Computational Tools and Guidance developed by the JRC

Klaus Daginnus
Institute for Health & Consumer Protection
Joint Research Centre, European Commission

CAESAR Final Workshop
Milan, 11 March 2009

http://ecb.jrc.ec.europa.eu/qsar/
1. Computational toxicology at the JRC
2. Role of computational methods in risk assessment
3. Filling data gaps: read-across and (Q)SARs
4. Documenting the results of read-across and (Q)SARs
5. Non-testing strategy – a stepwise approach
6. Conclusions
The European Commission’s Joint Research Centre

European Commission

Directorates-General

Directorates or Institutes

Units

Molecular Biology
In Vitro Methods
Nanobiosciences
Chemical Assessment & Testing
Systems Toxicology

JRC

Policy Areas

Genetically Modified Organisms
Alternative Methods & ECVAM
Nanotechnology
Health and Environment
Consumer Products & Nutrition
Overall aim: to promote the development, assessment, acceptance and implementation of computer-based methods potentially suitable for the regulatory assessment of chemicals


Main approaches: SAR, QSAR, molecular modelling, ranking

Computational methods provide information for use in hazard and risk assessment → “non-testing” or alternative methods

http://ecb.jrc.ec.europa.eu/qsar/
Risk assessment process

Exposure estimation
- Emission rates
- Environmental distribution
- Exposure levels, intakes

Data evaluation
- Data set

Hazard identification
- Toxicity data
- Extrapolation
- No-effect levels

Risk characterisation
- Exposure vs No-effect Level

Risk = Hazard \times Exposure
Role of computational methods

Exposure estimation
- Emission rates
- Environmental distribution
- Exposure levels, intakes

Data evaluation
- Data set

Hazard identification
- Dose-response assessment
- Toxicity data
- Extrapolation
- No-effect levels

Mechanistic information

Risk = Hazard \times Exposure

Threshold of No Concern

Risk characterisation
- Exposure vs No-effect Level
Information requirements are largely tonnage dependent, however …

“Information on intrinsic properties of substances may be generated by means other than tests, provided that the conditions set out in Annex XI are met” (Article 13)

(Animal) testing can be reduced or avoided by “replacing traditional test data with predictions or equivalent data”

… however a number of conditions apply
Integrated Testing Strategies (ITS)

**Endpoint-specific strategy**

- **in vitro tests**
- **Exposure information**
- **(Q)SARs**
- **read-across & chemical groups**
- **Other existing information**

**C&L, risk assessment, PBT (vPvB) assessment**

**Targeted testing**

**safe use of chemicals?**

**Risk management measures**
Filling data gaps by read-across

Known information on the property of a substance (source chemical) is used to make a prediction of the same property for another substance (target chemical) that is considered “similar”

<table>
<thead>
<tr>
<th>Property</th>
<th>Source chemical</th>
<th>Target chemical</th>
</tr>
</thead>
</table>

1,2-Benzenedicarboxylic acid, bis(2-ethoxyethyl) ester  
Known to be harmful: $1 < \log \text{LC50} < 2$

Acute fish toxicity?  
Predicted to be harmful

diethyl phthalate
The analogue approach refers to the grouping of chemicals and application of read-across for a single endpoint based on a relatively small number of analogues.

### One-to-One

<table>
<thead>
<tr>
<th>Property</th>
<th>Substance 1</th>
<th>Substance 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>⬤</td>
<td>⬤</td>
</tr>
</tbody>
</table>

- ⬤ reliable data point
- ⬤ missing data point

### Many-to-One

<table>
<thead>
<tr>
<th>Property</th>
<th>Substance 1</th>
<th>Substance 2</th>
<th>Substance 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
</tr>
</tbody>
</table>

- ⬤ reliable data point
- ⬤ missing data point
The **category approach** refers to a wider approach, based on more analogues, multiple endpoints, and in which trends are also apparent.
In order for a (Q)SAR result to be adequate for a given regulatory purpose, the following conditions must be fulfilled:

- the estimate should be generated by a valid (reliable) model

- the model should be applicable to the chemical of interest with the necessary level of reliability

- the model endpoint should be relevant for the regulatory purpose
The need for “adequate and reliable” documentation is met by using standardised reporting formats:

A (Q)SAR Model Reporting Format (QMRF) is a robust summary of a (Q)SAR model, which reports key information on the model according to the OECD validation principles.

A (Q)SAR Prediction Reporting Format (QPRF) is a description and assessment of the prediction made by given model for a given chemical.
A (Q)SAR should be associated with the following information:

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

QMRF captures information on fulfilment of OECD validation principles, but no judgement or “validity statement” is included

• Principles adopted by 37th Joint Meeting of Chemicals Committee and Working Party on Chemicals, Pesticides & Biotechnology; 17-19 Nov 2004
• ECB preliminary Guidance Document published in Nov 2005
• OECD Guidance Document published in Feb 2007
• OECD Guidance summarised in REACH guidance (IR and CSA)
QPRF captures information on the substance and its prediction, and is intended to facilitate considerations of the adequacy of a prediction

1. Substance information
2. General (administrative) information on QPRF
3. Information on prediction (endpoint, algorithm, applicability domain, uncertainty, mechanism)
4. Adequacy (optional, legislation-specific, and includes judgement and indicates whether additional information is needed for WoE assessment)

- Assessment of adequacy depends on reliability and relevance of prediction, but also on the availability of other information, and the consequence of being wrong
- Not just a scientific consideration, but also a policy decision
Outline of a non-testing strategy

1. Existing information
2. Preliminary assessment of reactivity & fate
3. Classification schemes & structural alerts
4. Preliminary assessment of reactivity, fate & toxicity
5. Chemical grouping & read-across
6. QSARs

Working Matrix

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metabolite 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metabolite 2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Assess adequacy
Conclusion or targeted testing
Step 1: Information collection

- Chemical composition (components, purity/impurity profile)
- Structure generation and verification
- Key chemical features (functional groups, protonation states, isomers)
- Experimental data: physicochemical properties, (eco)toxicity, fate
  - Freely-accessible web resources (ESIS, ChemSpider, PubChem, AMBIT2)
  - Databases in freely-available software tools (OECD Toolbox)
  - Commercial databases (Vitic, …)
- Estimated data: pre-generated QSAR or read-across estimates
  - Freely-accessible web resources (ChemSpider, Danish QSAR database)
  - Chemical category databases (OECD Toolbox)
European chemical Substances Information System

ESIS (European chemical Substances Information System), is an IT System which provides you with information on chemicals, related to:

- EINECS (European Inventory of Existing Commercial chemical Substances) O.J. C 146A, 15.6.1990.
- ELINCS (European List of Notified Chemical Substances) in support of Directive 92/32/EEC, the 7th amendment to Directive 57/548/EEC,
- NLP (No-Longer Polymers),
- BPD (Biocidal Products Directive) active substances listed in Annex I or IIA of Directive 98/8/EC or listed in the so-called list of non-inclusions,
- PBT (Persistent, Bioaccumulative, and Toxic) or vPvB (very Persistent and very Bioaccumulative),
- C&L (Classification and Labelling), substances or preparations in accordance with Directive 67/548/EEC (substances) and 1999/45/EC (preparations),
- Export and import of Dangerous Chemicals listed in Annex I of Regulation (EEC) No 304/2003,
- HPVCs (High Production Volume Chemicals) and LPVCs (Low Production Volume Chemicals), including EU Producers/Importers lists,
- IUCLID Chemical Data Sheets, IUCLID Export Files, OECD-IUCLID Export Files, EUSES Export Files,
- Priority Lists, Risk Assessment process and tracking system in relation to Council Regulation (EEC) 793/03 also known as Existing Substances Regulation (ESR).

http://ecb.jrc.ec.europa.eu/esis/
1. Structure Generation from Name – ACD Labs Name

2. Validation of structures by CAS-Registry random sample

3. Manually refinements, data mining e.g. DSSTOX

4. Merge, filter, standardisation of structure representation for organic substances

5. Descriptors: substance identification and estimated physical-chemical data (ACDLabs, Pipeline Pilot, ADMET Predictor)

Availability of Structures and public experimental data for Pre-registered substances

- All: 144953
- Registration 2010: 54816
- Registration 2013: 59718
- Registration 2018: 30419

PRS structure generated: 78556
DSSTOX: 30792

<table>
<thead>
<tr>
<th>Category</th>
<th>Prs</th>
<th>Pprs</th>
<th>Dsstox</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>144953</td>
<td>78556</td>
<td>30792</td>
</tr>
<tr>
<td>Registration 2010</td>
<td>54816</td>
<td>5115</td>
<td>6115</td>
</tr>
<tr>
<td>Registration 2013</td>
<td>59718</td>
<td>387</td>
<td>35660</td>
</tr>
<tr>
<td>Registration 2018</td>
<td>30419</td>
<td>8746</td>
<td>290</td>
</tr>
</tbody>
</table>
Validation of generated structures

NTS validation

Random sample 400
CAS structures 346
NTS structures 294
NTS ok 221
NTS Warning* 49
NTS *Dictionary* 46
NTS Warning* 24
Validation ok 12

total

Legend:
- total
- Validation ok
ENDDA – Endocrine Active Chemicals Database

web-accessible database under development
Step 2: Preliminary assessment of reactivity & fate

- Prediction of abiotic / biotic reactivity to identify reactive potential and possible transformation products / metabolites

- Freely-available software
  - CRAFT (Chemical Reactivity & Fate Tool)
  - START (Structural Alerts in Toxtree)
  - OECD Toolbox

- Commercial software and databases
  - CATABOL, TIMES, Meteor, Mexalert, MetabolExpert …
  - MetaPath, SciFinder, MDL Reaction Database …
- Collaboration with Molecular Networks (Germany)
- Toxtree plug-in
- Estimates biodegradation potential
• Collaboration with Molecular Networks (Germany)
• Generates & visualises reactions, ranks transformation products
• Initial emphasis on abiotic processes & microbial biodegradation
• Data model based on AMBIT technology
• User can modify knowledge base and rulebase
Models and rulebases for mode-of-action classification, hazard identification, hazard classification and potency prediction

Freely-available software
- Episuite, Toxtree, AMBIT2, OECD Toolbox …
- OpenTox framework (http://www.opentox.org)

Commercial software
- DEREK, MultiCASE, HazardExpert, ToxAlert, ToxBoxes …
- *Insilico* first consortium (Multicase Inc, Lhasa Ltd, Molecular Networks GmbH, Leadscope Inc)

QSAR Model Databases (QMDBs)
- JRC QSAR Model Database
- OECD Toolbox
Toxtree is a flexible, user-friendly, open source application, which is able to classify chemicals into modes of action and estimate toxic hazard by applying decision tree approaches.

Collaboration with Ideaconsult (BG)

Rulebases in version 1.51 (June 2008):

- Acute Fish Toxicity (Verhaar scheme)
- Oral systemic toxicity (Cramer scheme)
- Skin irritation & corrosion potential (BfR rulebase)
- Eye irritation & corrosion potential (BfR rulebase)
- Mutagenicity & carcinogenicity (Benigni-Bossa rulebase)

Main screen in Toxtree

- Compound properties
- Prediction
- Compound structure
- Reasoning
Toxtree Predictions for Carcinogenicity

Alerts / QSAR for carcinogenicity

Potential carcinogenic QSAR
Unlikely Carcinogenic QSAR
genotoxic alerts
non-genotoxic alerts

ADMET Risk estimates from ADMET Predictor
The implementation of the Cramer classification scheme in the Toxtree software was evaluated to evaluate its concordance and highlight potential software modifications.

The results were promising with an overall good concordance between the reported classifications and those generated by Toxtree.

Improvements for Toxtree were proposed. Notable of these is a necessity to update the lists of common food components and normal body constituents as these accounted for the majority of false classifications observed.

Step 5: Chemical grouping and read-across

- Chemical read-across within analogue and category approaches
- Biological read-across (between endpoints or species)

- Chemical grouping by a top-down approach
  - Supervised and unsupervised statistical methods
  - Ranking methods (DART)

- Chemical grouping by a bottom-up approach
  - Freely available tools with analogue-searching capability (Toxmatch, AMBIT2, AIM, PubChem, OECD Toolbox)

DART (Decision Analysis by Ranking Techniques) is a flexible, user-friendly, open source application, which is able to rank and group chemicals according to properties of concern.

- collaboration with Talete srl (Italy)
- supports priority setting of chemicals

Pavan M & Worth AP (2008). A set of case studies to illustrate the applicability of DART (Decision Analysis by Ranking Techniques) in the ranking of chemicals. EUR 23481 EN.
Toxmatch: chemical similarity tool

Collaboration with Ideaconsult (BG)

Supports:
- chemical grouping & read-across
- comparison of training & test sets

Many-to-one read-across of a quantitative property \((k\) Nearest Neighbours)
Example of read-across in Toxmatch

- BCF of aniline predicted on basis of effective diameter, maximum diameter and LogP
- Predicted LogBCF = 1.05
- Experimental LogBCF = 0.78 (Hazardous Substances Databank)

Training set of 610 chemicals

Aniline - test chemical

Pairwise similarity between aniline and training set compounds
Need to identify and use relevant, reliable and well documented (Q)SARs

The JRC QSAR Model Database is a searchable inventory of peer-reviewed information on (Q)SAR models

Developers and users of (Q)SAR models can submit information on (Q)SARs by using the (Q)SAR Model Reporting Format (QMRF)

Step 6: JRC QSAR Model Database

http://qsardb.jrc.it

QMRF (xml)
QMRF (sdf)

Access through Internet
Upload of QMRF
Upload of training & test sets
Download of QMRF
Ability to search QSAR database

QMRF pdf report
QMRF xml file
QMRF excel file
QMRF sdf file
Searching the QSAR database

- QMRF No.
- Free text
- Endpoint
- Algorithm
- Software
- Authors

- CAS No.
- Formula
- Chemical name
- Alias
- SMILES

http://qsardb.jrc.it
• Need to assess the toxicological significance of pesticide active metabolites and degradation products (not tested under *Directive 91/414/EEC*)

• Three projects funded by EFSA (2009-2010)
  • Applicability of QSAR analysis in assessing metabolite toxicity
  • Applicability of the TTC concept in assessing metabolite toxicity
  • Impact of metabolism on toxicological properties

• Next steps by EFSA
  • Opinion of the PPR panel (2010-2011)
  • Guidance document on pesticide residue definition for dietary risk assessment (2011-2012)
Optimized Strategies for Risk Assessment of Industrial Chemicals through Integration of Non-Test and Test Information (OSIRIS)

http://www.osiris-reach.eu/
Concluding remarks

- To optimise the use of non-testing data, a conceptual framework is provided in the REACH guidance documentation.
- There is a need to incorporate mechanistic knowledge in the models (e.g., based on chemical reactivity and “omic” data).
- An increasing number of models are being implemented in a range of software tools.
- There is a need to facilitate the use of multiple tools by developing automated workflows.
- Further guidance is needed on how to assess the adequacy of non-testing and alternative test data by weight-of-evidence approaches.