Use of QSAR models for REACH and the CAESAR approach

http://www.caesar-project.eu/
we will NOT discuss about QSAR theory in general

we will NOT discuss the utility of QSAR for REACH in general

we will present the specific CAESAR approach and results
Our target is NOT to solve all the issues related to QSAR but to provide some tools useful for a SPECIFIC APPLICATION.
THE CAESAR Endpoints

Bioconcentration
Skin Sensitization
Mutagenicity
Carcinogenicity
Developmental Toxicity

http://www.caesar-project.eu/
**Long-term bird toxicity**

**In vivo skin irritation**

**In vivo eye irritation**

**Acute oral toxicity**

**Acute dermal toxicity**

**Acute inhalation toxicity**

**Short-term fish toxicity**

**Long-term fish toxicity**

**Developmental toxicity screening**

**Accumulation**

**Short-term repeated dose**

**Carcinogenicity**

**Sub-chronic toxicity**

**Long-term repeated toxicity**

**Short-term fish toxicity**

**Acute inhalation toxicity**

**Acute dermal toxicity**

**Acute oral toxicity**

**In vivo eye irritation**

**In vivo skin irritation**

**Long-term bird toxicity**
CAESAR had a defined BUDGET and DURATION

we wanted to OPTIMIZE the use of models

we wanted to provide tools with maximum benefits for the USERS

we focused our attention on these endpoints, because they are IMPORTANT and, hopefully, FEASIBLE
THE USER INTEREST IS OUR TARGET

NOT a new algorithm

NOT a new chemical approach

NOT a new theory

NOT a general approach

NOT a proposal of models we had

All these targets are correct, but we aimed at SOMETHING ELSE
to take into account the specific REACH requirements in addition to QSAR model requirements
for REACH

Experimental data high quality

QSAR evaluations have to increase their horizons

not all inputs are equally good
not all outputs are equally good
QSAR IS LIKE A BRIDGE

between DATA and USE

between INPUTS and OUTPUTS

Are all the bridges the same?

NO
we also used previous experience

DEMETRA

Quantitative Structure-Activity Relationship (QSAR) for Pesticide Regulatory Purposes

Editor: Emilio Benfenati

http://www.demetra-tox.net/
in the case of *drugs*/*pesticides* we need data for ANIMALS

in the case of *cosmetics* we need data from ALTERNATIVE METHODS

for REACH we use ALL INFO

*DIFERENT STRATEGIES*

a) in case of single QSAR  
b) in case of COMBINED TOOLS
for REGULATORY PURPOSES it is important to minimize the false negatives.

INDUSTRY, during compound development, has other TARGETS. different QSAR needed.
the *balance* between *false positives* and *false negatives* depends on:

1. **THE USE OF THE METHOD** alone or combined, *see the legislation*

2. **POLITICAL AGREEMENT** it is not only scientific
CEASAR developed a strategy to turn the model in one direction (max accuracy) or another (max sensitivity)

Typically QSAR models are evaluated using square parameters ($R^2$, $Q^2$)

CEASAR developed new strategies for evaluation of regression models
the threshold values are relative to the LEGISLATION

false negatives are relative to thresholds

different optimization depending on to the LEGISLATION
within REACH different models may be preferable, depending on the tonnage.

For models for risk assessment, we need continuous values, to be compared with the exposure level.

in other cases we need a classification: different evaluations.
it has to be demonstrated that QSAR works in the specific case

Proof of principle

“Any theory has to be demonstrated”
CAESAR is a step in the direction of safer, more extensive use of QSAR for REACH.

YOUR INPUTS for future steps are very important.
Important features for REACH

according to the *REACH regulation (Annex XI)*
a QSAR is valid if:

- the model is recognized to be scientifically valid;
- the substance is included in the applicability domain of the model;
- results are adequate for classification, labelling and risk assessment;
- adequate documentation of methods is provided.
SCIENTIFICALLY VALID: proof, does it work?

within CAESAR validation done according to different procedures:

• sound statistical internal validation;
• external test set(s);
• predictivity assessed.
VALIDATION

- $R^2$, $R'^2$, slope, slope', leave-one-out;
- attention to false negatives;
- external test set: split of the data into a training and test (20%);
- when possible further test sets;
- Collaboration is welcome on validation and checking.
in most cases the target has been a model of wide applicability;

specific limitations as common in QSAR: neutral form, no chirality, no polymers;

boundaries also defined by descriptor range;

limitations for individual models: rules for higher uncertainty for certain chemical classes.
APPLICABILITY - UNCERTAINTY

- **A PRIORI** conditions, as typically performed: *chemometric tools, on model’s basis/inputs*;
- **A POSTERIORI** conditions, on the basis of the results/outliers, identifying chemical groups with higher uncertainty;
- not black/white, but uncertainty is given.
A POSTERIORI CHECK OF APPLICABILITY

- performances per chemical category;
- number of compounds;
- average error, compared with experimental value;
- number of false positives/negatives;
- presentation of the results for similar compounds.
DEFINITION OF INTENDED USE

- classification or continuous values;
- suitable for C&L or RA;
- attention to the EU thresholds.
Adequate Documentation - 1

- **Toxicity/property data on the web**
  All values given, including multiple values

- **Chemical structure given**

- **Chemical descriptors/fragment**
  This includes the formula for descriptor. In several cases it varies depending on the software or on the version.
ADEQUATE DOCUMENTATION - 2

• **Mathematical algorithm** of the *in silico* model is given;
• **The methodology** is provided and the **parameters** fixed;
• **Full results provided** (raw data and statistical evaluation).
since the algorithm is given (including chemical descriptors) the model is fully transparent and reproducible

the same value obtained in Italy, Finland, France, etc.

the same model used by regulators and industry
CAESAR MODELS: ADVANTAGES

- Planned for REACH: reference to legislation & QC
- Planned to be used, transparent, validated
- Some models can replace assays, others are supplementary info
- Models for the future: implementable
- Modern IT basis (Java, Python, ...)
- Public, easy access
Future Steps - 1

Feed-back from USERS

Your OPINIONS

OPTIMIZATION of the models

IMPLEMENTATION of more CAESAR models

INTEGRATION of the models
Future steps - 2

- collaboration with US EPA
- collaboration with CHEMPREDICT
- use and test within OSIRIS
- dissemination (ORCHESTRA)
- your SUPPORT, your IDEAS
GRAZIE!