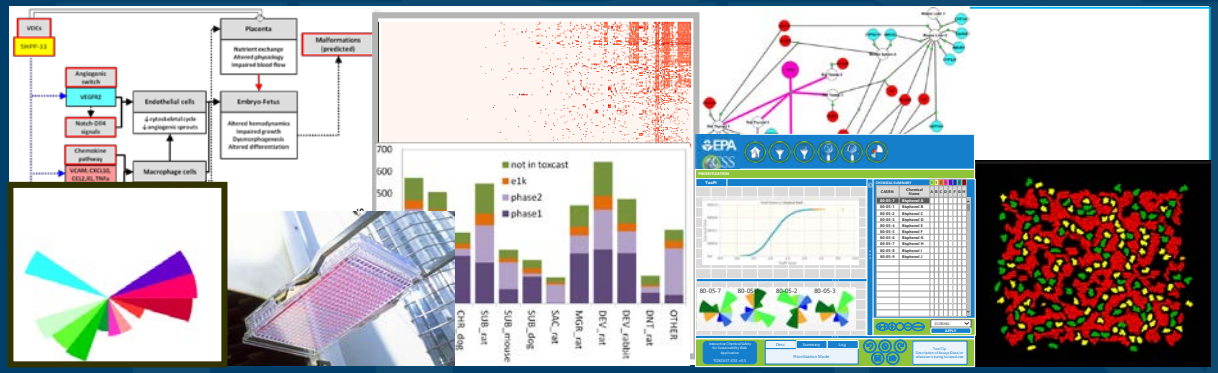
EPA United States Environmental Protection Agency

Use of chemoinformatics tools- nuts and bolts Challenges in their regulatory application



Informing and supporting mechanisms for cosmetic chemical space – Driving clarity on needs for new approaches to safety assessment
Brussels, June 2-3, 2015

Grace Patlewicz
National Center for Computational Toxicology

The views expressed in this presentation are those of the author and do not necessarily reflect the views or policies of the U.S. EPA

EPA United States Environmental Protection Agency

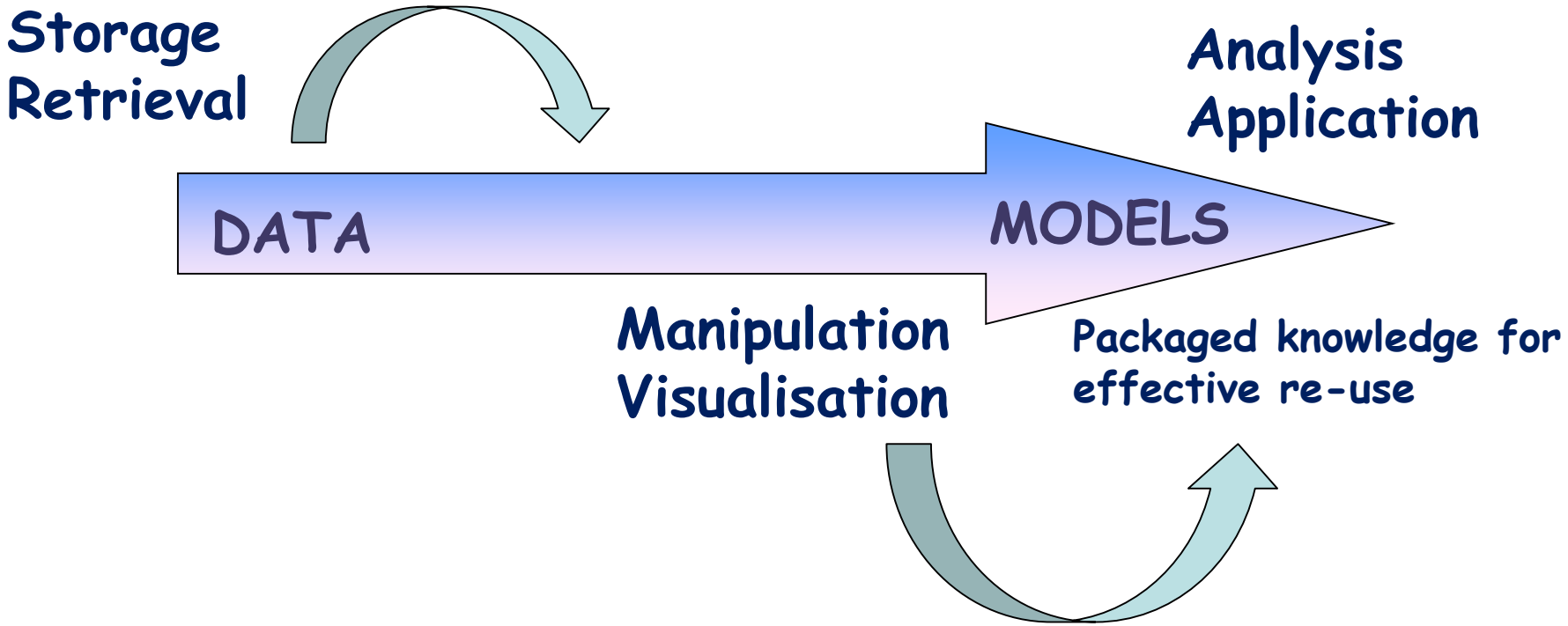
Outline

- What is Chem(o)informatics
- Decision contexts and Applications
- Screening level hazard identification and how this impacts current and emerging (Q)SAR development and application
- Chemical categories and associated read-across
- Issues with read-across
- Practical strategies to refine and enhance existing read-across approaches
- Take home messages

What is Chemoinformatics?

- Chemoinformatics or Cheminformatics?
- "..the mixing of those information resources to transform data into information and information into knowledge for the intended purpose of making better decisions faster in the area of drug lead identification…" Brown (1998)
- .."combining the scientific working fields of chemistry, computer science and information science…."

Cheminformatics – a continuum from data to knowledge



Cheminformatics tools add value to most regulatory decisions

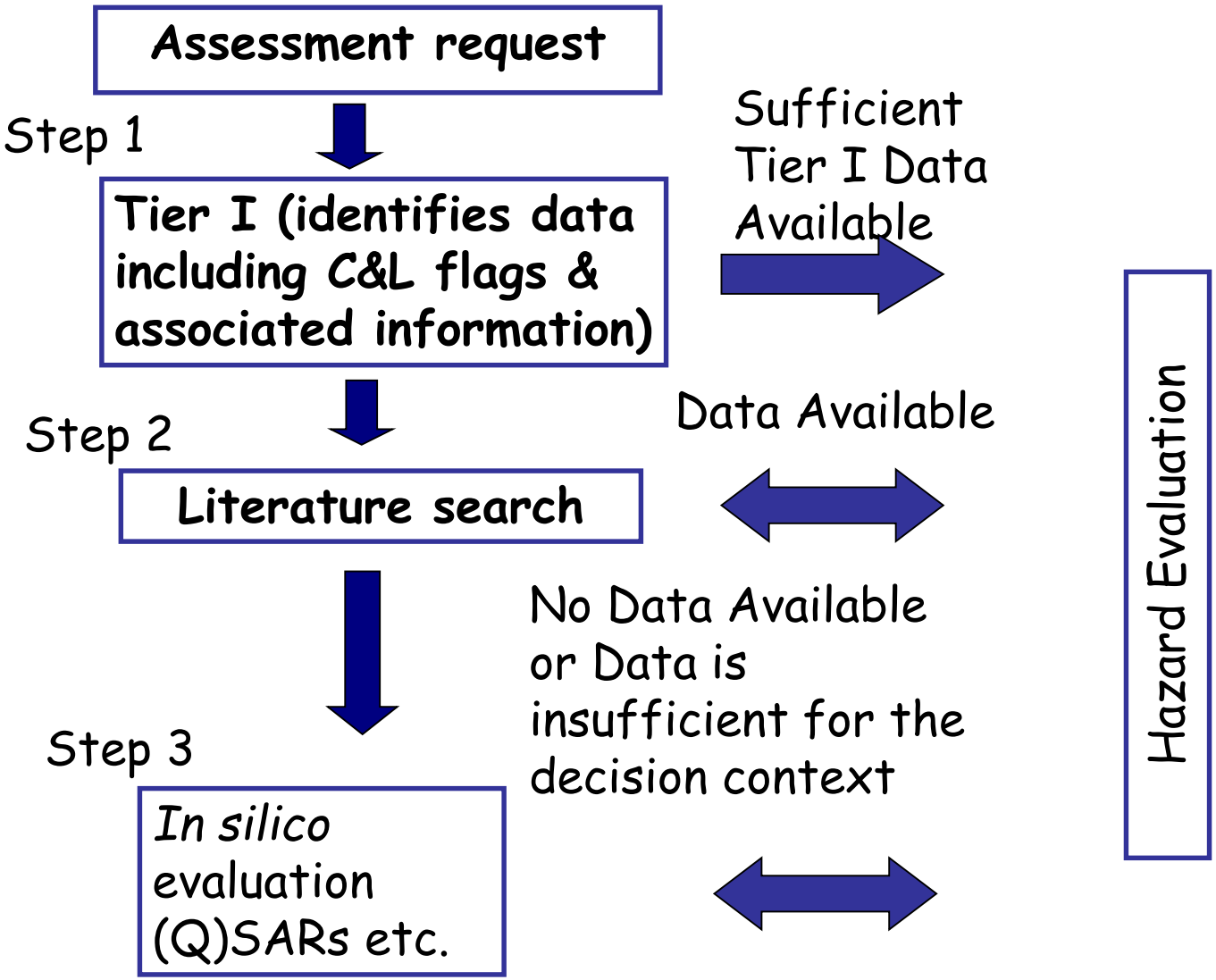
- Screening level hazard assessment
- Category formation for read-across
- Prioritisation
- Risk Assessment
- Exposure Assessment

Applications where Cheminformatics tools add value

- Screening level hazard assessment
- Cat
- Pric
- Risk
- Exposure Assessment
- ……

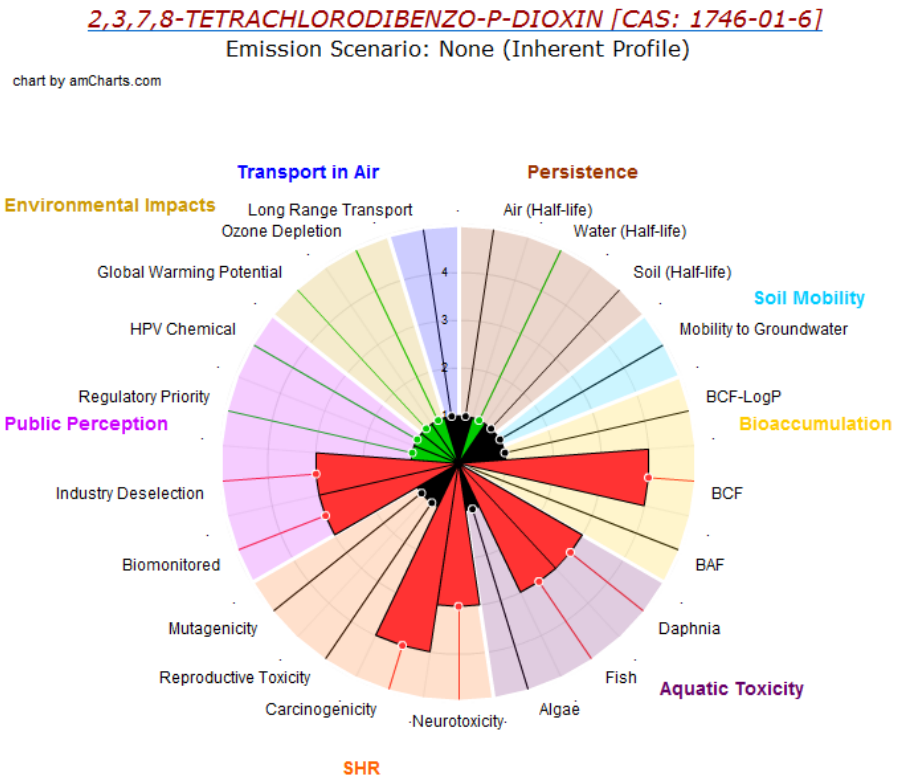
A Data gap analysis is typically the first step

Data gap analysis



Data gap analysis

Step 1: Tier I – Preliminary data search



Flags C&L information from EU, NZ, etc, Public perception lists (re: Green chemistry type considerations..)

From M Chen, DuPont

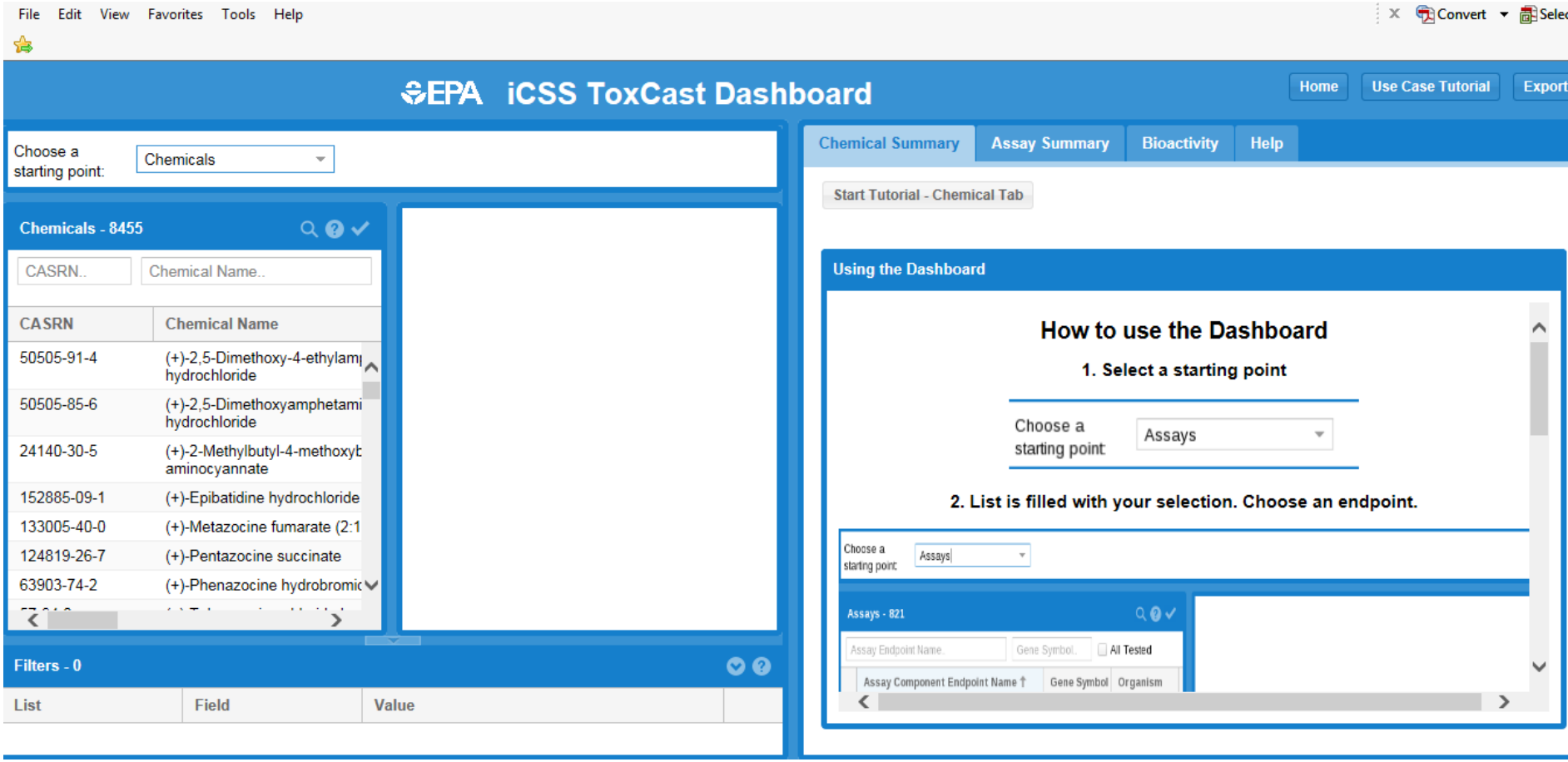
Data gap analysis

Step 2: Tier II – More extensive data search (typically traditional toxicity information)

- ACToR -http://actor.epa.gov/
- ECHA dissemination database
- eChemPortal http://www.echemportal.org/
- Scifinder
- OECD Toolbox
- Leadscope – www.leadscope.com

Data gap analysis

Step 2: In vitro - Bioactivity data

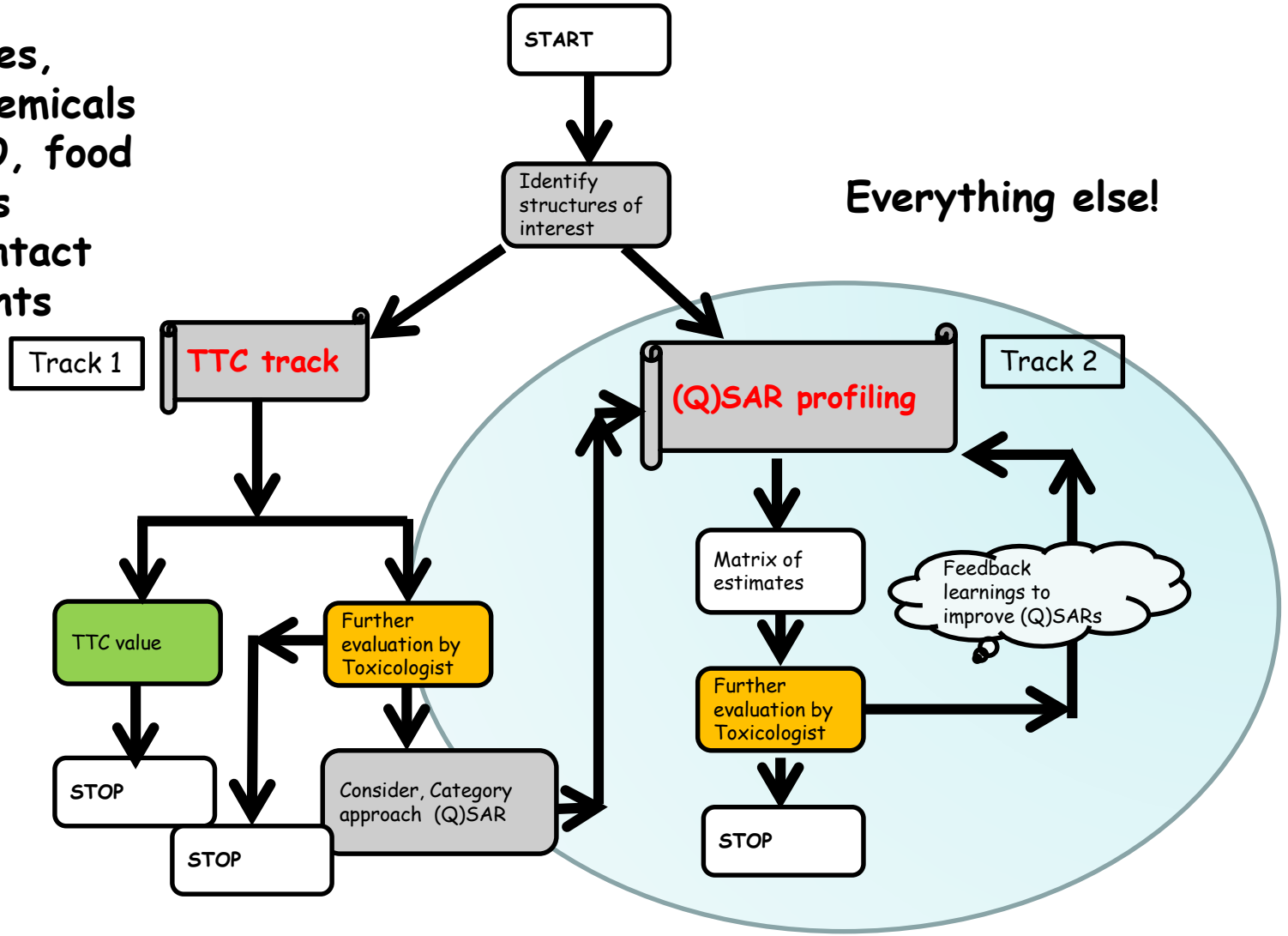


Data gap analysis

Step 3 – in silico evaluation

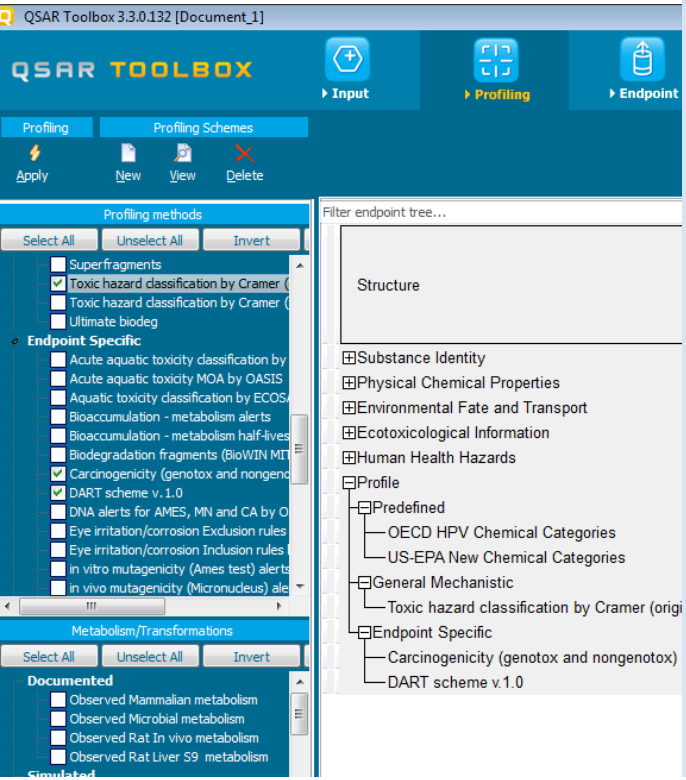
- Three approaches – depending on decision context and outcome of data gap analysis
- TTC approach if exposure is very low and is supported by use case
- or
- (Q)SARs to identify the likely endpoints of concern to help select more relevant analogues to address those endpoint data gaps
- and/or use
- Investigate an analogue/category approach

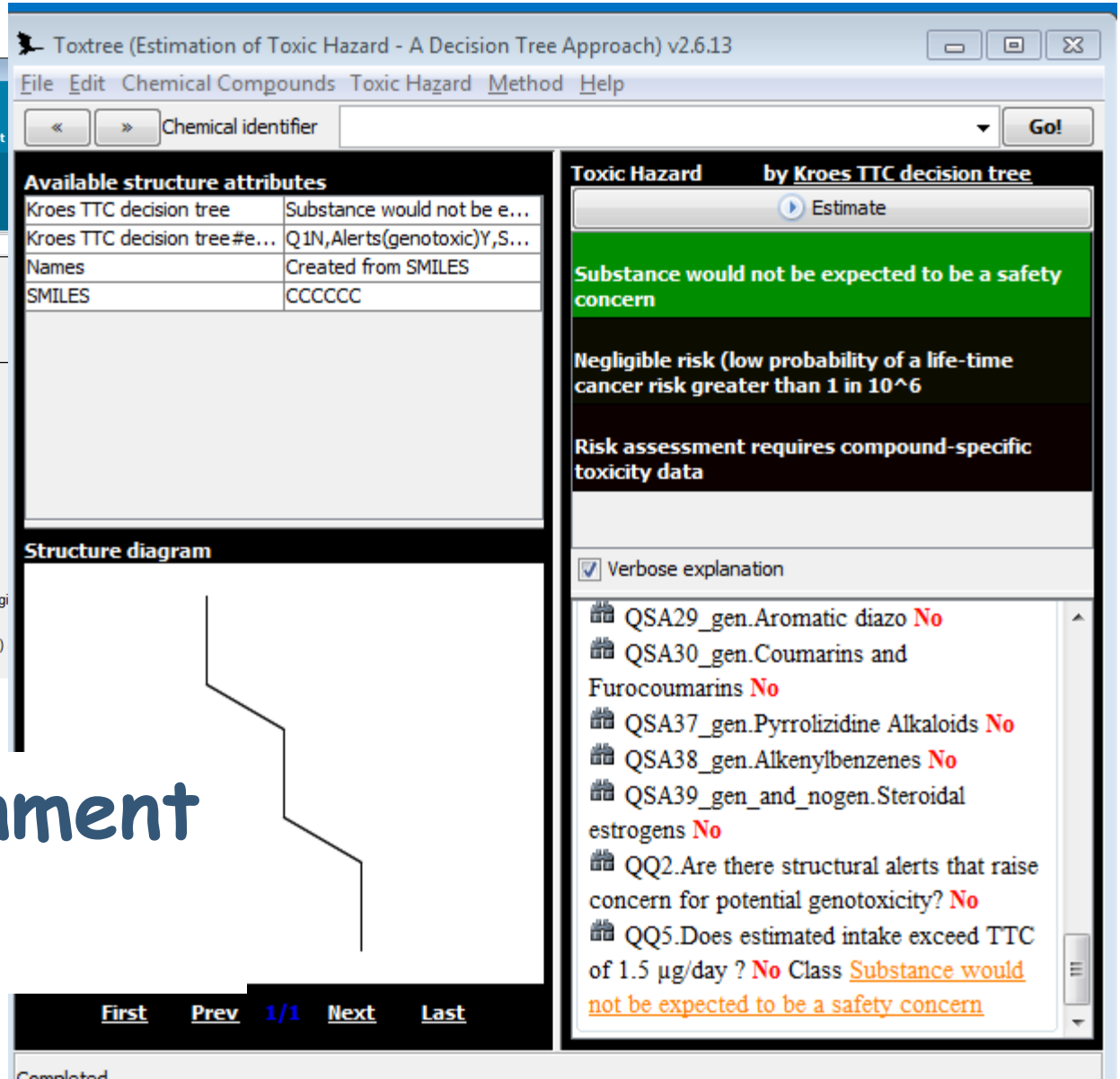
Impurities, novel chemicals for CMO, food additives Food contact ingredients



TTC via Toxtree

Cramer assignment via OECD TB

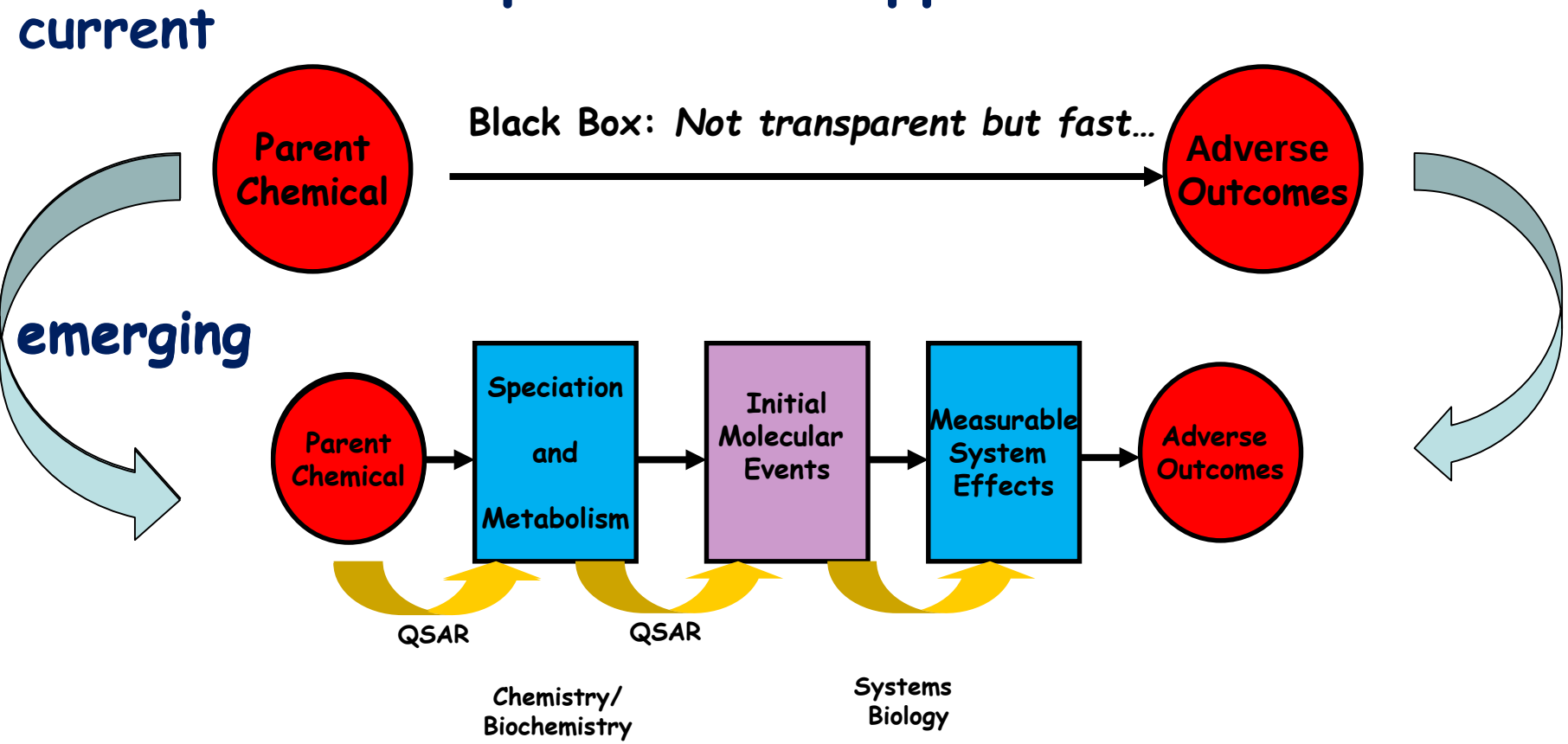




(Q)SAR profiling

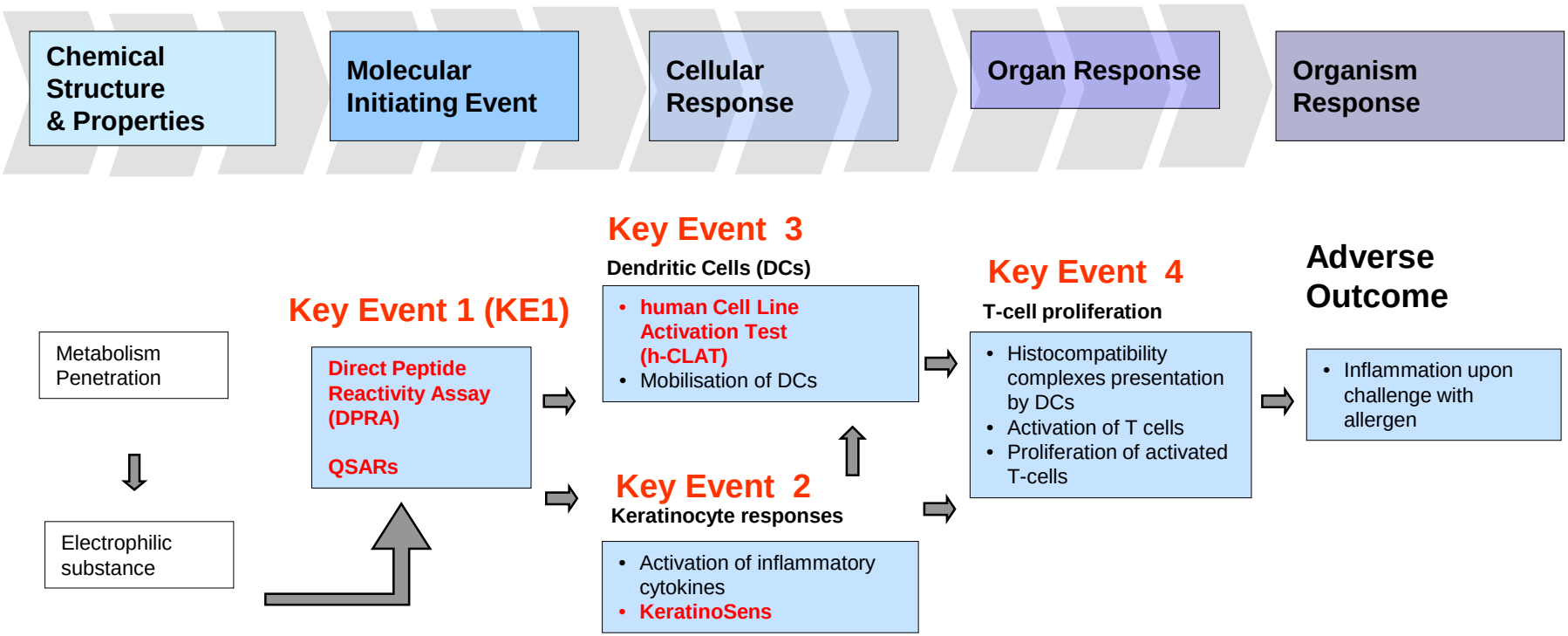
- (Q)SARs structured as "IATA" Pipelines informed by mechanistic understanding where feasible
  - Endpoint specific
    - e.g. Skin sensitisation informed by AOP
- or
- Various IATA coupled together to address several endpoints concurrently..
- Extend the approach to extract new SAR insights from bioactivity data

Conceptual approach for non-testing development and application



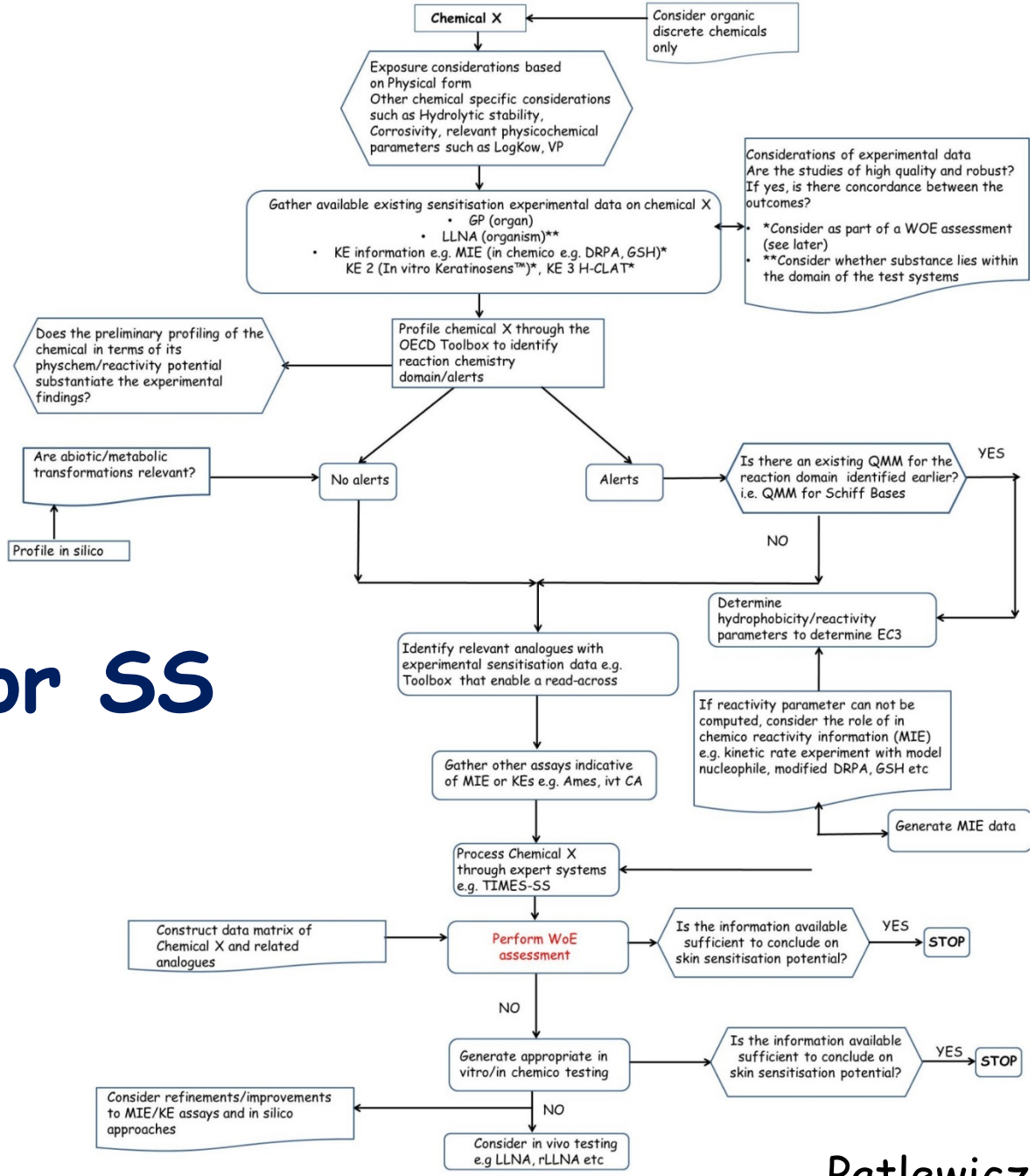
1. Identify plausible MIEs
2. Explore Linkages in Pathways to Downstream Effects
3. Develop QSARs to predict MIEs from Structure or characterise other KEs as SARs

AOP for skin sensitisation (SS) and assays mapped to KEs



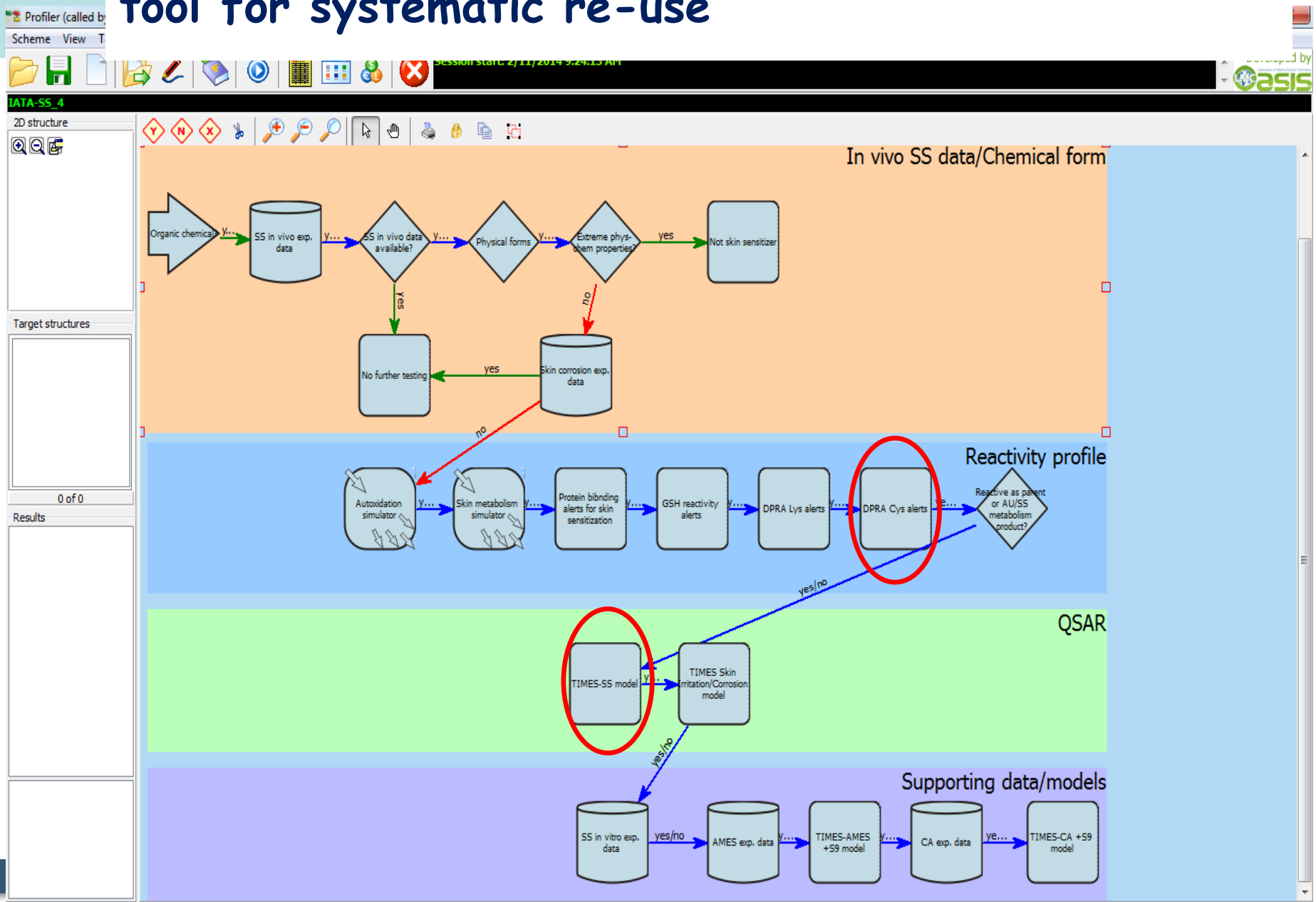
United States Environmental Protection Agency

IATA for SS



Patlewicz et al, 2014

Implementing the IATA-SS into a OASIS Pipeline tool for systematic re-use



Mechanistic basis – SAR Profiler for cysteine depletion

DPRA Cysteine peptide depletion (General Mechanistic) - Profiling Scheme Browser

Advanced

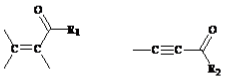
DPRA Cysteine peptide depletion - Category definitions

- DPRA Cysteine peptide depletion
  - High reactive
    - Activated haloarenes
    - alpha,beta-carbonyl compounds with polarize
    - Aromatic C-Nitroso compounds
    - Benzyl halides
    - Cyclopropenones
    - Diacylperoxides
    - Di-methacrylic acid esters
    - Halogenated isothiazolones
    - Halo-substituted dinitriles
    - Isothiazolinone derivatives
    - Organic disulfides
    - Quinones and quinone (di)imines
    - Thiols
    - Unsaturated acid anhydrides
    - Vinyl pyridines
  - Low reactive
    - Acyl halides
    - Alicyclic ketones
    - alpha-alkyl cinnamaldehyde derivatives
    - Long-chain aliphatic aldehydes
    - N-substituted aromatic amides
    - Primary haloalkanes
    - Saturated acid anhydrides
    - Special lactones
    - Sulfanilic acid derivatives
  - Moderate reactive
    - 1,2-Dicarbonyl compounds
    - Activated 1,3,5-triazine derivatives
    - Azalactones
    - Five-membered heterocyclic urea
    - Glycidyl ether epoxides
    - Mono-methacrylic acid esters
    - Phenyl substituted cinnamaldehydes
    - Saturated aldehydes

Profile Description

Potency category: High reactive

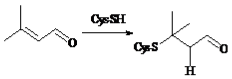
Chemical class: alpha,beta-carbonyl compounds with polarized multiple bond – **Under development**



R_1 = H, C{sp^2}, C{sp^3} R_2 = -OC, C, S, N

The possible mechanism of interaction of this structural alert with SH-group of Cysteine peptide is illustrated below:

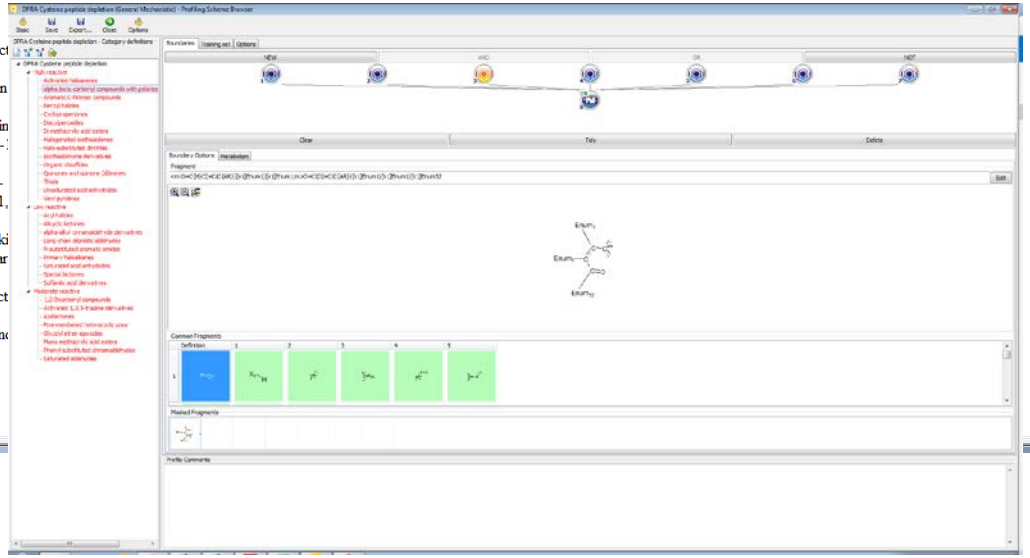
Scheme 1:



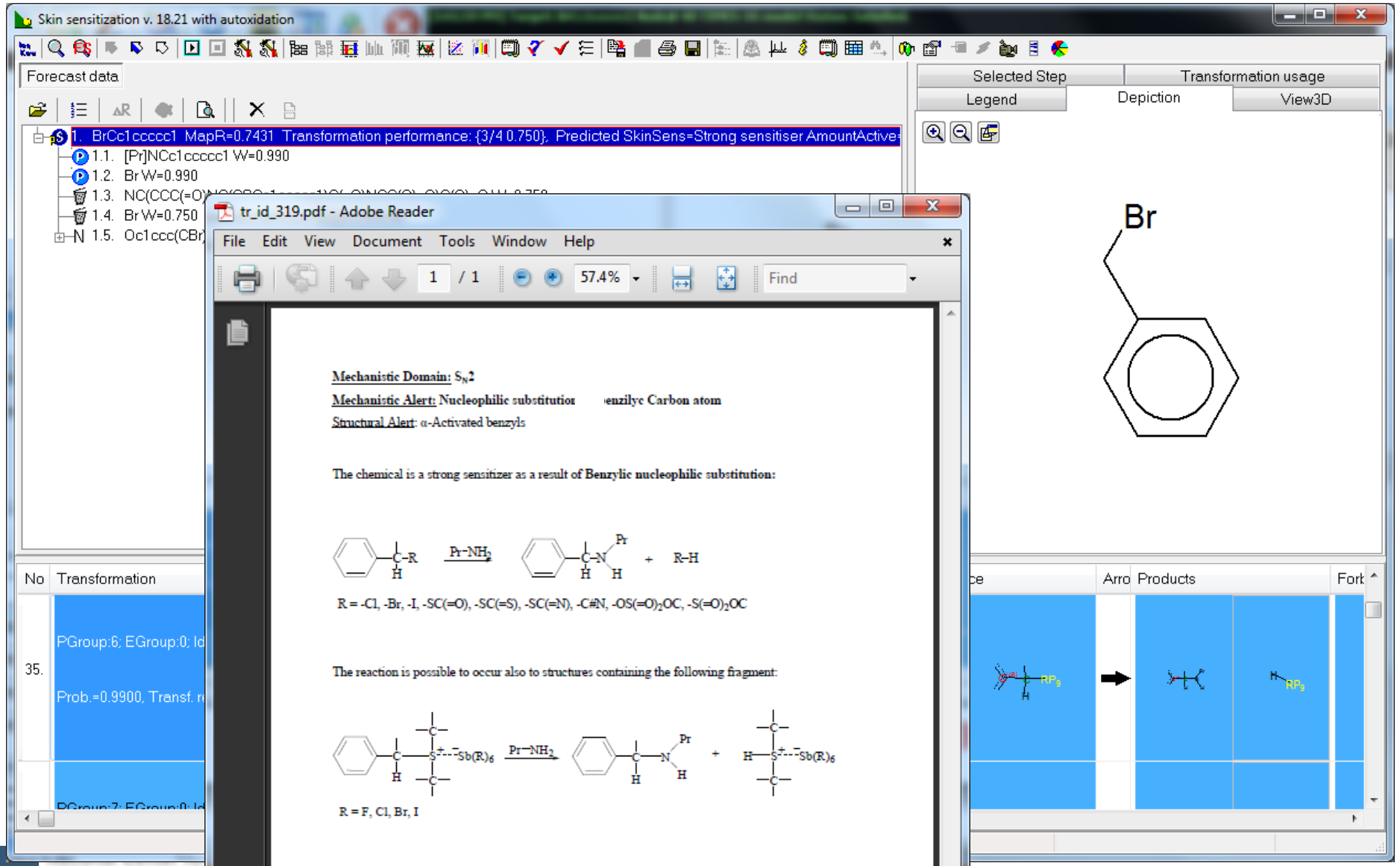
Michael acceptors are double or triple bonds with electron-withdrawing substituent such as carbonyl group. Michael-type addition provides a means of covalent adduct formation at an electrophilic center, without any leaving group. Direct addition of a nucleophile can take place across a double or triple carbon-carbon bond if it is attached to a highly polarized substituent that permits the resultant negatively charged transition state to be stabilized.
Compounds with double or triple bounds adjacent to a C=O group (in this case aldehyde-carbonyl group) are known as α,β-unsaturated carbonyl compounds. Nucleophiles will undergo conjugate additions with them. CHO-group profoundly affects the reactivity of the double bond. Proteins are good nucleophiles for conjugate addition reactions with these compounds.
Any α,β- unsaturated aldehydes can potentially act by the Schiff base mechanism.

References:

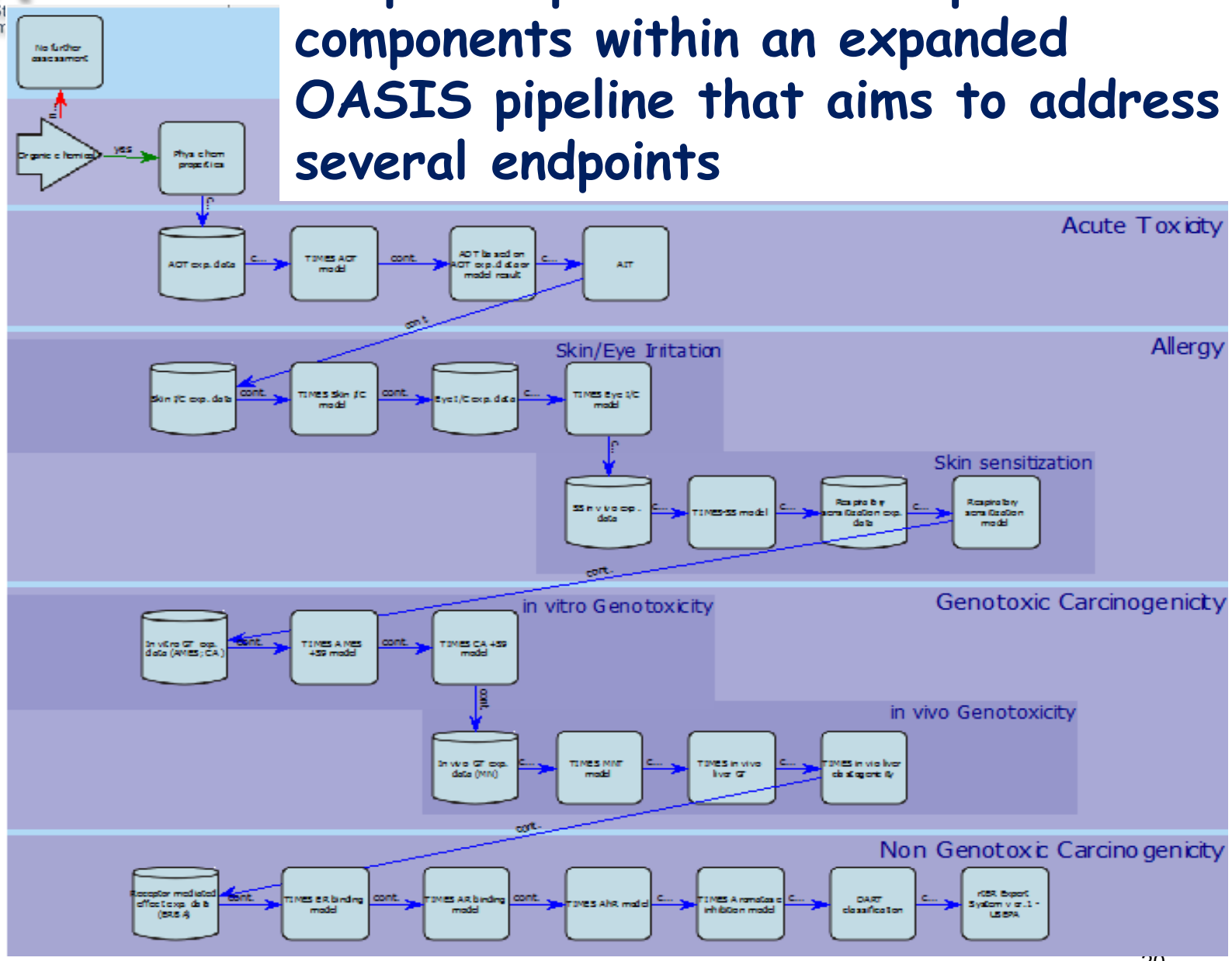
1. Aptula A, Roberts D. Mechanistic applicability domains for nonanimal-based predict
2. Roberts D.W., Patlewicz G., Kern P., Gerberick F., Kimber I., Dearman R.J., Ryan
3. Ashby J., Basketter D.A., Paton D., Kimber I., Structure activity relationships in skin sensitization using the murine local lymph node assay, Toxicology, 1995, 103, 177-
4. Patlewicz G., Basketter D.A., Smith C.K., Hotchkiss S.A.M., Roberts D.W., Skin-sensitisation structure-activity relationships for aldehydes, Contact Dermatitis, 2001,
5. Roberts D.W., Patlewicz G., Mechanism based structure-activity relationships for ski sensitisation -–the carbonyl group domain. SAR and QSAR in Enviromental Resear
6. Camilla K. Smith Pease, From xenobiotic chemistry and metabolism to better predict
7. Camilla K. Smith, Sharon A.M. Hotchkiss, Allergic Contact Dermatitis: Chemical an



TIMES-SS predictions



Endpoint specific IATA represent components within an expanded OASIS pipeline that aims to address several endpoints



Chemical category and read-across: Workflow

- Data gap analysis
- Overarching hypothesis
- Analogue identification
- Analogue evaluation
- Data gap filling
- Scientific justification

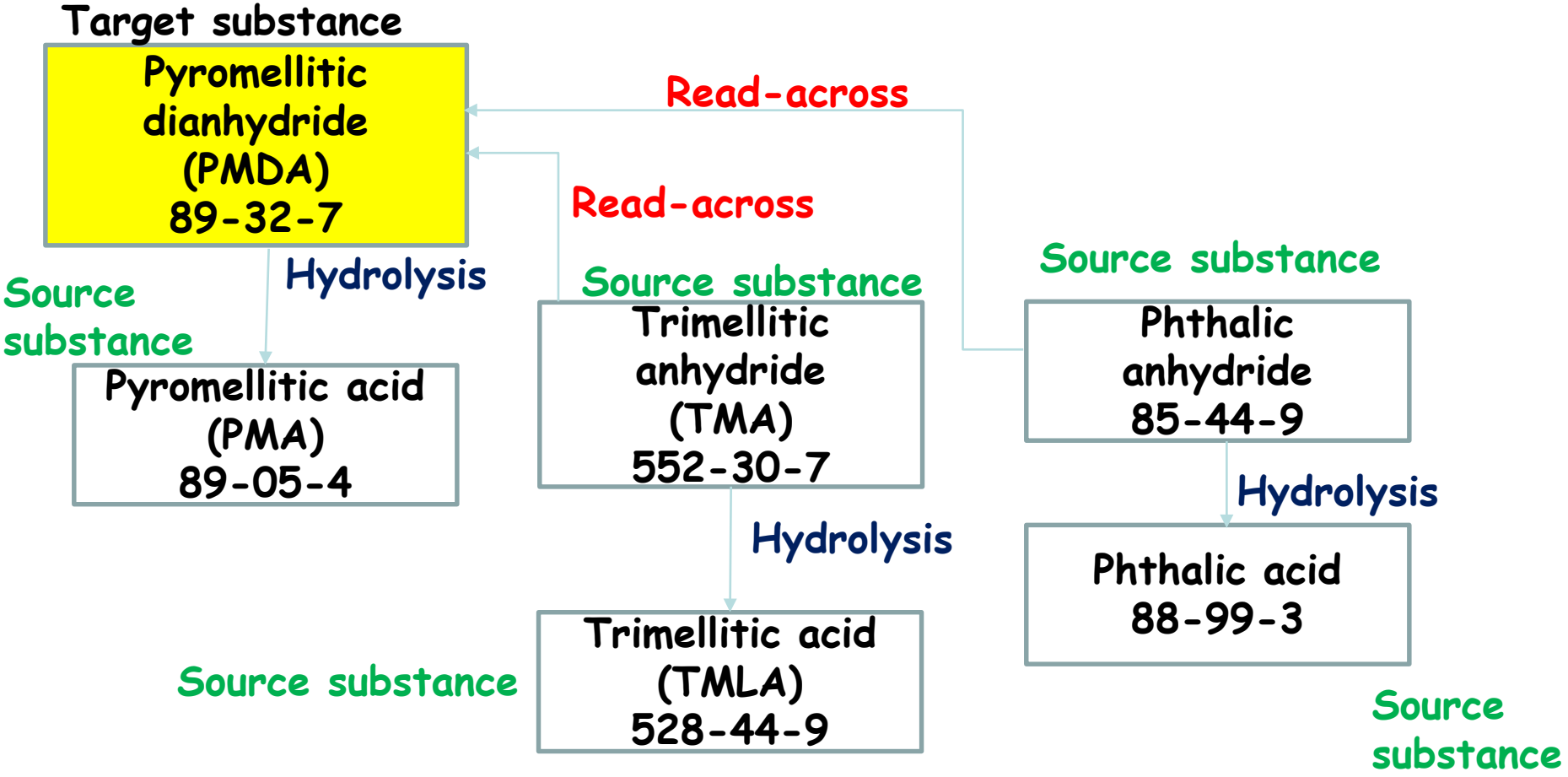
Overarching hypotheses - "similarity rationales"

- 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

  - The rationale underpinning the category/analogue approach might be based on 1 or more of these rationales

NB: Rationales extracted from the OECD & REACH Technical guidance

Overarching category rationale: the likelihood of common precursors and/or breakdown products, via physical or biological processes, which result in structurally similar chemicals



Analogue Identification - tools

- ChemID plus – structure searching for similar analogues with/without data
- eChemPortal – CAS, Name
- Leadscope – CAS, Structure (similar/exact)
- OECD Toolbox – structure, profilers..
- AMBIT v2- http://cefic-lri.org/lri_toolbox/ambit/
- ACToR – through DssTox

Search may be uninformed or informed by an overarching rationale

Analogue Evaluation

- Evaluating on the basis of reaction chemistry and mechanistic knowledge..to substantiate a proposed hypothesis or to establish a rationale for the grouping proposed

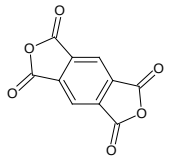
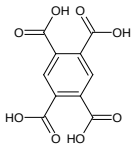
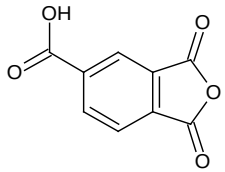
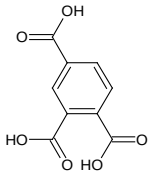
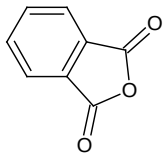
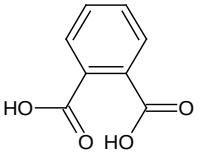
- OECD Toolbox
- Leadscope
- Toxtree
- Derek Nexus
- Meteor
- ..

Data gap filling approaches & tools

- Read-across can be:
  - Qualitative read-across
  - Quantitative read-across
  - Trend-analysis
  - External QSARs

- Tools may include:
- OECD Toolbox
- Toxmatch
- AMBIT
- Qualitative inferences using the data matrix directly

Data gap filling approaches: Data matrix

Name	Pyromellitic dianhydride	Pyromellitic acid	Trimellitic anhydride	Trimellitic acid	Phthalic anhydride	Phthalic acid
Role in category	Target	Source	Source	Source	Source	Source
Abbrev	PMDA	PMA	TMA	TMLA		
Cas	89-32-7	89-05-4	552-30-7	528-44-9	85-44-9	88-99-3
Structure						
Physicochemical properties	X	X	X	X	X	ND
Toxicological endpoints e.g. acute oral toxicity	Read-across	X	X	ND	X	ND
Ecotoxicological endpoints	X	ND	X	X	ND	ND
Environmental fate properties	Read-across	ND	ND	X	ND	X

Key challenges associated with read-across

- 'Negative read-across' – reading across the 'absence' of toxicity – burden of proof is higher
- → what is the mechanism of action for the absence of toxicity…
- How to estimate uncertainty?
- Not possible to remove uncertainty entirely
- how much residual uncertainty is acceptable?
- or what type of uncertainty is acceptable and does this differ for different endpoints and for different decision contexts?
- Can Uncertainty be addressed without (additional) *in vivo* testing?
- Read-across remains a subjective expert judgement assessment

How to address the challenges with read-across?

- Several public
  - constru
  - and
    - G
      - E
    - Fr
      - e
      - 2013
    - Frame
      - (ECH
      - 2014,

Read-across Assessment Framework

The ECHA Read-Across Assessment Framework (RAAF) structures the scientific

Regulatory Toxicology and Pharmacology xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

homepage: www.elsevier.com/locate/yrtph

SCIRADE

Building scientific confidence in the development and evaluation of read-across

G. Patlewicz [a,*], N. Ball [b], P.J. Boogaard [c], R.A. Becker [d], B. Hubesch [e,f]

[a] DuPont Haskell Global Centers, 1090 Elkton Road, Newark, DE 19711, USA
[b] Toxicology & Environmental Research and Consulting (TERC), The Dow Chemical Company, Midland, MI 48674, USA
[c] Shell Health, Shell International b.v., PO Box 162, 2501 AN The Hague, The Netherlands
[d] Regulatory and Technical Affairs Department, American Chemistry Council (ACC), Washington, DC 20002, USA
[e] CEFIC – The European Chemical Industry Council, 4 Avenue E. Van Nieuwenhuyse, B-1160 Brussels, Belgium
[f] Hubesch Consult BVBA, Madelieffeslaan 10, B-1600 Sint-Pieters-Leeuw, Belgium

ARTICLE INFO

Article history:
Received 18 March 2015
Available online xxxx

Keywords:
Read-across
(Quantitative) Structure Activity Relationship (Q)SAR
Quantitative Mechanistic Models (QMM)
Chemical categories
Analogue approach
Adverse Outcome Pathway (AOP)
Scientific confidence

ABSTRACT

Read-across is an alternative approach exploited to address information requirements for risk assessment and for regulatory programmes such as the European Union's REACH regulation. Whilst read-across approaches are accepted in principle, difficulties still remain in applying them consistently in practice. Recent work within Cefic LRI and ECETOC attempted to summarize the state-of-the-art and identify some of the barriers to broader acceptance of read-across approaches to overcome these. Acceptance is undoubtedly thwarted partly by the lack of a systematic framework to characterize the read-across justification and identify the uncertainties particularly for complex regulatory endpoints such as repeated-dose toxicity or prenatal developmental toxicity. Efforts are underway by the European Chemical's Agency (ECHA) to develop a Read-Across Assessment Framework (RAAF) and private sector experts have also considered the development of a similar framework. At the same time, mechanistic chemical categories are being proposed which are underpinned by Adverse Outcome Pathways (AOPs). Currently such frameworks are only focusing on discrete organic substances, though the AOP approach could conceivably be applied to evaluate more complex substances such as mixtures. Here we summarize the deliberations of the Cefic LRI read-across team in characterizing scientific confidence in the development and evaluation of read-across.

© 2015 Elsevier Inc. All rights reserved.

explanations and examples. For each assessment element, an assessor is guided through a series of questions to select the most appropriate assessment option (conclusion) for that element.

How to address the challenges with read-across?

- Reliance on prior knowledge and expertise regarding structure and function
  - Does not work as well with data poor substances
- No clear guidance on what to do to decrease uncertainty – what studies, data, etc.
- Although activities/projects are ongoing: LERAT, SEURAT, CAAT, AIMT-4
- → to target Uncertainty

SCIRADE proposals

Table 2
Scientific confidence considerations in Read-across evaluation.

Data issues	Similarity rationale
Analogue/category approach	Similarity rationale/hypothesis that underpins the analogue/category approach – Metabolic transformation – Structural similarity
Completeness of data matrix – No of data gaps e.g. source analogue(s) have many data points to address, target substance has a handful of data gaps.	Analogue validity – Analogue similarity with respect to general and endpoint specific considerations – Rationalization of why structural differences do not impact the toxicity
Quality of data for source analogues – e.g. Klimisch scores of 1 or 2	Concordance of effects and potency (if relevant) per endpoint • Presence or absence of adverse effects • Type of read-across (Qualitative, Quantitative, Trend Analysis) Concordance of effects and potency (if relevant) across endpoints

Patlewicz et al, 2015 (SCIRADE)

SCIRADE proposals

United States
l Protectior

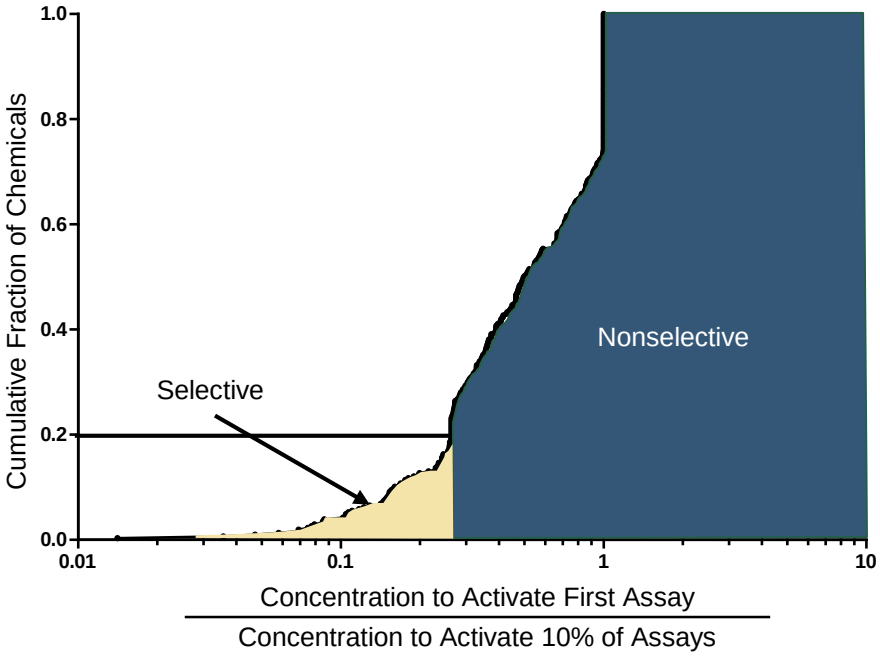
Table 3
Practical strategies for addressing uncertainties.

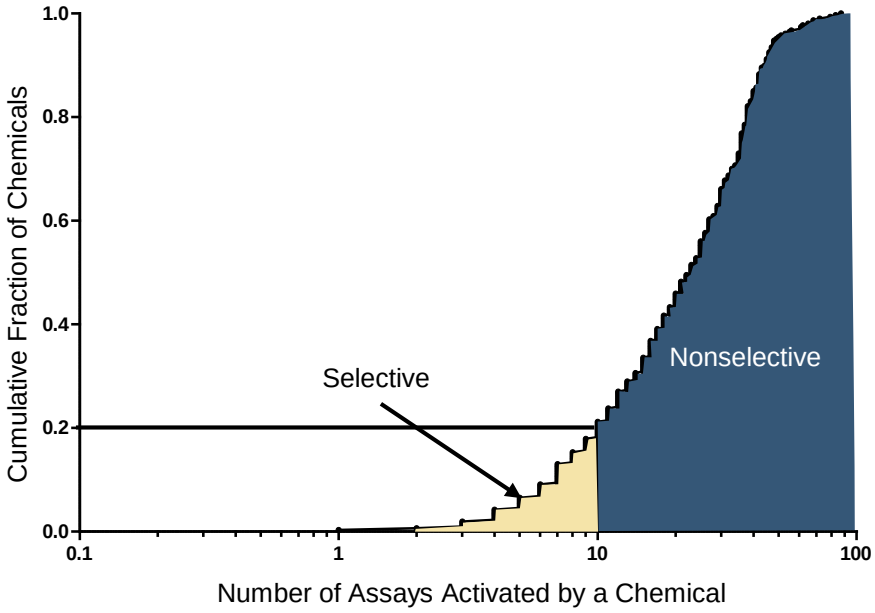
Similarity rationale element	Strategies to address uncertainty	Examples
Metabolic transformation	• Predict likely metabolite using *in silico* tools • Assessing metabolism through one or another systems. E.g. precision-cut tissue slices, subcellular fractions such as the microsomal fraction, primary cells in suspension, primary monolayers of cells in culture, continuous cell lines, immortalized primary cells, liver-derived cell	• e.g. OECD Toolbox contains simulators of skin and liver metabolism • e.g. Use of the rat/human *in vitro* hepatocyte assay to substantiate transformation hypothesis in terms of identity of metabolite(s) formed, and kinetics of transformation
		assays associated with key events within the AOP for skin sensitization
	• Are there specific *in vitro* assays that are associated with a mechanistic pathway not necessarily affiliated with a defined AOP?	• e.g. Neurotransmitter inactivation mechanisms – assays for acetylcholinesterase inhibitors; assays to measure mitochondrial dysfunction
	• Are there specific target organ effects that can be linked	• e.g. Available 28 day study in target and source ana-
Unspecific toxicity	• Is there concordance in a broad spectrum of cell based phenotypic assays? • Is there concordance in high content (HC) imaging profiles?	• e.g. HTS assays such as those within the EPA ToxCast™ programme • e.g. HC to evaluate cell death, apoptosis, oxidative stress, mitochondrial membrane potential, DNA damage, cell cycle inhibition etc
	• Is there concordance in gene expression signatures? • Is there a role for organotypic systems • Is there utility in the use of non mammalian systems such as in insects, nematodes, and zebrafish models?	• e.g. L1000 • e.g. liver, skin, lung
Evaluating whether structural differences of the source analogue may impact the toxicity relative to the target substance	• Are there specific structural alerts identified for the structural features that are not common between the target and source analogues? • Do the structural differences translate to significant differences to the reactivity profile between source and target analogue that could result in differences in toxicity? • Do the structural differences translate to significant differences to the metabolic pathway between source and target analogue that could result in differences in toxicity? • Do the structural differences result in significant differences to the physicochemical properties that could impart differences in bioavailability?	• e.g. Use of systems such as the OECD Toolbox, Derek Nexus can be helpful in identifying specific structural alerts • e.g. The same systems may be helpful in rationalizing where structural differences may translate into differences in reactivity profile – an activated carbonyl vs a stable carbonyl • e.g. Use of the OECD Toolbox's metabolic simulators or Meteor Nexus may prove helpful in exploring the metabolic pathways and their differences • e.g. Estimation of log Kow and MW can provide useful insights into potential differences in bioavailability

Using AOPs to provide the roadmap for mechanistic information

Using bioactivity information

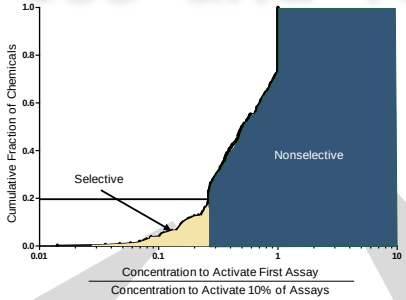
Most Chemicals are Promiscuous





Thomas et al, 2013

Using Selectivity to address a practical strategy in enhancing read-across and to define PODs



Selective Chemical

Selectively Activated *In Vitro* Assays

Define Mode-of-Action

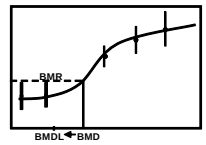
Confirm Human Relevance and Derive Point-of-Departure

Nonselective Chemical

Bioactivity

Define Point-of-Departure

Key Events



Thomas et al, 2013


OECD
EPAA
OECD IATA for Skin sensitisation DG
ECETOC TF
WNT
TFHA
ITS
TR116
Manuscript
Chemical categorization and read-across
TG development
EAGMST
OECD Toolbox
Cefic LRI LERAT
AOP Handbook
AOP-KB
IATA Case studies
Grouping guidance v2
Cefic LRI AIMT-4
AOP-Wiki
SCiRADE
Effectopedia
ECHA
OECD non-guidance in vitro test methods
ACC Comp Profiling & Risk Assessment WG
IATA Roadmap Workshop
CAAT
Uncertainty in read-across
Manuscript
PoToMAC
CAAT Read-across
RAAF
Scientific Confidence Framework (SCF)
RAAF Workshop October 2014
Food for Thought
AhR AOP
Assays
Prediction Models
REACH experiences
Negative read-across
Addressing Uncertainty
BIAC AOP WG
AOPs
Local validity
Mechanistic information from biological activity e.g. ToxCast
P&G Uncertainty Framework
NCCT
NICEATM Bioactivity-Based Read-Across (BaBRA)
AOPs
Enhancing read-across in current OECD HPV categories
Tropsha CBRA
Human Toxicology Consortium
(Dow) Low toxicity e.g. propylene glycols
SEURAT-1 (Read-across)
Framework for Intelligent Non-Animal Alternative Methods for Safety Assessment
Known MOA
Case studies
Guidance for read-across
Risk21
Low toxicity e.g. sat alcohols
ILSI
W meta act - beta-unsat alcohols
Direct acting PFOA

Objectifying read-across

- Addressing uncertainty in read-across and promoting a more systematic approach to evaluating read-across performance
- Using AOPs
- Using biological activity data
    1. Local validity approach – hybrid "QSAR" nearest neighbour similarity distance to establish a baseline performance and quantify uncertainty i.e extension of CBRA approach
        - Extend and refine by codifying expert insights
    2. Explore bioactivity data as a means of enhancing existing chemical categories

CBRA (Low et al, 2013)

- Chemical Biological Read Across
- Predict RA activity of chemical as similarity-weighted activity of neighbours:

EPA United States Environmental Protection Agency

"GeneRA"

- Generalised Read Across (GeneRA)
- Predict chemical activity as similarity-weighted activity of neighbours across different descriptor spaces:

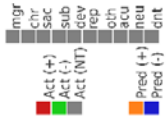
Jaccard similarity:

Shah et al, in prep

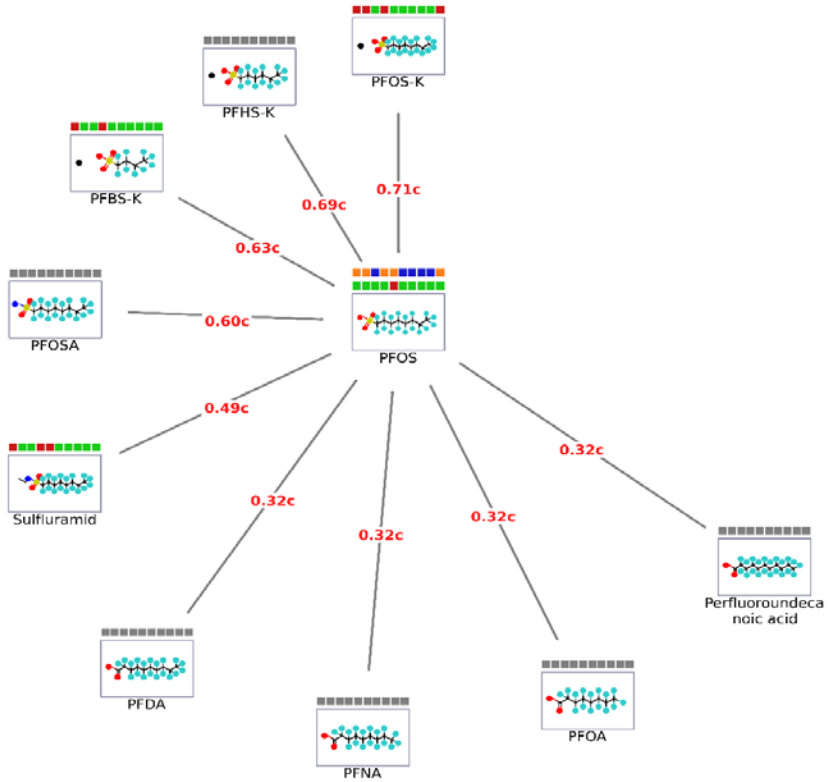
GeneRA: Clustering chemicals

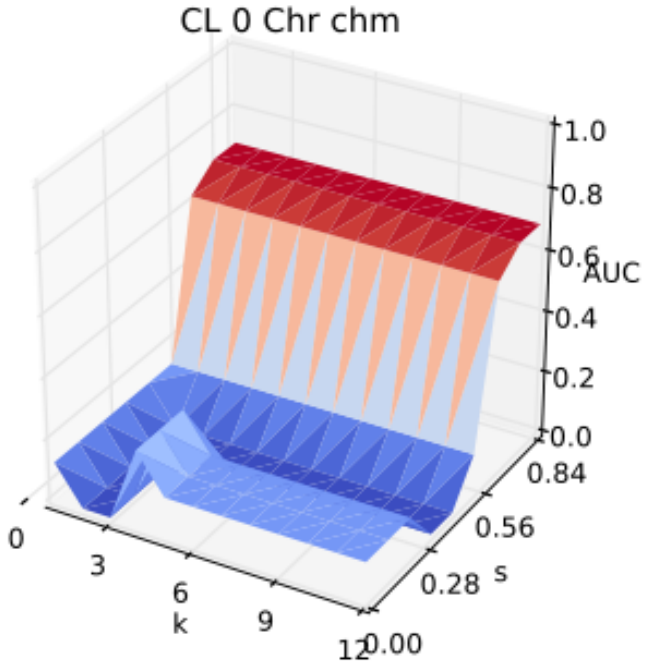
Shah et al, in prep

GeneRA: Nominal cluster



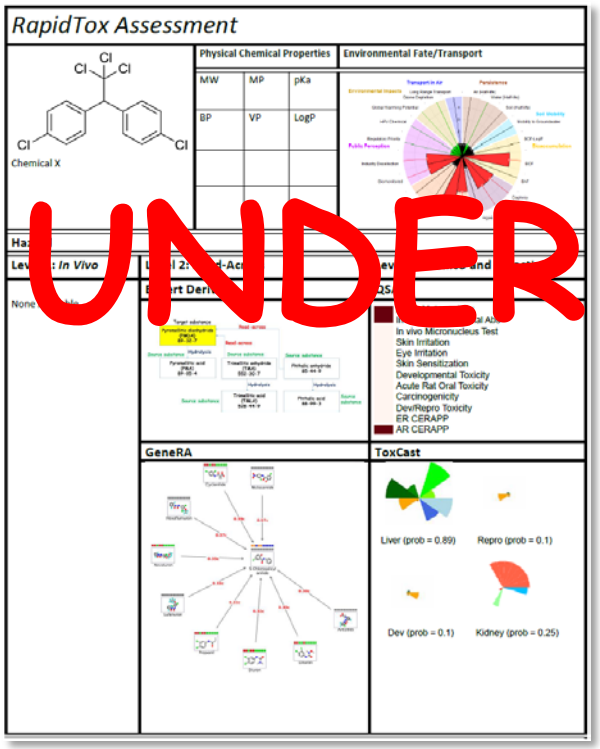
Infer AUC for chronic effects
Using chemical
/bioactivity or hybrid descriptor

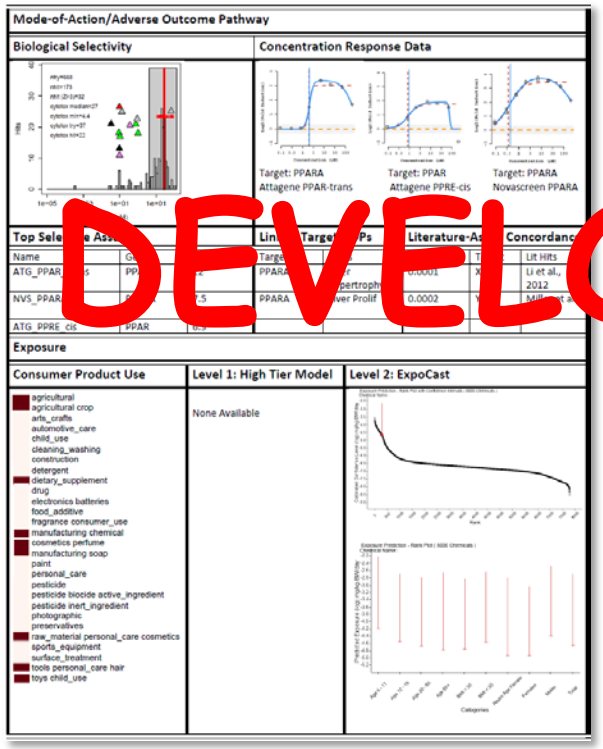


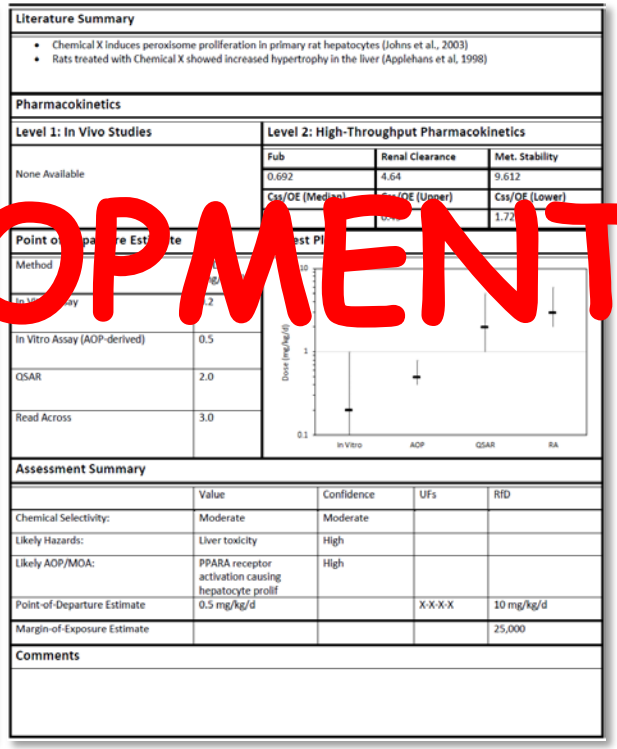


Shah et al, in prep

Prototype Implementation – within a Dashboard

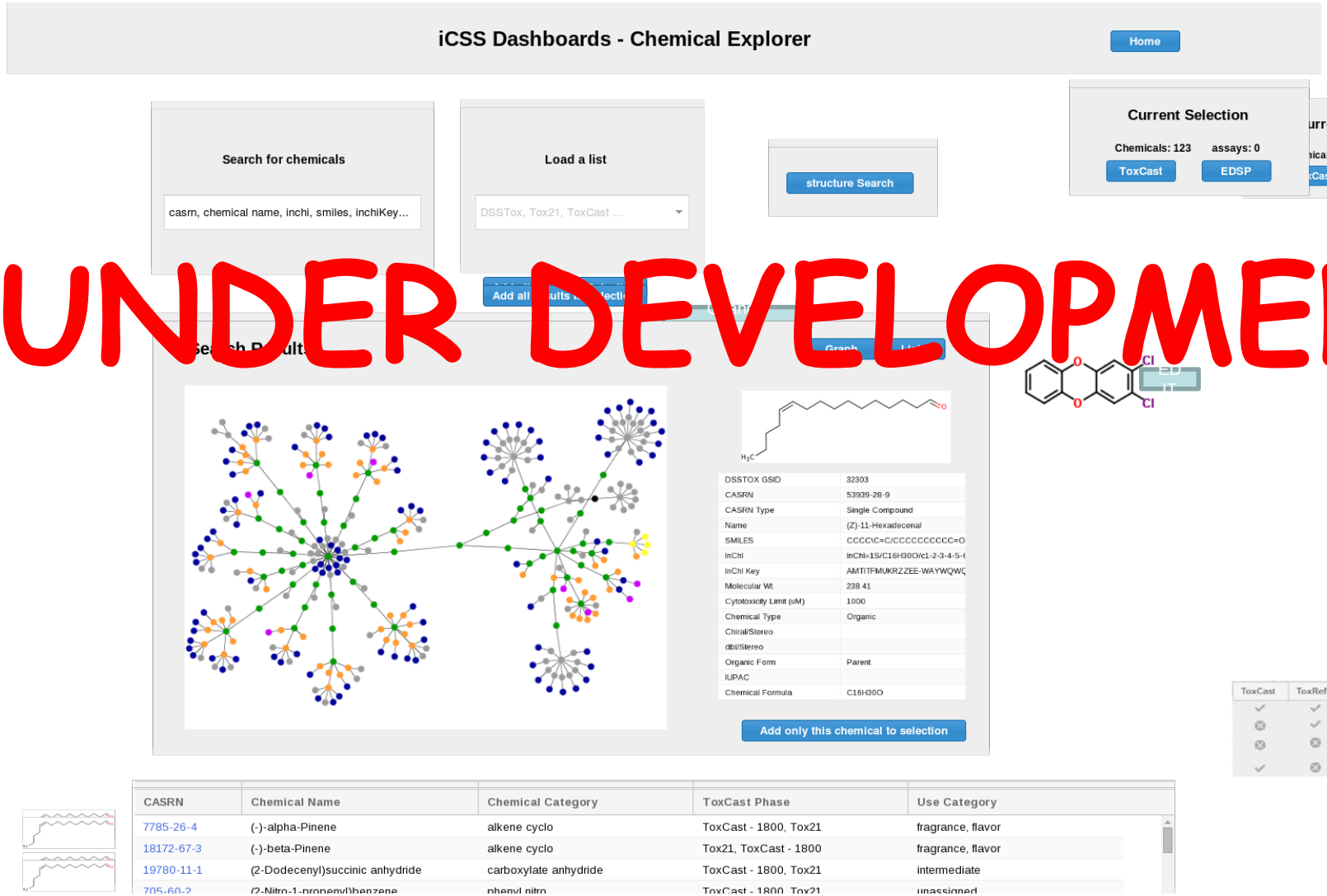






UNDER DEVELOPMENT

Prototype Implementation – within a Dashboard



EPA United States Environmental Protection Agency

Take home messages

- Scope of cheminformatics is very broad
- Focused on specific tools which facilitate screening level hazard assessments and read-across within chemical categories
- Illustrated how mechanistic information from AOPs can be helpful to derive new (Q)SARs

Take home messages

- Highlighted the issues with read-across and suggestions for how in the absence of adversity or AOPs, in vitro bioactivity data could be helpful in quantifying performance and shifting read-across away from a subjective expert driven assessment (at least for specific decision contexts)