Workshop on the use of the QSAR Toolbox

Feedback from Industry Users and Development Needs

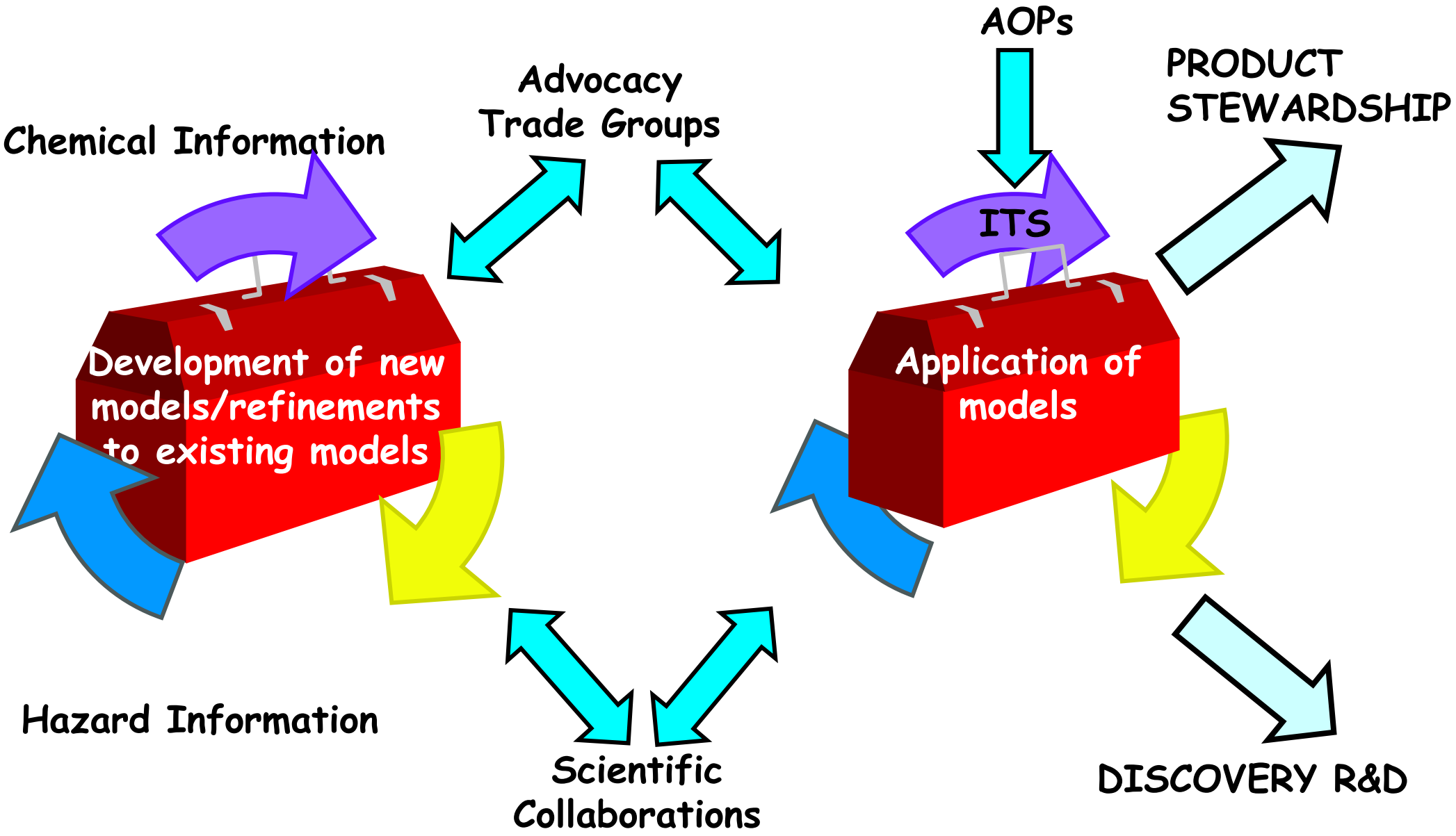
Grace Patlewicz
DuPont Haskell Global Centers
Newark, DE

DUPONT
The miracles of science™

Outline

- Predictive Toxicology at DuPont
- Use of (Q)SARs/grouping approaches for REACH 2010
- Approaches employed and challenges faced
- Next steps short & mid term
- Concluding remarks

Predictive Tox @ DuPont



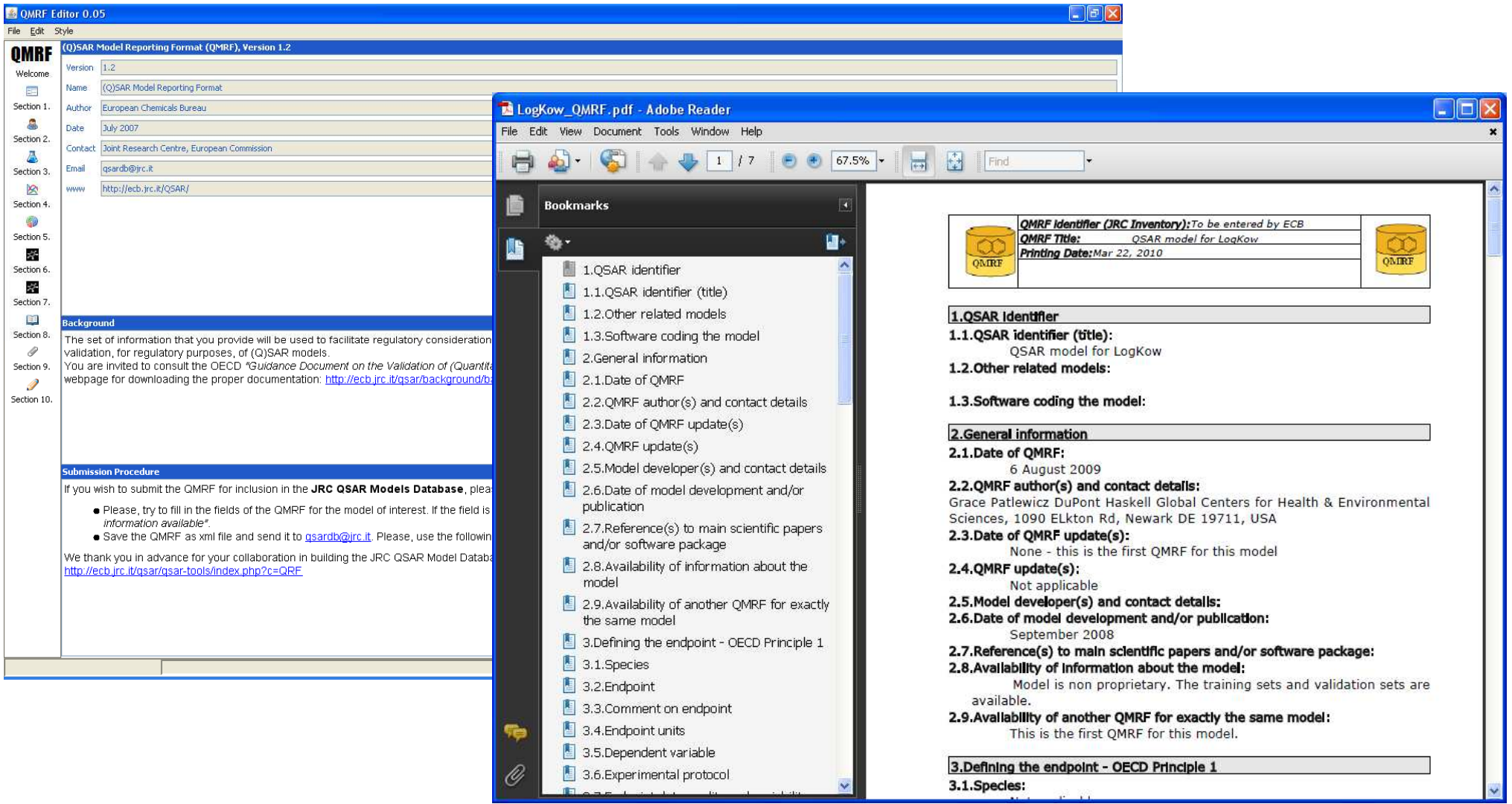
Use of (Q)SAR/grouping approaches under REACH 2010

Use of (Q)SAR/grouping approaches under REACH 2010: QSARs

- For certain physchem properties, notably LogKow we used external QSARs as full replacements to experimental testing
- An external model would be characterised in accordance with the OECD Principles as far as possible
- If no domain was described by an external model, one would be defined and this together with as much information as feasible regarding the OECD Principles would be described in an associated QMRF
- The domain could take the form of structural domain based on fragments, descriptor ranges or mode of action information depending on the basis of the QSAR model
- For the LogKow model, a structural domain was extracted on the basis of structural fragments

Use of (Q)SAR/grouping approaches under REACH 2010: QSARs

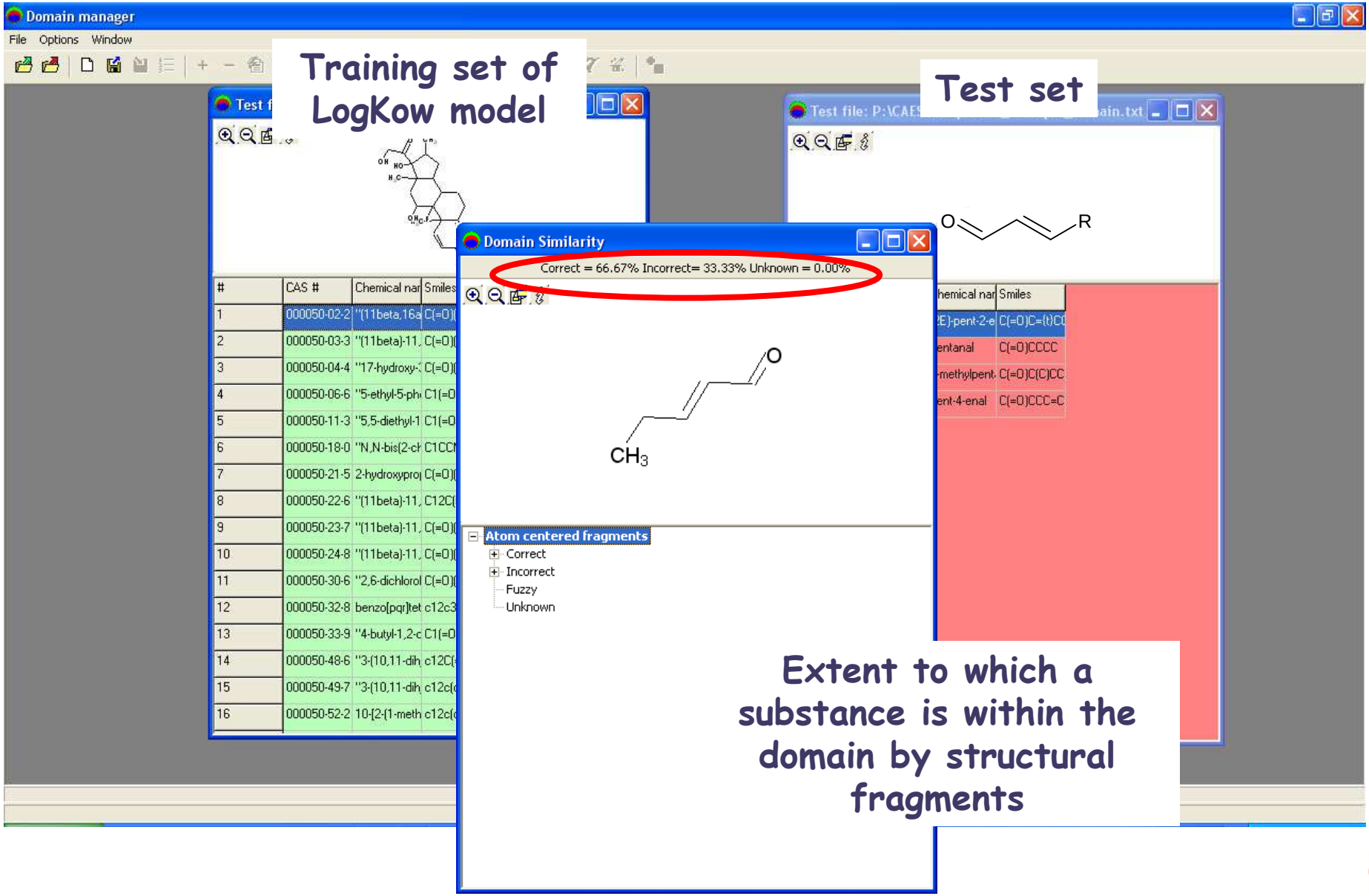
- The JRC editor was used as a mean of generating QMRFs



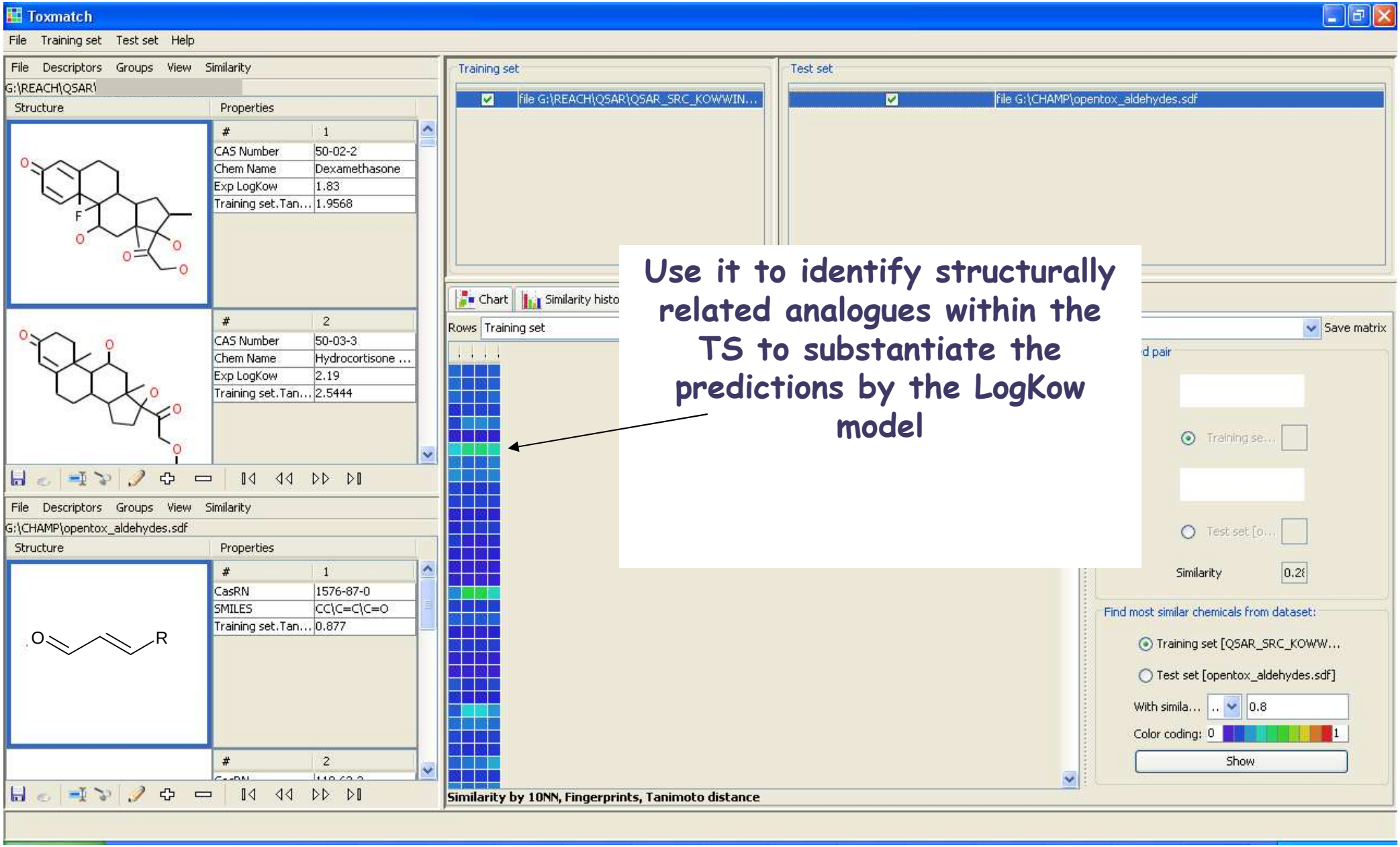
Use of (Q)SAR/grouping approaches under REACH 2010: QSARs

- Compiled many QPRFs to justify as far as possible the relevance of a given QSAR in terms of it satisfying the domain criteria and by showing that "similar" analogues had predictions which were in good agreement with their experimental values
- Similar analogues were found either from the underlying training sets or examples were identified using Leadscope or the OECD Toolbox
- Toxmatch and Leadscope were found to be a convenient means of identifying structurally related analogues through similarity indices or clustering approaches

QSARs: Structural Domain assessment documented in associated QMRF & QPRF



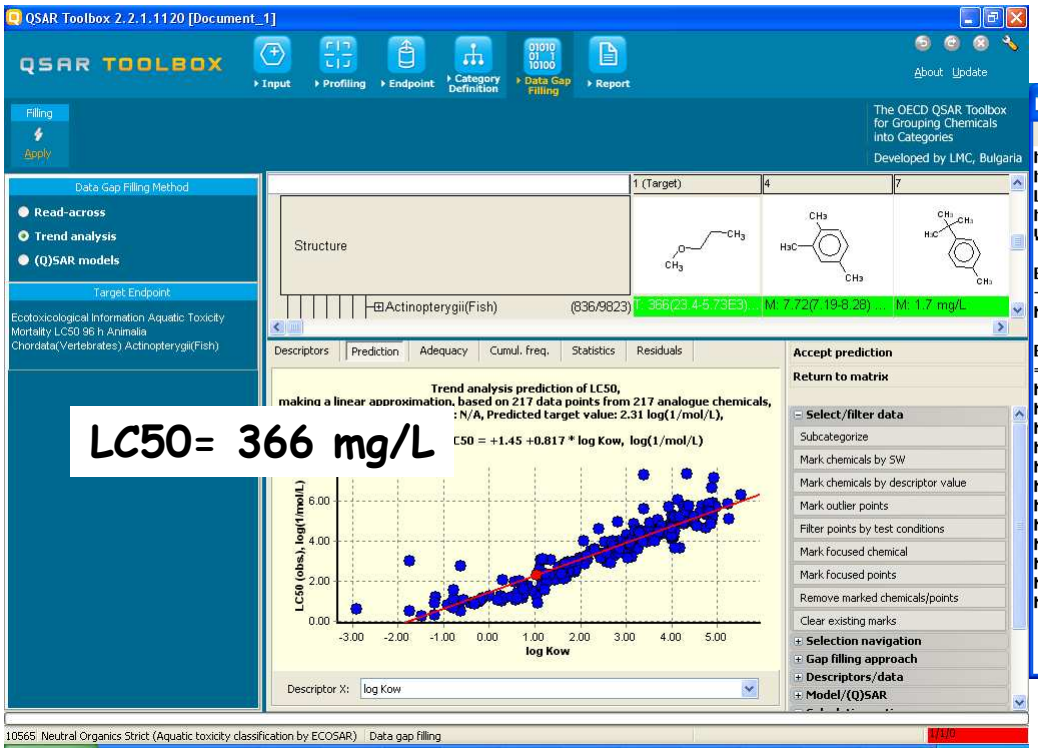
QSARs: Structural analogues for inclusion into the QPRF



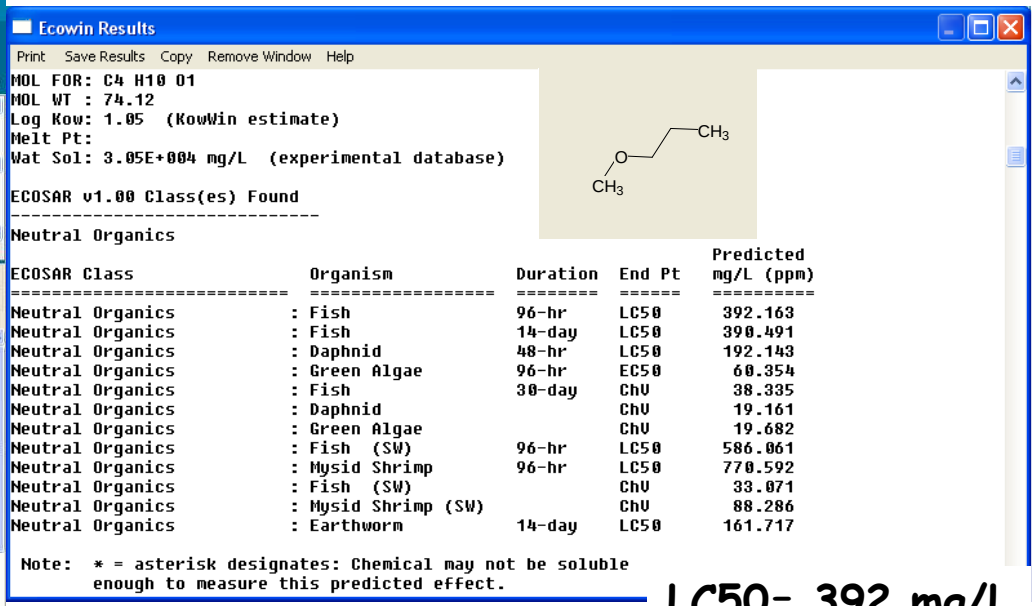
Use it to identify structurally related analogues within the TS to substantiate the predictions by the LogKow model

Use of (Q)SAR/grouping approaches under REACH 2010: QSARs

- QSARs were also extensively used for aquatic toxicity endpoints as replacement values. A combination of external QSARs and endpoint specific categories developed within the Toolbox were applied to fulfill datagaps for acute aquatic toxicity to fish, daphnid or algae.



LC50= 366 mg/L



LC50= 392 mg/L

Use of (Q)SAR/grouping approaches under REACH 2010: QSAR

- The OECD Toolbox was sometimes used as a source of data from which to develop new QSAR models outside of the Toolbox environment – particularly if more than 1 descriptor was needed to derive an algorithm or if an approach merited descriptors not implemented in the Toolbox
- QSARs were also used as supporting information to substantiate studies of less than ideal quality (per Klimisch codes) as part of a WOE approach or to provide more justification for waivers (e.g. biodegradation e-fate) or as a means to substantiate the context of similarity for an endpoint as part of a category approach (see later)

Use of (Q)SAR/grouping approaches under REACH 2010: QSAR

- Extensive use of QSAR and the Toolbox for physchem and aquatic toxicity endpoints as replacement values
- QSAR/Toolbox applied to provide supporting information for e-fate and mammalian endpoints

Use of (Q)SAR/grouping approaches under REACH 2010: Grouping

- Grouping approaches (categories) have been used in a handful of cases where there have been several data gaps to fill and where traditionally QSARs are less well developed
- Typically the categories were small – more like an analogue approach or else a limited category (2 or 3 members at most)
- Whilst obviously a larger category is considered more robust (a trendline with >3 data points is better, more connective tissue to substantiate the similarity..) there were practical challenges of deriving categories of larger sizes e.g. cost of data access, complexicity within IU5, level of information needed for source analogues (robust study summaries)…

Use of (Q)SAR/grouping approaches under REACH 2010: Grouping

- Scientifically it makes sense to form larger groups, and the Toolbox is geared to facilitate this in terms of endpoint specific categories but from a practical perspective, it has not proved to be feasible

Annex XI of REACH – grouping and read-across

A chemical category is a group of chemicals whose physico-chemical 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. The similarities may be based on the following:

- common functional group(s) e.g. aldehyde
- common constituents or chemical classes, similar carbon range numbers e.g. UVCB substances
- an incremental and constant change across the category e.g. a chain-length category for boiling point range;
- the likelihood of common precursors and/or breakdown products, via physical or biological processes, which result in structurally similar chemicals

Annex XI of REACH – grouping and read-across

If the group concept is applied, substances shall be classified and labelled on this basis.

In all cases results should:

- be **adequate** for the purpose of classification and labelling and/or risk assessment
- have adequate and reliable coverage of the key parameters addressed in the corresponding test method referred to in Article 13(3)
- cover an exposure duration comparable to or longer than the corresponding test method referred to in Article 13(3) if exposure duration is a relevant parameter, and
- **adequate and reliable documentation** of the applied method shall be provided

Use of (Q)SAR/grouping approaches under REACH 2010: Grouping

- Small categories or analogue approaches have been so far constructed using either structural similarity + similarity in functionality or structural similarity + breakdown products
- Other source analogues are often discussed to help substantiate the expected effects for different endpoints on an endpoint per endpoint basis. Thus these source chemicals were often relied upon as supplementary information to add a pseudo weight of evidence as to the validity of the grouping
- QSARs and the OECD Toolbox profilers were extensively relied upon to provide a context of similarity that could be discussed with respect to the observed endpoint effects

Use of (Q)SAR/grouping approaches under REACH 2010: Grouping

- Approach was to provide as much discussion for all endpoints regardless of whether the data gap needed to be filled or not. Aim was to try and demonstrate consistency of effects across a range of endpoints
- Adequate and reliable documentation was interpreted to mean providing an extensive CRF/ARF to describe the inferences and justify the similarity between the target and source substance(s)
- Drafting the ARF/CRF and providing what was perceived to be the necessary information has proven to be a very manual exercise not facilitated by either IU5 or the Toolbox e.g. a data matrix export from the Toolbox would be great to provide a snapshot

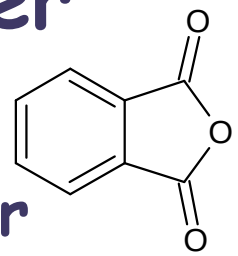
Use of (Q)SAR/grouping approaches under REACH 2010: Grouping

- Not yet able to use own REACH data in the Toolbox as mapping between IU5 and Toolbox is not optimised The export from TB to IU5 works great, the other way around is not, lots of mappings still needs to be teased out. e.g. own data is typically mapped to an "undefined field" even if exported IU5 is pre version 5.3
- Overall approach had been to formulate a hypothesis for why the grouping was relevant and then substantiate it with reference to QSAR/Toolbox profiler information coupled with empirical data. Other analogues with associated data were used as supporting information

Use of (Q)SAR/grouping approaches under REACH 2010: Grouping

- Since toxicokinetic information is not typically readily available, groupings were reasoned based on available toxicity information coupled predominantly by chemical reactivity inferences especially for endpoints where covalent binding could be considered a molecular initiating event (MIE) in the context of an AOP
- There is merit and interest to group on the basis of common transformation route e.g. hydrolysis, metabolism To that end may be useful to have a hydrolysis simulator within the Toolbox..
- Approach has been to make the hypothesis and write the justification independent of the Toolbox and supplement with what qualitative TK data might be available or simply based on other data experimental or estimated

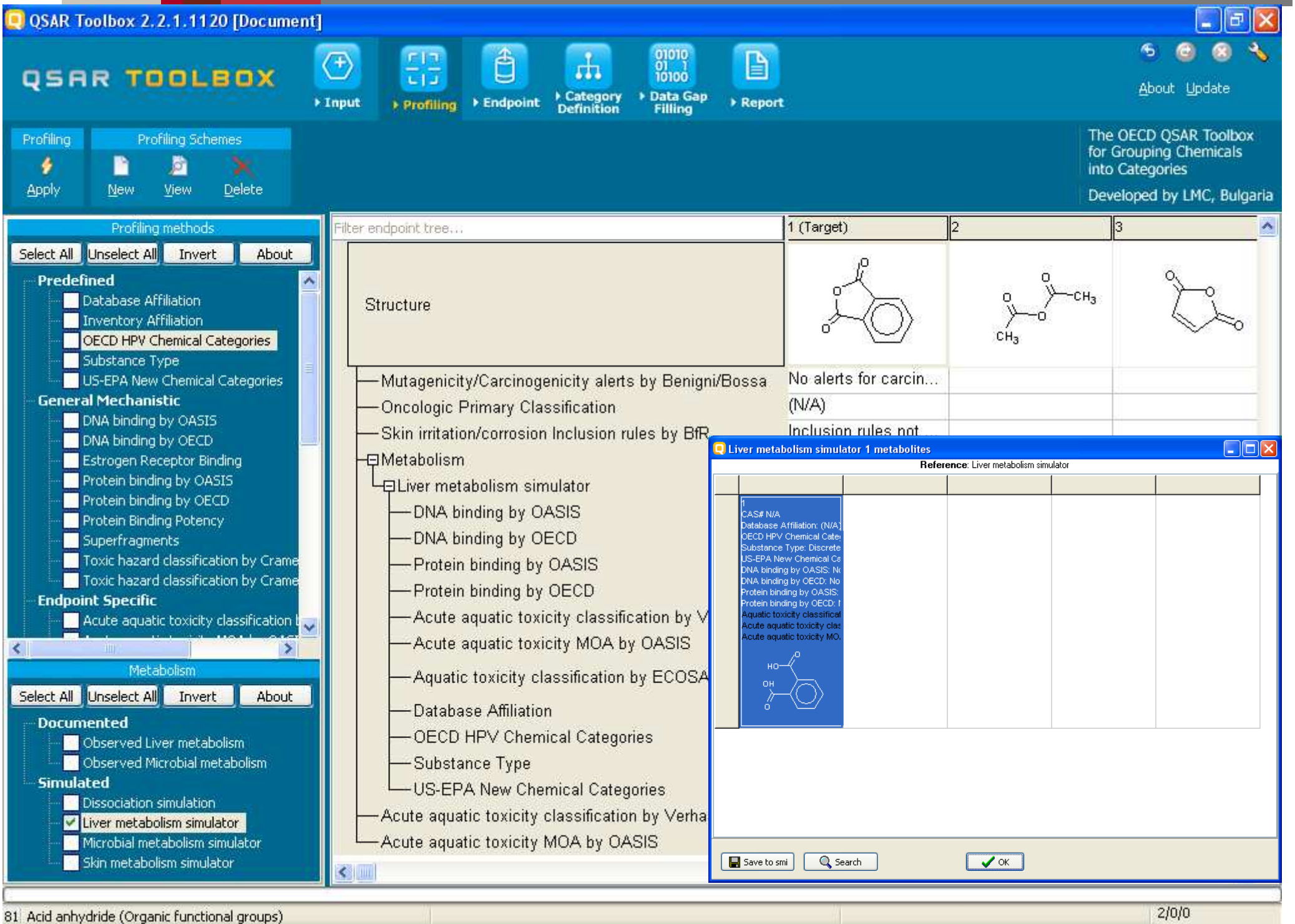
Use of (Q)SAR/grouping approaches under REACH 2010: Grouping



- Example acid anhydrides e.g. phthalic anhydride – for sensitisation read-across is not appropriate between hydrolysis products and the parent anhydrides. Acids are non electrophilic whereas anhydrides are capable of acting as acylating agents

- On the otherhand any aquatic toxicity is likely to be due to the hydrolysis products, equally systemic toxicity is likely to be driven by the degradate acids rather than the parent anhydrides

- Non trivial to approach this in the Toolbox – manually add each degradate?..develop 2 separate categories? but experiments may have been conducted on the parent..2 categories based on acids and acid anhydrides merged…? How to construct data matrix? 2 targets in the category?


QSAR Toolbox 2.2.1.1120 [Document]
QSAR TOOLBOX
Input
Profiling
Endpoint
Category Definition
Data Gap Filling
Report
About
Update
Profiling
Apply
Profiling Schemes
New
View
Delete
The OECD QSAR Toolbox for Grouping Chemicals into Categories
Developed by LMC, Bulgaria
Profiling methods
Select All
Unselect All
Invert
About
Predefined
Database Affiliation
Inventory Affiliation
OECD HPV Chemical Categories
Substance Type
US-EPA New Chemical Categories
General Mechanistic
DNA binding by OASIS
DNA binding by OECD
Estrogen Receptor Binding
Protein binding by OASIS
Protein binding by OECD
Protein Binding Potency
Superfragments
Toxic hazard classification by Crame
Toxic hazard classification by Crame
Endpoint Specific
Acute aquatic toxicity classification
Metabolism
Select All
Unselect All
Invert
About
Documented
Observed Liver metabolism
Observed Microbial metabolism
Simulated
Dissociation simulation
Liver metabolism simulator
Microbial metabolism simulator
Skin metabolism simulator
Filter endpoint tree...
Structure
1 (Target)
2
3
Mutagenicity/Carcinogenicity alerts by Benigni/Bossa
No alerts for carcin...
Oncologic Primary Classification
(N/A)
Skin irritation/corrosion Inclusion rules by BfR
Inclusion rules not ...
Metabolism
Liver metabolism simulator
DNA binding by OASIS
DNA binding by OECD
Protein binding by OASIS
Protein binding by OECD
Acute aquatic toxicity classification by V
Acute aquatic toxicity MOA by OASIS
Aquatic toxicity classification by ECOSA
Database Affiliation
OECD HPV Chemical Categories
Substance Type
US-EPA New Chemical Categories
Acute aquatic toxicity classification by Verha
Acute aquatic toxicity MOA by OASIS
Liver metabolism simulator 1 metabolites
Reference: Liver metabolism simulator
1
CAS# N/A
Database Affiliation: (N/A)
OECD HPV Chemical Cate
Substance Type: Discrete
US-EPA New Chemical Ca
DNA binding by OASIS: N
DNA binding by OECD: No
Protein binding by OASIS:
Protein binding by OECD:
Aquatic toxicity classifica
Acute aquatic toxicity clas
Acute aquatic toxicity MO.
Save to smi
Search
OK
81 Acid anhydride (Organic functional groups)
2/0/0

Short term next steps

- Integration of IU5 and the Toolbox
- Adding more data either other literature data, C&L data, REACH dissemination data
- Resolving some of the practical difficulties between endpoint specific categories with analogue/small category approaches

Mid-Longer term

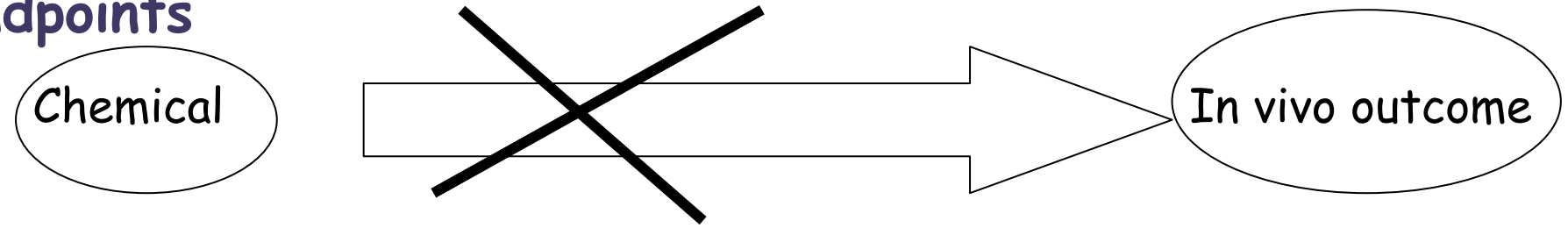
- Profilers for MIEs for associated AOPs
- Capability of integrating non standard data e.g. HTS data such as that from Toxcast

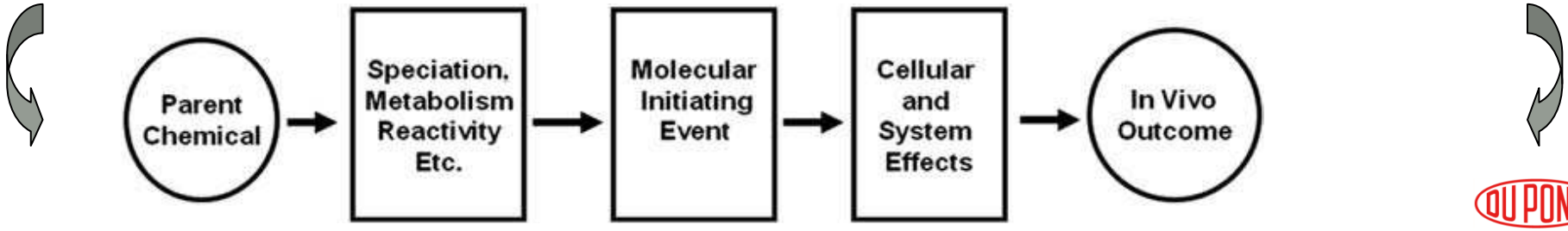
Concluding remarks

- Aimed to take advantage of (Q)SAR/grouping approaches for REACH submissions
- OECD Toolbox has been a tremendous tool to assist in filling datagaps and providing a context of similarity in the types of analogue/category approaches attempted
- The Toolbox has also proved invaluable for any sort of read-across question for any purpose whether it be REACH, other regulatory programmes or general internal product stewardship. We have used it in support of registrations in the other regions
- Has much wider application, utility and value than just REACH

Concluding remarks

- Evolvement with more data, some additional utility to facilitate the exchange of data between IU5 and Toolbox is critical to exploit its functionality fully
- Resolvement of the apparent disconnect between how to form categories within the Toolbox and how categories can practically be developed for REACH in terms of the information that needs to be provided within IU5 for source analogues
- Future work has to focus on AOPs, creating libraries for MIEs to help develop meaningful categories for more complex endpoints





Workshop on the use of the QSAR Toolbox
Feedback from Industry Users and Development Needs: Technical Features

Gina Blankenship, Grace Patlewicz

DuPont Haskell Global Centers for Health & Environmental Sciences

Newark, DE 19711

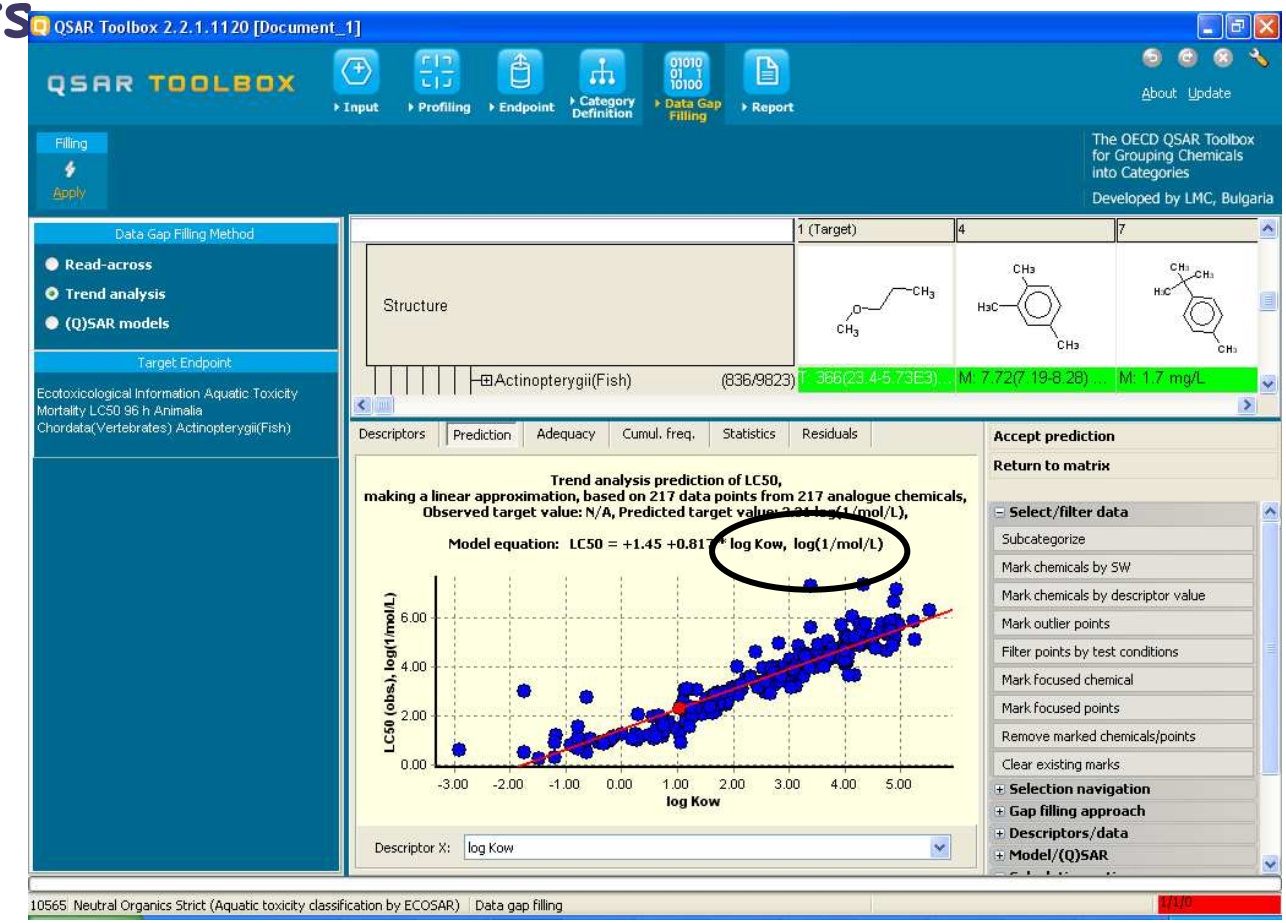
USA

Technical aspects of the functionality of the Toolbox: Guidance, Documentation & Version Comparison

- Lots of new features are added which is great, adds to the capability of the Toolbox
- Some features have been apparently removed and it is not always clear whether there was a rationale for their removal or whether the capability still exists
- E.g. Importing of a local database in version 1 a file would be created as explained in the Guidance document
- In version 2, no local file is created – is it merged with the main database? How can a local database be efficiently shared with another person? Does that lead to a merging of propietary with public within a standalone version?

Technical aspects of the functionality of the Toolbox: Guidance, Documentation & Version Comparison

- Having the predicted outcome in recognisable units e.g. mg/L was very helpful, now the default is in log units



Technical aspects of the functionality of the Toolbox: Guidance, Documentation & Version Comparison

- Discussion Forum is great but difficult to navigate through unless have been a follower from Day 1 – need a means of archiving the discussions or categorising based on the different versions
- On-line guidance (as in integrated with the Toolbox or fired up from within the Toolbox) would be a useful addition
- Guidance documentation is a mix of both versions

Searching/Data gathering

- Incremental saving or saving of session: need to complete a task or start again each time
- Can save a document which has the list of structures started with or can save a model or report…but can't save the workflow when it is still in progress
- Sharing work across multiple installation (importing of local files – how to share databases efficiently)
- Searching by structural features – can a flexible search query be constructed elements + acyclics?

Exports & Reports

- Endpoint specific category export to IU5 works great
- Are there plans for additional exporting options?
- E.g data matrix would be a convenient means of populating a ARF/CRF

Structure				
⊞Substance Identity				
⊞Physical Chemical Properties (2/9)	M: 196 °C, -0.54, -76 °C, 0.126 mm Hg, 1.00E+6 mg/L	M: 285 °C, -7 °C, 0.00132 mm Hg, 1.00E+6 mg/L		
⊞Environmental Fate and Transport (2/5)	M: 2.23E-8 atm-m3/mole, 5.72E-11 cm3/molecule-sec	M: 1.7E3, 1.7E3 L/kg wet, 58.5 %		
⊞Ecotoxicological Information (2/180)	M: >5E3 mg/L, 1.07E+4(5.58E3-2.09E+4) mg/L, 1.24...	M: >5E3 mg/L, >1E4 mg/L, >1E4 mg/L, 5 parts per million, ...		
⊞Human Health Hazards (3/82)	M: 23.4 mg/kg/day, 75.3 mg/kg/day, 292 mg/kg/day, 7...	M: 1.57E3 mg/kg/day, 3.92E3 mg/kg/day, 3.92E3 mg/kg/da...		M: Negative, not converted
⊟Profile				
—DNA binding by OASIS	No binding	No binding	No binding	No binding
—DNA binding by OECD	No Binding	No Binding	No Binding	No Binding
—Estrogen Receptor Binding	Non binder, non cyclic structure	Non binder, non cyclic structure	Non binder, non cyclic structure	Non binder, non cyclic structure
—Protein binding by OASIS	No binding	No binding	No binding	No binding
—Protein binding by OECD	No binding	No binding	No binding	No binding
—Protein Binding Potency	Not possible to classify according to these rules	Not possible to classify according to these rules	Not possible to classify according to these rules	Not possible to classify according to
—Superfragments	Has superfragment C(O)COR{*}	Has superfragment C(O)COR{*} C(O)COR{*}	No superfragment	Has superfragment C(O)COR{*}
—Toxic hazard classification by Cramer (original)	High (Class III)	High (Class III)	High (Class III)	High (Class III)
—Toxic hazard classification by Cramer (with extension)	High (Class III)	High (Class III)	High (Class III)	High (Class III)
—Acute aquatic toxicity classification by Verhaar	Class 5 (Not possible to classify according to these rul...	Class 5 (Not possible to classify according to these rules)	Class 5 (Not possible to classify according to ...	Class 5 (Not possible to classify acc
—Acute aquatic toxicity MOA by OASIS	Basesurface narcotics	Basesurface narcotics	Basesurface narcotics	Basesurface narcotics
—Aquatic toxicity classification by ECOSAR	Neutral Organics	Neutral Organics	Neutral Organics	Neutral Organics
—Bioaccumulation – metabolism alerts	Aliphatic alcohol [-OH] Methyl [-CH3] -CH2- [linear] Aliphatic ether [C-O-C]	Aliphatic alcohol [-OH] -CH2- [linear] Aliphatic ether [C-O-C]	Aliphatic alcohol [-OH] -CH2- [linear]	Aliphatic alcohol [-OH] Methyl [-CH3] -CH2- [linear] Aliphatic ether [C-O-C]
—Bioaccumulation – metabolism half-lives	Very fast	Very fast	Very fast	Very fast
—Biodegradation fragments (BioWIN MITI)	Aliphatic alcohol [-OH] Methyl [-CH3] -CH2- [linear] Aliphatic ether [C-O-C]	Aliphatic alcohol [-OH] -CH2- [linear] Aliphatic ether [C-O-C]	Aliphatic alcohol [-OH] -CH2- [linear]	Aliphatic alcohol [-OH] Methyl [-CH3] -CH2- [linear] Aliphatic ether [C-O-C]
—Eye irritation/corrosion Exclusion rules by BfR	(Undefined)Group All Lipid Solubility < 0.01 g/kg	(Undefined)Group All Lipid Solubility < 0.01 g/kg	(Undefined)Group All Lipid Solubility < 0.01 g/kg	(Undefined)Group All Lipid Solubility
—Eye irritation/corrosion Inclusion rules by BfR	Inclusion rules not met	Inclusion rules not met	Inclusion rules not met	Inclusion rules not met

IUCLID Import

- Mapping is an issue between IU5 and the Toolbox – data from IU5 is mapped to the right overall endpoint but as "undefined" => Phrase_T102 error
- In V5,1 basic features import works but in v5.2 numerous error messages – these are presented in an open log window but are not saveable or logged in any file – hence very difficult to resolve errors
- Undefined endpoint suggests that the data is not being picked up from the right places in IU5 e.g. aquatic toxicity mortality is selected but the LC50 results are not pulled across
- Also means that can not exploit data from both the Toolbox and IU5 – as no conversion scale can be added e.g. how to translate non-sensitising in IU5 with negative in the Toolbox

Database import and management

- Where does the db file for user imports of proprietary data reside? Is this a separate file that can be shared or is it merged with the overall database?
- How to handle import of data and endpoints that are not already in the Toolbox e.g. Internal GHS classification database has field names such as comments or the classifications themselves but this is not readily associated with a specific endpoint – can some flexibility be added to accommodate such additions
- Capture/export/printing of database import errors – is there a log file that can be accessed to help resolve such errors

Database import and management

- If a db needs to be mapped as it is imported, is there a means to save this as a template to facilitate future updates or be used/modified for subsequent similar databases?

Concluding remarks

- Lots of positives with the Toolbox and it has become an integral tool as part of any predictive tox/read-across query
- Our wish list for the short-term:
  - Mapping between IU5 and the Toolbox to facilitate use of our REACH data
  - Saving sessions
  - Exporting the data matrix