C, Daston G, Blackburn K. Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants. Chem Res Toxicol. 2013 Dec 16;26(12):1840-61.

DART SAR model - overview

The screenshot displays the QSAR Toolbox software interface. The top navigation bar includes buttons for 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling' (highlighted with a red '1'), and 'Report'. The left sidebar shows 'Data Gap Filling Method' with options: 'Read-across', 'Trend analysis', and '(Q)SAR models' (highlighted with a red '2'). Below this, the 'Target Endpoint' is set to 'Human Health Hazards Developmental Toxicity / Teratogenicity Developmental and Reproductive Toxicity (DART)'. A red circle highlights the '<< CREATE A NEW QSAR >>' button in the 'Relevant (Q)SAR Models' list (highlighted with a red '4'). A context menu is open over this button, with 'Predict' selected (highlighted with a red '5'). The central 'Filter endpoint tree...' shows a tree structure with 'Developmental and Reproductive Toxicity' selected (highlighted with a red '3'). A warning dialog box is displayed over the data matrix, asking 'Domain is not defined for selected QSAR! Do you want to continue?' with 'Yes' and 'No' buttons (highlighted with a red '6'). The data matrix table shows a prediction for 'Q: Known precedent reproductive and develop...' (highlighted with a red '7') and a callout box stating 'Prediction obtained by DART model'.

1. Go to "Data Gap Filling"
2. Select "QSAR model" (DART)", then the model will appear in the list with QSARs(4)
3. Select node "Developmental and Reproductive Toxicity"
4. Right click over the model
5. Select "Predict Current Chemical"
6. The software informs the user that the domain is not defined. Click "Yes"
7. Prediction obtained by DART model appears on datamatrix

DART SAR model - overview

QSAR TOOLBOX

Input Profiling Endpoint Category Definition Data Gap Filling Report

Filter endpoint tree... [target]

Structure

Substance Identity
Physical Chemical Properties
Environmental Fate and Transport
Ecotoxicological Information
Human Health Hazards
Acute Toxicity
Bioaccumulation
Carcinogenicity
Developmental Toxicity / Teratogenicity
Developmental and Reproductive ... (1/1)
Developmental Toxicity (8/10)
Reproductive Toxicity (8/9)
Teratogenicity (FDA TERIS)
Genetic Toxicity
Immunotoxicity
Irritation / Corrosion
Neurotoxicity
Photoinduced Toxicity
Repeated Dose Toxicity
Sensitisation
ToxCast
Toxicity to Reproduction
Toxicokinetics, Metabolism and Distribution

Known precedent reproductive and developmental potential

1

2

3

Known precedent reproductive and developmental toxic potential

Polyhalogenated benzenes (8c)

N-substituted urea (2b-4)

1. **Right click** over the prediction result and select "Explain prediction".

3. **Click** "Details"

2. **Select** first category "Known..."

DART SAR model – Explain results

Node responsible for the assigning the “Known precedent reproductive and devopmental...” is selected

ID	Node	Result	Label
0	Is the chemical organic ?	Satis...	
2	Contains a ring ?	Satis...	
4	ER/AR receptor binding rules	Satis...	
8	Retinoic acid receptor rules	Satis...	
9	Rules for nicotinic acetyl choline receptor and acetyl cholinesterase ...	Satis...	
11	Rules for ion channel opener/inhibitor, beta-adrenergic, ACE/ARA or ...	Satis...	
12	Rules for opioid/tubline receptor binding	Satis...	
13	Rules for nudeotide & nudeobase derivatives	Satis...	

DART SAR model – Explain results

Node responsible for assigning the “Known precedent reproductive and devopmental...” category to the target is selected

* Blue tick marked categories are required to be met in order “Known precedent reproductive and devopmental...” to be assigned. With green ticks are marked results obtained for current target. As it can be seen the general category “Known precedent...” is assigned to the target, because it has some of the required (blue ticks) categories.

1. **Click** on “Query tree” panel
2. **Scroll down** and **find** category marked with green tick
3. Marked category “Non-steroid nucleus derived estrogen receptor (ER)(2b-4)” is responsible for DART effect*. The category “Known precedent repro..... potential” obtained as a DART SAR results is due to this category and the other “Polyhalogenated benzene derivatives (8c)” category. More details for both schemes is provided on next slides
4. **Close** the window

DART SAR model – Explain results

N-aryl substituted urea(2b-4)

The screenshot displays the QSAR Toolbox interface. On the left, the 'Filter endpoint tree...' is expanded to show 'Developmental Toxicity / Teratogenicity' and 'Reproductive Toxicity'. The main workspace shows a grid of chemical structures. A 'Profiling results' dialog box is open, showing the following text:

ceptor (AR)
 ceptor (AR) >> N-aryl substituted urea, carbamate and amide derived androgen receptor (AR) (2b-4)

Below this text, there are two buttons: '? Details' and 'Close'.

1. Select second category "N-aryl substituted urea, carbamate and amide derived androgen receptor (AR)(2b-4)"
 2. Click "Details"

DART SAR model – Explain results

N-aryl substituted urea(2b-4)

The image shows two windows from the DART SAR model software. The left window, titled 'DART scheme v.1.0 (Endpoint Specific) - Profiling Scheme Browser', displays a complex flowchart of nodes and decision paths. A blue callout '1' points to a specific node in the 'Inorganic and organometallic' section. A blue callout '2' points to the 'View Scheme' button at the bottom. The right window, titled 'Category 2 Estrogen receptor (ER) and androgen receptor (AR) binding compounds', shows a 'N-aryl substituted urea, carbamide and amide derived androgen receptor (AR) (2b-4)' analysis. It features a 'Target' field with a chemical structure, a 'Boundaries' section with 15 numbered icons, and a 'Common Fragments' table. A blue callout '3' points to the 'View Scheme' button. A blue callout '4' points to the 4th boundary icon, which is highlighted in green. A blue callout '5' points to the chemical structure of the target compound. A blue callout '6' points to the 'Profile Description' section, which contains the following text:

Category 2: Estrogen receptor (ER) and androgen receptor (AR) binding compounds

2b. Non-steroid nucleus derived estrogen receptor (ER) and androgen receptor (AR) Binders

2b-4. N-aryl substituted urea, carbamide and amide derived androgen receptor (AR) binders.

Most androgenic chemicals activate AR-mediated transcription in mammalian cells through receptor mediated mechanisms. For example, in the sub-category of N-aryl substituted ureas, carbamides and amides, many are AR binders which display developmental toxicity potential. The general core structural requirement of Ph-N-CO-X in (2b-4-1) and (2b-4-2) as shown in Figure S9, appears to be important for AR binding. The substituents associated with activity are further enumerated below. SAR analysis indicates that electron-withdrawing groups on the benzene ring, such as F, Cl, NO₂, or CN favor AR binding.^{1,2} In some cases, the substituents on X in (2b-4-1) could be fused with the NH to form five membered heterocyclic

1. Node responsible for DART effect, which assigns the respective category is selected
 2. **Click** "Referential node" tab
 3. **Click** "View Scheme"
 4. **Select** green marked boundary #4. It is marked green because the target met the criteria of the boundary.
 5. Definition of structural boundary
 6. Textual description associated with DART category
- Details about the category is presented on next slide

DART SAR model – Explain results

N-aryl substituted urea(2b-4)

Profile Description

Category 2: Estrogen receptor (ER) and androgen receptor (AR) binding compounds

2b. Non-steroid nucleus derived estrogen receptor (ER) and androgen receptor (AR) Binders

2b-4. N-aryl substituted urea, carbamide and amide derived androgen receptor (AR) binders.

Most androgenic chemicals activate AR-mediated transcription in mammalian cells through receptor mediated mechanisms. For example, in the sub-category of N-aryl substituted ureas, carbamides and amides, many are AR binders which display developmental toxicity potential. The general core structural requirement of Ph-N-CO-X in (2b-4-1) and (2b-4-2) as shown in Figure S9, appears to be important for AR binding. The substituents associated with activity are further enumerated below. SAR analysis indicates that electron-withdrawing groups on the benzene ring, such as F, Cl, NO₂, or CN favor AR binding.^{1,2} In some cases, the substituents on X in (2b-4-1) could be fused with the NH to form five membered heterocyclic ring moieties (2b-4-2), exemplified by imidazolidine-2,4-dione, oxazolidine-2,4-dione, and pyrrolidine-2,5-dione derivatives. For these cyclic compounds, X can be nitrogen, oxygen or carbon, and substituents (R, R₁, R₂ and R₃) are listed in Figure S9 in structure (2b-4-2). These N-aryl substituted heterocyclic ring derivatives, such as vinclozolin (CAS# 50471-44-8), iprodione (CAS# 36734-19-7) and proyamidone (CAS# 32809-16-8), have a range of developmental and reproductive effects linked to activity as an anti-androgen.^{3, 4} Another androgen antagonist is prochloraz (CAS# 67747-09-5), which appears to have a distinct pattern of toxicity. Prochloraz had been reported to have multiple effects on the development of male rodents, and *in vitro* data show not only anti-androgen but also anti-estrogen effects as well as interaction with Ah receptors and inhibition of aromatase^{5, 6}.

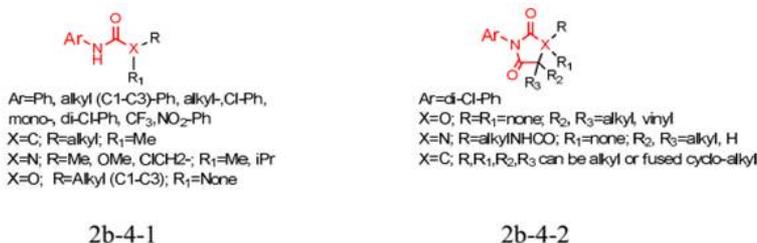


Fig. S9. The general structural features of N-aryl substituted ureas, carbamides and amides like chemicals.

Original reference:

Shengde, W., Joan, F., Jorge N., Michael L., Cathy L., George D., and Karen B.,(2013) Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants. *Chem.Res.Toxicol.* 26(12), 1840-1861.

References cited in the original article:

1. Fang, H., Tong, W., Branham, W., Moland, C., Dial, S., Hong, H., Xie, Q., Perkins, R., Owens, W.,Sheehan, D. (2003) Study of 202 natural, synthetic, and environmental chemicals for binding to the androgen receptor. *Chem Res Toxicol* 16 1338-1358

DART SAR model – Explain results

Polyhalogenated benzene derivatives (8c)

The screenshot displays the QSAR Toolbox interface. The top navigation bar includes buttons for Input, Profiling, Endpoint, Category Definition, Data Gap Filling, and Report. The left sidebar shows the 'Data Gap Filling Method' (Read-across, Trend analysis, QSAR models) and 'Relevant (Q)SAR models'. The main area shows a 'Filter endpoint tree...' on the left and a grid of chemical structures in the center. A 'Profiling results' dialog box is open, showing a 'Chemical profile' with a 'DART scheme v.1.0' section. A red circle highlights the category 'Polyhalogenated benzene derivatives (8c)' in the DART scheme. A blue callout box with the number '1' points to this category, and another blue callout box with the number '2' points to the 'Details' button in the dialog box.

1. **Select** second category "Polyhalogenated benzene derivatives (8c)"
2. **Click** "Details"

DART SAR model – Explain results

Polyhalogenated benzene derivatives (8c)

The image shows two overlapping software windows. The left window is the 'DART scheme v.1.0 (Endpoint Specific) - Profiling Scheme Browser', displaying a complex flowchart of SAR nodes. A red arrow points from a node in the 'Endocrine toxicity' section to a callout box labeled '1'. A 'View Scheme' button is highlighted with a callout box labeled '3'. The right window is 'Category 8: Aromatic compounds with alkyl, multi-halogen and nitro groups', showing a 'Polyhalogenated benzene derivatives (8c)' boundary. A green boundary is highlighted with a callout box labeled '4'. A chemical structure of a polyhalogenated benzene derivative is shown with a callout box labeled '5'. A 'Training set' table is highlighted with a callout box labeled '7'. A 'Profile Description' section is highlighted with a callout box labeled '6'. A 'Referential Node' tab is highlighted with a callout box labeled '2'.

1. Node responsible for DART effect, which assigns the respective category is selected
 2. Click "Referential node" tab
 3. Click "View Scheme"
 4. Select green marked boundary #1. It is marked green because the target met the criteria of the boundary.
 5. Definition of structural boundary
 6. Textual description associated with DART category
 7. Training set associated with category
- Details about the category is presented on next slide

DART SAR model – Explain results

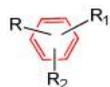
Polyhalogenated benzene derivatives (8c)

Category 8: Aromatic compounds with alkyl, multi-halogen and nitro groups

Chemicals within this category include the following five sub-categories: 8a. toluene and small alkyl toluene derivatives; 8b. NO₂-alkyl/NO₂-benzene derivatives; 8c. polyhalogenated benzene derivatives; 8d. polyhalogenated-, NO₂/halogenated-oxydibenzene; 8e. dihalogen-, dinitro-phenol and their ester derivatives. The general core structures of these chemicals include the toluene, oxydibenzene and phenol ring with alkyl, halogen and/or nitro substituents as shown in (8a to 8e) in Figure S24.

8c. Polyhalogenated benzene derivatives

For sub-category 8a, toluene and a single alkyl chain substituent (< 5 carbon atoms) present on toluene are included. The alkyl substituents can be at ortho, para or meta-positions. For subcategory 8b, the majority of chemicals are mono-, di-, tri-nitrobenzene or nitrotoluene with ortho, para or meta relative substituent placement. Members of 8a and 8b without nitro substituents appear to be primarily developmental toxicants, while addition of a nitro group may be associated with a distinctive pattern of male reproductive toxicity. Sub-category 8c includes multi-chlorinated benzene derivatives containing from 2 to 6 chlorine atoms. Other possible substituents include methyl or nitrile groups. The members of this class included here are primarily developmental toxicants (Appendix 1). Sub-category 8d includes multi-substituted oxydibenzene with halogen or halogen/nitro substituents. Because these chemicals normally do not readily form co-planar structures, they are not anticipated to bind to the AhR, (see section 3b-2), nor do they have a mode of action dependent on the AhR interaction.²⁹



R=R₁=R₂=Cl
 (# of Cls from
 1 to 6)

8c

Fig. S24. The structural scope of alkyl substituted benzene, alkyl/NO₂-substituted benzene, polyhalogenated benzene, oxydibenzene, poly-halogenated, poly-NO₂phenol and their esters

Original reference:

Shengde, W., Joan, F., Jorge N., Michael L., Cathy L., George D., and Karen B. (2013) Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants. *Chem.Res.Toxicol.* 26(12), 1840-1861.

References cited in the original article:

1. ATSDR (Agency for Toxic Substances and Disease Registry) (2004) Toxicological profile for polybrominated biphenyls and polybrominated diphenyl ethers.

Interpretation of SAR results

- Obtained SAR results show, that target chemical may elicit developmental and reproductive toxic potential based on belonging to the two DART toxic categories:
 - N-aryl substituted urea(2b-4)
 - Polyhalogenated benzene derivatives (8c)
- Both DART toxic categories are characterized with mechanistic interpretation and training set chemicals with observed DART data.

Summary

- Toolbox 3.3 includes two application of DART scheme as discussed by Wu S paper:
 - DART scheme that can be used as a profiler for category formation or
 - DART SAR model for obtaining results based on DART profiling scheme and DART training set database.

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - Endpoint
 - Category definition
 - Support the prediction by DART model
- **Save predictions**

Saving the prediction result

- This functionality allows storing/restoring the current state of Toolbox documents including loaded chemicals, experimental data, profiles, predictions etc., on the same computer. The functionality is implemented based on saving the sequence of actions that led to the current state of the Toolbox document and later executing these actions in the same sequence in order to get the same result(s).
- Saving/Loading the file with TB prediction is shown on next screenshots.

Saving the prediction result

1

2

3

4

8 Aryl<AND> Aryl halide<AND> Urea derivatives (Organic Functional groups) 1/0/0

1. **Go** to "Input" section; 2. **Click** on "Save" button; 3. **Browse** and **put** name of the file; 4. **Click** "Save" button

Open saved file

The screenshot displays the QSAR Toolbox software interface. On the left, the 'Documents' panel shows a new document being created, indicated by callout '1'. The top toolbar features the 'Open' icon, highlighted by callout '2'. A 'Select file' dialog box is open, showing a file list with 'Tutorial 24.tbw' selected, indicated by callout '3'. The dialog's 'Open' button is highlighted by callout '4'. The main interface shows a 'Filter endpoint tree...' panel with 'Substance Identity' selected.

1. **Create** new document; 2. **Click** "Open"; 3. **Find** and **select** file; 4. **Click** "Open"

Open saved file

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes options like Input, Profiling, Endpoint, Category Definition, Data Gap Filling, and Report. Below this is a toolbar with icons for New, Open, Close, Save, and various data management functions. The main workspace is divided into several panes:

- Documents:** A sidebar on the left showing a file tree with folders like 'Document', 'CAS: 330-54-1', and 'Tutorial 24.tbw'.
- Filter endpoint tree...:** A central pane showing a tree view of endpoints such as 'Substance Identity', 'Physical Chemical Properties', 'Environmental Fate and Transport', 'Ecotoxicological Information', 'Human Health Hazards (8/17)', 'Profile', 'Endpoint Specific', 'Empiric', and 'Organic Functional groups'.
- Chemical List:** A table with 8 columns, each containing a chemical structure. The first column is labeled '1 [target]'. Below the structures, there are labels like 'M: Known develop...'. A blue callout bubble with the number '1' points to an 'OK' button in a dialog box.
- Information Dialog:** A small window in the center with the message 'The file was executed successfully' and an 'OK' button.
- Bottom Panel:** A section containing a chemical structure, its SMILES string CN(C)C(=O)Nc1ccc(Cl)c(Cl)c1, and buttons for 'select filter type ..', 'Create', and 'Apply'.

The file is opened successfully

1. Click "OK"