CADASTER

CAse studies on the Development and Application of in-Silico Techniques for Environmental hazard and Risk assessment

Grant agreement no.: 212668

Collaborative Project

Sub-Priority ENV2007 3.3.1.1: In-silico techniques for hazard-, safety-, and environmental risk-assessment

Evaluation of QSAR models in the legal framework

Due date of deliverable: 31 December 2011
Actual submission date: 22 December 20111

Start date of project: 1 January 2009
Duration: 4 years

Lead Contractor: National Institute of Public Health and the Environment (RIVM), Laboratory for Ecological Risk Assessment

Corresponding authors of document: Magnus Rahmberg[1], Annika Tegeback[1], Linus Bergfors[1]

1. IVL Swedish Environmental Research Institute, Box 210 60, SE - 100 31 Stockholm, Sweden (magnus.rahmberg@ivl.se)

Deliverable no.: 4.5 Evaluation of QSAR models in the legal framework
Nature: Report

Project co-funded by the EU Commission within the Seventh Framework Programme		
Dissemination Level		
PU	Public	X
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

General

CADASTER is a project that was granted within the 7th Research Framework Programme of DG Research of the European Commission. CADASTER aims at providing the practical guidance to integrated risk assessment within REACH by carrying out a full hazard and risk assessment for chemicals belonging to four compound classes. The main goal is to exemplify the integration of information, models and strategies for carrying out safety, hazard and risk assessments for a selected number of compounds within four specific chemical domains. Real hazard estimates will be delivered according to the basic philosophy of REACH of minimizing animal testing, costs, and time. CADASTER will show how to increase the use of non-testing information for regulatory decision whilst meeting the main challenge of quantifying and reducing uncertainty.

CADASTER has officially started on the 1st of January, 2009. The project officer on behalf of DG Research of the European Commission is Dr. Georges Deschamps, the project is coordinated by Dr. Willie Peijnenburg (RIVM).

Table of Content

1. Objective and background

The objective stated in the Description of Work, DoW, for task 4.4 “QSAR models in the legal framework” is: “Evaluation of methods and decision points for the establishment of scientific validity and applicability domains for QSAR models. Evaluation of the need for documentation, with regard to current progress in the OECD and REACH implementation”. Since the start of the project and the actual time frame for this task, the first registration round in REACH for chemicals produced and used at the highest tonnages, has been finalized. This means that information now is available on the outcome of the use of QSAR in REACH. We will in this report review this information and also summarize the status from different other sources and projects about the use of QSAR in REACH.

2. Evaluation of the use of QSAR in REACH

This section will give a background of the guidelines and recommendations that are available for the use of QSAR in REACH and thereafter a review of the use of QSAR in the first round of registration in REACH.

2.1. Guidelines for QSAR models

There are several guidelines available for the use of QSAR. These guidelines describe how to use and report QSAR in regulatory purposes. They all are in various aspects based on the OECD principles of QSAR. These principles were agreed on in 2004 and published in the Guidance Document on the Validation of (Quantitative) Structure Activity Relationship [(Q)SAR] Model, 2007.

The OECD principles are as follows: to facilitate the consideration of a QSAR model for regulatory purposes, the model should be associated with the following information:

1. a defined endpoint;

   where endpoint refers to any physico-chemical property, biological effect (human health or ecological) environmental fate parameter that can be measured and thereafter modelled. The intent of this principle is to ensure transparency in the endpoint being predicted by a given model, since a given endpoint could be determined by different experimental protocols and under different experimental conditions.

2. an unambiguous algorithm;

   The intent of this principle is to ensure transparency in the description of the model algorithm.

3. a defined domain of applicability;

   The need to define an applicability domain expresses the fact that (Q)SARs are reductionist models which are inevitably associated with limitations in terms of the types of chemical structures, physico-chemical properties and mechanisms of action for which the models can generate reliable predictions.

4. appropriate measures of goodness-of-fit, robustness and predictivity;

   This principle expresses the need to provide two types of information: a) the internal performance of a model (as represented by goodness-of-fit and robustness), determined by using a training set; and b) the predictivity of a model, determined by using an appropriate test set.

5. a mechanistic interpretation, if possible.

   The intent of this principle is to ensure that there is an assessment of the mechanistic associations between the descriptors used in a model and the endpoint being predicted, and that any association is documented. Where a mechanistic interpretation is possible, it can add strength to the confidence in the model already established on the basis of Principles 1-4.

The use of QSAR is in the REACH framework treated in Article 13, 25 and in Annex XI. Article 25 declares that tests on animals should be a last resort when all other options, such as QSAR, are considered. In Article 13 the rules are given for how to generate necessary information by actual testing or by QSAR models. In Annex XI it is stated that a result of QSARs may be used if certain criteria are fulfilled:

- results may be derived only from a (Q)SAR model whose scientific validity has been established
- the substance must fall within the applicability domain of the (Q)SAR model
- results must be adequate for the purpose of classification and labelling and/or risk assessment
- adequate and reliable documentation of the applied method needs to be provided

Considering the use of QSAR in REACH there are a few guidance documents and a number of practical guides published by ECHA:

- Regulation (EC) No 1907/2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH)

- Practical guide 5: How to report (Q)SARs
- Practical guide 10: How to avoid unnecessary testing on animals
- Guidance on information requirements and chemical safety assessment – Chapter R.6: QSARs and grouping of chemicals

We will not go into details in these guidelines. In principle they are all referring to the OECD criteria's for a valid QSAR model and that a QSAR model should be documented in the QSAR Model Reporting Format (QMRF). Also, the prediction from a QSAR model should be reported in a QSAR Prediction Reporting Format (QPRF).

In a recent report from JRC-ICHP (Joint Research Centre – Institute for Health and Consumer Protection) the frame work for documenting QSARs and their predictions are reviewed and it is documented in this report how to develop the frame work for documenting QSARs for better guidance in regulatory purposes, JRC (2011). A check list is suggested of 10 key questions that the risk assessor should go through when evaluating a QSAR model in a regulatory purpose. Not all questions should be answered if they are not needed in the actual regulatory context, and in the same way additional questions could be added. These questions will assess the practical applicability of the QSAR and the adequacy of the predictions. The questions are presented in Appendix 2.

The focus in this report is on the use of QSAR in regulatory purposes and especially the application under REACH. Parallel to the more direct use, in the regulatory frame work as data gap filler and replacement of animal test, QSAR models have been and are used in the identification and prioritization of compounds that are of concern but not yet recognized or tested. The Danish EPA has used QSAR model in this approach in the

screening of possible reproductive toxicants from a structure set of 57,014 chemicals from the European Inventory of Existing Chemical Substances (EINECS), (Jensen *et al.*, 2008). The result from the screening was that 9.2 % of the chemicals evaluated were predicted as reproductive toxicants. Based on this the advisory classification list with the suggested classification for reproductive toxicity could be updated.

2.2. Evaluation of QSARs used in REACH

2.2.1. Status of reported chemicals in the first registration round for REACH

The deadline for the first registration round in REACH was the 30th of November 2010. The outcome of this registration round has now been evaluated and reported by ECHA. The more general evaluation is the Report on the Operation of REACH and CLP 2011, ECHA 2011a. The total number of registered substances was 4 300 in around 25 000 different dossiers. The evaluation consists of two parts, the compliance check and the examination of test proposals. The compliance check examines if the submitted dossier fulfils the REACH legislation requirements; it is not an evaluation of the quality of the submitted data. In REACH at least 5% of the submitted dossiers will be evaluated according to the compliance check. This work has started in ECHA and the first checks show that many of the submitted dossiers are missing information required or that the reported information is inadequate. Regarding the testing proposals submitted in the dossiers to fill information gaps in the requirements, ECHA has almost finalised the evaluation. It can be concluded that the submitted proposals are in almost every case accepted by ECHA, some with minor changes, and thus they are needed and necessary

for the fulfilment of the REACH requirements. ECHA will also start an evaluation based on individual substance in 2012.

Based on the submitted dossiers ECHA has published a report, The Use of Alternatives to Testing on Animals for the REACH Regulation 2011, ECHA 2011b. In this report, the outcome of used alternative methods in the first registration round is summarized. The analysis starts at all the dossier submitted up to the deadline at the 30th of November 2010 and for:

- Phase-in substances at or above 1 000 tonnes per annum, tpa
- Phase-in substances at or above 100 tpa and
- Non-phase-in substances at or above 100 tpa.

Since the owner of the dossier has the possibility to update the information in his dossier at any time, ECHA set a cut off date by the 28th of February 2011 for retrieval of data from the database. This selection resulted in 17 062 dossier and from this set reported 'chemicals categories', i.e. non-specific substances (85 dossiers), were excluded. Also substances classified as intermediates were left out of the analysis. Since many of the dossiers are results of Substance Information Exchange Forums (SIEFs) only lead dossiers are included and also dossiers submitted individually for specific chemicals. The final set of dossiers for analysis was 1 862, which corresponds to 1 789 substances. The total amount of substances reported in the first registration round is 4 599.

ECHA has selected three different approaches in the analysis of the data set described above.

1. Endpoint Study Record (ESR) approach

   This perspective gives the cumulative picture for each endpoint reported in the submitted dossiers

2. Substance approach

   Analysis of how the requirements on a substances base are fulfilled with the help of alternatives methods.

3. Studies conducted or proposed for the purpose of REACH

   Retrieves data in the dossiers if a test on vertebrate level is proposed to fill data gaps or if existing test results, reported in 2009 or later, are used to fulfil the requirements.

There are 33 test types to be reported in the REACH registration process. This includes different routes of the chemical into the test object and also the duration of the test. But to compromise the result, the different routes and the test duration are merged together in one resulting test type for each category. This resulted in 20 test types presented.

In the reporting of the substances seven different categories can be chosen for how the information of the current test type is derived. They are:

- Experimental studies.
- Testing proposal.
- Read-across.

- IUCLID flags to omit the study. These options are to be used to indicate when testing does not appear to be: scientifically necessary; technically not possible; or not necessary based on low exposure considerations (abbreviation FO).
- Weight of Evidence.
- QSAR.
- Miscellaneous. None of the above fit the information inserted in the registration process.

The result for the ESR approach is presented in Appendix 1. In figure 1, this percentage is shown for the different methods per test type. Here the different tonnage span and the phase-in and non-phase-in are summarized to get an overview of the alternative methods used. For a more detailed picture refer to Appendix 1 where the full results for the ESR approach are given in a table.

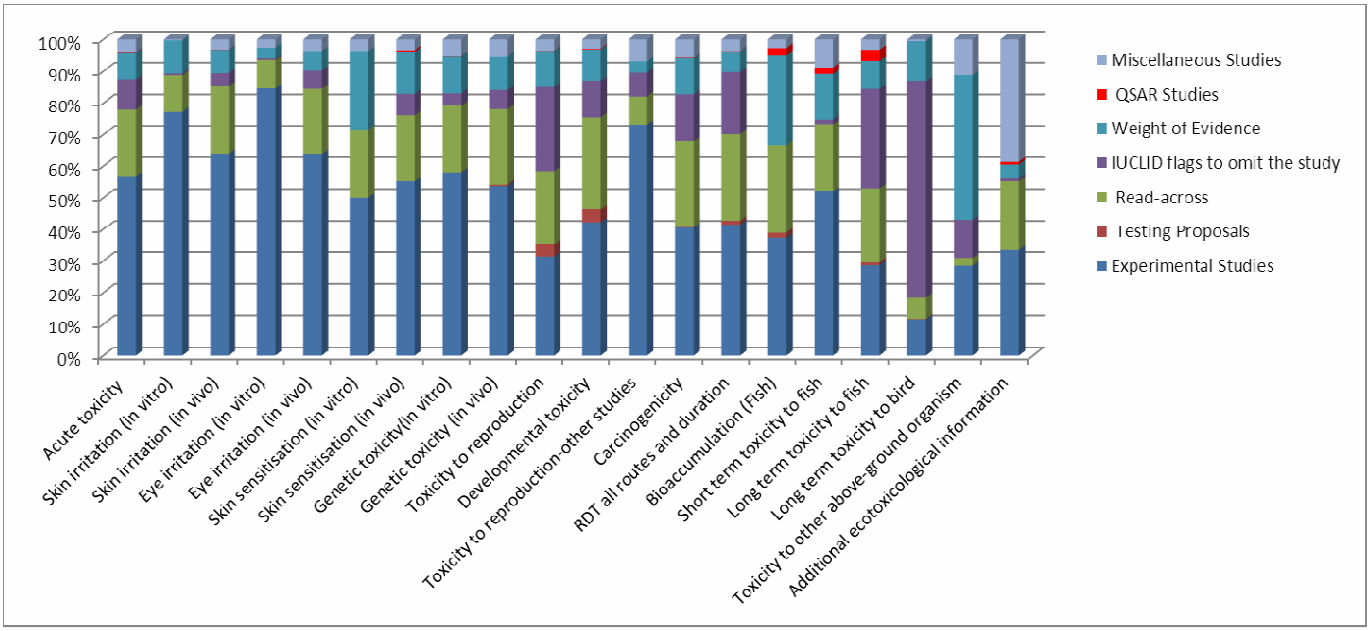

Figure 1. The percentage of methods used for each of the 20 test types.

Registrants in the first registration round in REACH have almost in every test type given experimental studies as the main source. The dominating alternative method reported is read-across. For the use of QSAR, which is our focus, there are very little results reported. In twelve of the test types QSAR has been reported but the number is very small, less than 1 percentage. These test types and endpoints include: Acute toxicity, Skin irritation (in vivo), Skin sensitisation (in vivo), Genetic toxicity (in vitro), Toxicity to reproduction, Developmental toxicity, Carcinogenicity, RDT all routes and duration, and "Additional ecotoxicological information". The figures are a little bit higher for the test types concerning the environmental compartments: Bioaccumulation (Fish), Short term toxicity to fish, Long term toxicity to fish. For these types the use of QSAR is about 2-3 percent. This corresponds to respectively 25, 267 and 153 actual number of QSAR models used. In total 433 QSAR models were reported.

The substance approach was not reported as detailed as the ESR since the alternative methods were counted together and hence no conclusions can be drawn on the amount of QSAR used for different substances or groups of substances.

The last evaluation is on studies conducted and proposed for the purpose of REACH. As described above all tests that have been reported during 2009 and later are classified as being performed for the REACH requirements. The outcome of this analysis is presented in table 1.

Table 1. Conducted and planned experimental studies

Experimental studies	in vitro	in vivo	Total
Conducted	1 491	1 849	**3 340**
Proposed	-	711	**711**
Total	1 491	2 560	**4 051**

In total there are 4051 conducted and planned experimental studies that are connected to the REACH requirements. The tests that have been performed for fulfilling the requirements are a very little part of the total amount of test results used in the registration process.

In the CADASTER workshop in Maribor, Slovenia, in September 2011 Evelin Fabjan from ECHA held a presentation about the status of REACH and especially the use of QSAR based on the reports “The Use of Alternatives to Testing on Animals for the REACH Regulation 2011” and “Report on the Operation of REACH and CLP 2011”. ECHA had some remarks on the QSAR models reported that are not mentioned in these reports but are known since ECHA has started to evaluate the reported QSARs. The findings are:

- Often, limited information is included in the dossier about the (Q)SAR model used (e.g. version unclear, data on the model not transparent/ available);
- The scientific validity of the models is not always demonstrated;
- The applicability domain of the models often is not analysed (or only partially analysed);
- The model and/or endpoint is not relevant for regulatory purposes (e.g. the endpoint predicted is not suitable to meet the information requirements of REACH).

2.2.2. Questionnaire

We have send out a questionnaire with 5 questions. Being representative for the view of industry, the answer from International Flavors & Fragrances Inc. (IFF) is presented below.

1. Did IFF make any registration in REACH, if so have you used any In Silico methods?

IFF has done several registrations in 2010 both as lead registrant and as co-registrant. For human health endpoints, we have used in silico methods:

1a) We used a QSAR for eye irritation in a Weight of Evidence (WoE) approach in combination with an in vitro eye corrosion test.

1b) We used read across approaches to fill data gaps on repeated dose toxicity and/or reproductive toxicity.

- For ecotoxicity we have not used QSARs in our registrations, but we have used them to identify potential R50/53 classifications for our > 100 ton chemicals. This was done because when a substance is/had to be labelled as R50/53, the substance also needed to be registered in 2010.

- For physico-chemical properties testing is preferred above QSARs, especially for those properties which are important to estimate the environmental fate.

- For biodegradation we use the models of US-EPA to indicate the biodegradability potential of substances, but a test is performed for REACH as well.

- BCF (Q)SARs are often used to estimate the potential for bioaccumulation within risk characterisation and PBT assessment. For C&L the criteria for QSAR use for this endpoint seem more stringent.

2. In the report "The Use of Alternatives to Testing on Animals for the REACH Regulation 2011" from ECHA it is clear that the use of QSAR is very limited. In your opinion what is the reasons for this and why is the use of read-across greater compared to QSARs?

- One reasons for the limited use of (Q)SAR predictions as standalone information is that QSARS may have been included in the IUCLID sections for WoE.

- QSAR are generally more valid for acute and/or local endpoints than for long term (eco)toxicological endpoints. The 2010 registered chemicals are likely to be data rich considering acute toxicity, skin and eye irritation (local) endpoints and possibly also acute aquatic toxicity.

- Ecotoxicological (Q)SARs for long term toxicity (e.g. ECOSAR) are not always valid e.g. R^2 is < 0.7 and/or based on a limited number of chemicals. ECOSAR may overpredict the toxicity of chemicals used as fragrances, which is indicated in: Daniel T Salvito, Ronald J Senna, Thomas W Federle 2002, a framework for prioritizing fragrance materials for aquatic risk assessment, published in Environmental Toxicology and Chemistry.

- Another reason may be that the translation of human health (Q)SAR predictions, insufficiently meets the REACH need for classification and labelling and risk characterisation.

-There could also be a financial drawback for using (Q)SARs for short term (eco)toxicological endpoints, because it takes quite a lot of expert judgment and efforts to make QMRFs and QPRFs for adequate documentation. Hence, this may be more expensive than doing the actual test.

-Read across may often be more convincing and more explanatory compared to (Q)SARs, although many (Q)SARs share the principle that similar chemicals are expected to have similar properties.

3. What are the organizational risks of using QSAR?

- For acute endpoints, testing gives more certainty. This is especially important for classification and labelling (C&L). C&L of substances has a large impact on handling and marketing the substance and thus over-classification has important consequences. This is in contrast to risk characterisation where the uncertainty of the testing or (Q)SAR results can be more easily taken into account.

- In addition, as indicated above using (Q)SARs for acute (eco)toxicological endpoints is not necessarily cheaper than testing because QMRF and QPRFs need to be written and the uncertainty of (Q)SARs is generally considered higher.

Added by IFF: 3.a. What can be the potential organizational benefits of using QSAR?

- (Q)SARs for acute fish toxicity may be used more when it is shown that fish is generally not more toxic than Daphnia as is being presented in the Cadaster project for (benzo) triazoles interspecies model (oral communication with Paola Gramatica)

- (Q)SARs for long term fish toxicity may also be used when the result is similar to the long-term for NOEC Daphnia or Fish, because in that case the Assessment Factor for risk characterisation can be lowered. On the other hand, when the (Q)SAR indicates higher toxicity (lower NOEC) compared to the available testing data, the data may not be used because of the impact on C&L.

- (Q)SARs for BCF will be used for PBT or vPvB assessment, e.g. EPA PBT profiler or directly the QSAR for PBT Index (Papa and Gramatica, Green Chemistry, 2010).

- When QMRs are already presented on the JRC website, this will facilitate their use. The QSARs of the CAESAR project I have used for screening sensitisation and BCF. They provide also read across chemicals which help checking the reliability of the results. The reliability of the BCF results for fragrances needs to be verified.

4. What are the organizational risks of using Read-Across?

Read across is mostly used for human health endpoints: systemic dose and reproductive toxicity and may also be used for the long-term ecotox endpoints e.g. fish toxicity and/or bioaccumulation.

4.a. What are the potential organizational benefits for using Read-Across?

- A read across for human health provides generally data that can be used both for hazard assessment, C&L, PBT (vPvB) and risk characterisation.

- A read across for ecotoxicity may be important for the longer term endpoints such as long-term fish toxicity and BCF. For 2010 we have not used it, but we may do more for 2013.

5. What do you think of the OECD principles for reporting and using QSARs?

- The OECD principles for the evaluation of (Q)SARs are very helpful in finding the strengths and limitations of the (Q)SARs. It is also of great help when developers of

(Q)SARs already document their (Q)SARs according to these principles, in QMRFs or QPRFs, because the user of the model only needs to document the prediction.

- These principles are also very helpful for the evaluation of read across.

2.2.3. Review of performed interviews/questionnaires about use of QSAR in industry

ORCHESTRA is a project funded by the EU that aims at promoting a wider understanding, awareness and appropriate use of in-silico methods. The project is funded to disseminate recent research on computer-based in-silico methods for evaluating the toxicity of chemicals (ORCHESTRA 2011b). Within the ORCHESTRA project, stakeholders have been interviewed and the results are of interest for the CADASTER project. Therefore, a review of the outcome of ORCHESTRA has been carried out.

REACH provides possibilities to use existing data or alternative assessment methods (e.g. In-silico methods). In-silico methods can be used to replace animal testing to some extent. In-silico methods use findings from both in vivo and in vitro laboratory tests, but the repetition of animal testing can be reduced as further tests are replaced by computer modelling. Animal experiments are associated with costs but some in-silico tools are available online, free of charge, which can be useful for e.g. users and importers of chemicals. REACH supports the use of alternative methods, but the European Chemicals Agency (ECHA) is more restrictive in accepting their use as serious consequences may rise due to defective assessments (ORCHESTRA 2010).

Read-across is a simplified version of a QSAR model. The property of one or few chemicals is predicted on the basis of one or more similar compounds (ORCHESTRA 2011c).

ORCHESTRA interacts with several stakeholder organisations and individuals, including EU and member state regulators, industry, small companies, QSAR developers, policy makers, scientists, educators, NGOs and citizens (ORCHESTRA 2011a).

The use of QSAR is disputed. According to ORCHESTRA (2011d), the major concerns deal with the reliability of in silico models and the complexity of the mechanisms and processes underlying toxicity. There are also views on animal experiments and a major criticism is that real results only can be provided by real experiments. There is a discussion on validation of the in silico models. The possible use of a model should be defined clearly. Work is carried out to enhance the models and on the integration of the models.

A documentary, QSARs in REACH, has been produced within the ORCHESTRA project. This documentary is based on interviews with regulators, industry and developers. The documentary is available online (ORCHESTRA 2011e). This section includes a summary of the content of the documentary:

Regulators:

- Under REACH, the industry must submit registration dossiers to ECHA that document the safe use of chemicals. ECHA is there to ensure the safe use of a substance (Professor Wim de Coen, Head of Evaluation 1 at the European Chemicals Agency (ECHA), in *QSARs in REACH*).

- Under previous legislation there was no need for in-silico methods but with REACH many chemicals are assessed in a short time and there is no choice but to use alternative methods (Bob Diedrich, Principal Administrator at OECD, in *QSARs in REACH*).

- The challenge for industry is to ensure reliability and not just speed (Dr Simon Pardoe, PublicSpace Ltd. Research Communication, speaker, in Q*SARs in REACH*)).

- REACH ensure the safe use of substances for man and the environment and the properties of substances have to be documented by industry. Animal testing is the last resort. When documenting properties, alternative methods should also be used (Professor Wim de Coen, ECHA, in *QSARs in REACH*).

- QSAR need to be very well documented so that authorities can perform a transparent and independent assessment of all the criteria for the in-silico model (Professor Wim de Coen, ECHA, in *QSARs in REACH*).

- In-silico models are used more and more but for e.g. long-term mammalian toxicology endpoints, available data is not robust enough (Professor Wim de Coen, ECHA, in *QSARs in REACH*).

- QSAR is not a standalone replacement for testing. Weight of evidence approaches should be used which incorporate all the existing information (Magnus Lofstedt, Environmental Protection Agency, Denmark in *QSARs in REACH*)

- Still in the acceptance phase and developers of in-silico models need to understand how they are used in regulatory decision-making to make it easier

for the authorities and industry (Professor Wim de Coen, ECHA, in *QSARs in REACH*).

- QSARs are based on experimental data and there is a degree of uncertainty, but the numbers are not invented (Marco Valentini, Inspector, Health and Safety authority, Ireland, in *QSARs in REACH*)

Industry:

- L'Oréal uses in-silico methods to design molecules – to find new compounds for cosmetic applications. In-silico models are used to screen compounds early in the product development process. L'Oréal is integrating non-testing methods, i.e. in-silico methods, in vitro methods and physical-chemical properties, to evaluate the safety of products without relying on animal testing. (Dr Stéphanie Ringeissen, L'Oréal: Head of modelling and biological simulation, in *QSARs in REACH*)

- Cambrex are using results from QSAR models concerning physical-chemical endpoints to classify substances and prepare Safety Data Sheets. For human toxicity and ecotoxicological endpoints, Cambrex prefer to use QSAR models as supporting evidence. They use QSAR to try to fill the data gap in literature with results obtained from the modelling (Dr Laura Bigini, Cambrex Profarmaco Milano, in *QSARs in REACH*).

- Only a limited number of companies feel confident with (and understanding the approaches) using in-silico methods. The models are not a simple toolbox and QSAR results need to be evaluated. Companies need to evaluate the limitations of individual QSARs used, and that is probably one of the greatest challenges

that industry will be confronted with (Dr Erwin Annys, European Chemical Industry Council (CEFIC), in *QSARs in REACH*)

- A list is needed of laboratories that can produce documentation on the decision behind a number (Dr Maurizio Colombo Lamberti SPA: HSE Manager, in *QSARs in REACH*)

- There is uncertainty and doubt on regulatory acceptance (Dr Simon Pardoe, PublicSpace Ltd. Research Communication, speaker, in Q*SARs in REACH*).

- The forthcoming REACH registrations are for chemicals used in smaller quantities and will thereby involve a much wider range of companies, including Small Medium Industries (SMIs) (Dr Simon Pardoe, PublicSpace Ltd. Research Communication, speaker, in Q*SARs in REACH*).

- QSAR models are of great use especially for SMIs. The SMIs have to pay experts to help them with the dossiers, but the costs associated with expert help are lower than the costs associated with in vitro and in vivo experiments (Dr Alexandre Péry, Institut National de l'Environnement Industriel et des Risques (INERIS), in Q*SARs in REACH*).

- Companies are worried about the responses from member states authorities (Dr Simon Pardoe, PublicSpace Ltd. Research Communication, speaker, in Q*SARs in REACH*).

- The success of QSARs and other in silico methods depends on the acceptance from member states authorities (Dr Erwin Annys, European Chemical Industry Council (CEFIC), in *QSARs in REACH*).

- The reason for Industries to not use QSAR is that they do not believe that the method will be accepted by the authorities (and not because it is more expensive or lack of access) (Philippe Hubert, Institut National de l'Environnement Industriel et des Risques (INERIS), in Q*SARs in REACH*).

- A connection between the industries, ECHA and the member states authorities is of interest, to be surer, to follow the same procedures and to speak the same language (Dr Laura Bigini, Cambrex Profarmaco Milano, in *QSARs in REACH*).

Within the ORCHESTRA project, a survey on stakeholders' views, needs and practices regarding in-silico methods, was conducted. The result was presented in a report (ORCHESTRA 2011f) including responses received from September 2010 through April 2011. The total number of responses was considered small and thereby not robustly validated; however, some trends appeared. In total, three questionnaires were developed:

1. Benefits and barriers to the use of QSAR methods (for those involved in Toxicology or Chemicals Regulation – regulators, industry users, academics and consultants)

2. Use of QSAR/In-silico methods (the same group of stakeholders as in number 1)

3. Policy issues around in-silico methods as alternatives to animal testing (all interested persons and particularly those without specialist knowledge).

The development of QSAR models appears to take place in each stakeholder context. The specialists responding displayed different roles, duties and activities. The ORCHESTRA project grouped the sample into three stakeholder categories;

1. ACACON (academics and/or consultants)

2. REGUL (having a regulatory mission)

3. INDUS (The research, development and/or application of in-silico methods take place directly within a commercial industrial context. Chemical manufacturers or their organizations, having a direct stake in the outcome of specific REACH dossiers)

The populations numbered 13 (ACACON), 12 (REGUL) and 8 (INDUS). The majority of the respondents have used QSAR/In-silico methods. All respondents have plans to test or use QSAR in the future. In total, 28 models, methods or software have been applied by the respondents. The two most cited suites are OECD Toolbox and EPIsuite and those are cited by all three stakeholder communities. CAESAR is also cited by the three stakeholder communities. ECOSAR, SPARC and TOXTREE were cited more than once by two stakeholder categories.

Regulators diversify their attention to the range of endpoints, whereas ACACON shows the highest rate of application in the area of physico-chemical properties of compounds. Unlike the other two categories, REGUL have not often used QSAR models for prioritisation. There is a consensus across stakeholders that in-silico models will be attractive for identifying and prioritizing substances of concern.

ACACON and INDUS do not consider economic costs to be a major barrier to the use of in-silico methods. The two stakeholder categories suggest that toxicologist may not find what they need for decision making. ACACON comment on the inability of in-silico models to address the challenges of cocktail compounds or chronic toxicity. INDUS wants reassurance that the scientific quality of any given tool is acceptable and they would like some indications on the best available QSAR model. ACACON are interested in technical and scientific aspects of in-silico applications. REGUL are

looking for a “good grasp” of software outputs (the results of the model and their meaning). Regulators require relatively more understanding of the correct applications for REACH. The survey revealed a role-related gap, i.e. the demand of ACACON is scientific, INDUS want to know which models to choose with confidence and the key issue for REGUL is information on the possible application for the specific use (REACH).

3. Discussion

Results in the report “The Use of Alternatives to Testing on Animals for the REACH Regulation 2011” and the “Report on the Operation of REACH and CLP 2011” are based only on information extracted from the IUCLID database. The information is not in any way evaluated according to the guidelines and recommendations that were published by ECHA, but just a picture of actual records reported. Hence there could be no conclusion drawn on the 433 dossiers with QSAR models applied, with regard to the question whether the models are reported according to the OECD principles, or whether the predictions will be accepted by ECHA. Even if the OECD principles and guidelines are followed, it does not mean that results from a QSAR model are automatically accepted by ECHA. The models and predictions will be evaluated by ECHA and from the indication given above it is obvious that in some cases the guidelines are not followed when applying a QSAR model.

Read-across and Weight of Evidence (WoE) approaches as alternatives methods are more applied in the first registration round, than QSAR models. However, a QSAR model could nevertheless have been applied in a Read-across approach to fill data gaps. When it comes to WoE, QSARs are probably included in this approach. The outcome of a WoE approach includes different types of information collected, for a specific end-point; if a QSAR model is used the results are evaluated together with the

additional information that is available. The QSAR model is not alone in providing the information needed in the REACH requirements. This will probably be one reason for the registrant to choose the WoE approach instead of "just" one or more QSAR models since the justification and evaluation are based on more information than just the QSAR model, thus probably reducing the uncertainty on how ECHA will treat the reported QSAR model. Therefore, the registrants put all information available in the WoE approach to be on the safe side in the evaluation process. This explanation, that QSAR could be used in the WoE approach, is also given by IFF in the questionnaire.

In the statistics from ECHA, the QSAR models applied in Read-across and WoE are not included. It is hard at this stage to give any number of QSAR models inside of these approaches but these numbers certainly will be high in WoE approaches since in the guidelines there is a recommendation to incorporate all possible information on individual end-points. It will be interesting to see if ECHA will further analyse the QSARs used in these two approaches and how these models are reported.

One reason not to use QSAR for short term acute toxicity could be purely economical. When there exist no QMRF reports for the actual QSAR model utilised and this need to be compiled together with the QPRF the cost for this could be higher than performing an actual test for the end point studied.

The establishment of Data Sharing & Substances Information Exchange Forums (SIEFs) has to a great extend prevented new and unnecessary testing on animals. The use of alternative methods has also contributed in the efforts to minimize these tests. However, the uses of QSAR models are very low and a wider use of them would even more decrease experimental tests on animals. In the CADASTER project the focus is to develop and facilitate QSAR models that are developed according to all the needed

requirements to be accepted in REACH. The models are therefore documented in a proper way. If a potential registrant in the REACH process will use the functionality developed for the four chemical groups chosen in CADASTER for predicting end points needed in the dossiers, and if the compounds for which the QSAR model is run are inside the applicability domain of the QSAR model, then the registrant has a very high probability that these results will be accepted by ECHA. This is mainly because the predictive quality of the local CADASTER models for the four classes of chemicals selected within CADASTER is demonstrated to be higher than other widely used software, as EPISUITE for instance.

In the next registration round when lover tonnage substances will be registered, it is anticipated that a larger amount of Small and Medium sized Enterprises, SMEs, will submit dossiers. Then the availability of reliable tools that are in line with the requirements in REACH, is even more important. After all, the in house expertise in these SMEs is likely to be lower than in larger enterprises. This implies that SMEs will rely more on the available tools. In that sense, projects like CADASTER are of value in improving the use of adequate QSAR models.

4. Bibliography

ECHA (2011a). Report on the Operation of REACH and CLP 2011. ECHA-11-R-003-EN. European Chemicals Agency, Helsinki, Finland

ECHA (2011b). The Use of Alternatives to Testing on Animals for the REACH Regulation 2011. ECHA-11-R-004.2-EN. European Chemicals Agency, Helsinki, Finland

JRC (2011). A Framework for assessing in silico Toxicity Predictions: Case Studies with selected Pesticides. Report from the European Commission's Joint Research Centre - Institute for Health and Consumer Protection.

OECD (2007). Guidance Document on the Validation of (Quantitative) Structure Activity Relationship [(Q)SAR] Models. OECD Series on Testing and Assessment No. 69. ENV/JM/MONO(2007)2. Organisation for Economic Cooperation and Development, Paris, France. 154 pp.

G.E. Jensen, J.R. Niemelä, E.B. Wedebye & N.G. Nikolov (2008). QSAR models for reproductive toxicity and endocrine disruption in regulatory use – a preliminary investigation, SAR and QSAR in Environmental Research, 19:7-8, 631-641

ORCHESTRA (2010). In silico methods for testing the toxicity of chemicals, an introduction. http://www.orchestra-qsar.eu/sites/default/files/Introductory_leaflet_on_in-silico_methods.pdf

ORCHESTRA (2011a). http://www.orchestra-qsar.eu/stakeholders-addr (7/12/11)

ORCHESTRA (2011b). http://www.orchestra-qsar.eu/overview (7/12/11)

ORCHESTRA (2011c). http://www.orchestra-qsar.eu/about-in-silico-methods (8/12/11)

ORCHESTRA (2011d). http://www.orchestra-qsar.eu/about-in-silico-methods/the-debate (8/12/11)

ORCHESTRA (2011e). http://www.orchestra-qsar.eu/qsars-in-reach-video (8/12/11)

ORCHESTRA (2011f). D2.2, A report of barriers to disseminating results of environmental research, and stakeholders needs related to the use of in silico models. Report of on-line stakeholder survey. Proposed due date of deliverable: 2010-10-31.

Appendix 1

Results of ESR per test type and for phase and tonnage.

Test type	Phase and tonnage	Total ESR	Experimental Studies, %	Testing Proposals, %	Read-across, %	IUCLID flags to omit the study, %	Weight of Evidence, %	QSAR Studies, %	Miscellaneous Studies, %
Acute toxicity	**Phase-In>1000**	**12 874**	**56.9**	**0.0**	**21.4**	**9.2**	**8.7**	**0.1**	**3.7**
	Phase-In 100 to 1000	**1 649**	**59.9**	**0.0**	**20.7**	**10.8**	**6.9**	**0.2**	**1.5**
	Non-Phase-In>100	**396**	**38.6**	**0.0**	**12.9**	**20.2**	**5.1**	**0.0**	**23.0**
Skin irritation (in vitro)	**Phase-In>1000**	**329**	**76.6**	**0.0**	**11.9**	**0.6**	**10.6**	**0.0**	**0.3**
	Phase-In 100 to 1000	**24**	**83.3**	**0.0**	**8.3**	**0.0**	**8.3**	**0.0**	**0.0**
	Non-Phase-In>100	**1**	**100.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
Skin irritation (in vivo)	**Phase-In>1000**	**5 216**	**64.1**	**0.0**	**21.3**	**4.1**	**7.7**	**0.1**	**2.6**
	Phase-In 100 to 1000	**600**	**67.0**	**0.0**	**21.8**	**4.7**	**5.2**	**0.2**	**1.2**
	Non-Phase-In>100	**157**	**45.9**	**0.0**	**14.6**	**8.9**	**4.5**	**0.0**	**26.1**
Eye irritation (in vitro)	**Phase-In>1000**	**172**	**86.6**	**0.0**	**7.0**	**0.6**	**2.9**	**0.0**	**2.9**
	Phase-In 100 to 1000	**27**	**70.4**	**0.0**	**22.2**	**0.0**	**7.4**	**0.0**	**0.0**
	Non-Phase-In>100	**1**	**100.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
Eye irritation (in vivo)	**Phase-In>1000**	**4 221**	**64.3**	**0.0**	**20.9**	**5.2**	**6.6**	**0.0**	**3.0**
	Phase-In 100 to 1000	**524**	**65.5**	**0.0**	**19.5**	**10.1**	**3.6**	**0.0**	**1.3**
	Non-Phase-In>100	**140**	**45.0**	**0.0**	**11.4**	**10.7**	**5.0**	**0.0**	**27.9**
Skin sensitisation (in vitro)	**Phase-In>1000**	**21**	**47.6**	**0.0**	**28.6**	**0.0**	**23.8**	**0.0**	**0.0**
	Phase-In 100 to 1000	**4**	**100.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
	Non-Phase-In>100	**3**	**0.0**	**0.0**	**0.0**	**0.0**	**66.6**	**0.0**	**33.3**
Skin sensitisation (in vivo)	**Phase-In>1000**	**3 754**	**55.4**	**0.0**	**20.8**	**7.0**	**13.7**	**0.5**	**2.6**
	Phase-In 100 to 1000	**488**	**58.0**	**0.0**	**24.4**	**0.0**	**14.8**	**0.6**	**2.3**
	Non-Phase-In>100	**176**	**41.5**	**0.0**	**15.3**	**19.9**	**2.3**	**0.0**	**21.0**
Genetic toxicity(in vitro)	**Phase-In>1000**	**10 322**	**57.2**	**0.0**	**22.0**	**3.8**	**12.1**	**0.1**	**4.8**
	Phase-In 100 to 1000	**1 745**	**64.6**	**0.0**	**17.7**	**3.0**	**11.8**	**0.0**	**2.9**
	Non-Phase-In>100	**351**	**51.3**	**0.0**	**10.3**	**9.1**	**2.8**	**0.3**	**26.2**
Genetic toxicity	**Phase-In>1000**	**3 533**	**52.4**	**0.5**	**24.8**	**6.3**	**11.0**	**0.0**	**5.0**

(in vivo)	**Phase-In 100 to 1000**	**596**	**61.4**	**0.3**	**21.5**	**4.4**	**10.1**	**0.0**	**2.3**
	Non-Phase-In>100	**94**	**50.0**	**0.0**	**5.3**	**7.4**	**1.1**	**0.0**	**36.2**
Toxicity to reproduction	**Phase-In>1000**	**3 535**	**31.7**	**4.2**	**23.8**	**25.6**	**12.1**	**0.1**	**2.5**
	Phase-In 100 to 1000	**487**	**30.0**	**1.8**	**24.2**	**28.3**	**9.7**	**0.0**	**6.0**
	Non-Phase-In>100	**156**	**26.3**	**4.5**	**7.1**	**41.0**	**3.8**	**0.0**	**17.3**
Developmental toxicity	**Phase-In>1000**	**4 217**	**42.3**	**3.6**	**29.7**	**10.9**	**10.7**	**0.2**	**2.6**
	Phase-In 100 to 1000	**589**	**44.1**	**5.8**	**29.5**	**12.1**	**5.4**	**0.3**	**2.7**
	Non-Phase-In>100	**121**	**29.8**	**10.7**	**9.9**	**33.1**	**3.3**	**0.0**	**13.2**
Toxicity to reproduction-other studies	**Phase-In>1000**	**390**	**75.1**	**0.0**	**9.5**	**5.6**	**2.1**	**0.0**	**7.7**
	Phase-In 100 to 1000	**41**	**53.7**	**0.0**	**4.9**	**24.2**	**17.1**	**0.0**	**0.0**
	Non-Phase-In>100	**3**	**66.7**	**0.0**	**0.0**	**33.3**	**0.0**	**0.0**	**0.0**
Carcinogenicity	**Phase-In>1000**	**3 559**	**38.7**	**0.1**	**27.9**	**14.9**	**12.2**	**0.2**	**6.1**
	Phase-In 100 to 1000	**451**	**56.3**	**0.2**	**22.2**	**13.1**	**5.8**	**0.0**	**2.4**
	Non-Phase-In>100	**29**	**13.8**	**0.0**	**24.1**	**48.3**	**0.0**	**0.0**	**13.8**
RDT all routes and duration	**Phase-In>1000**	**10 790**	**42.1**	**1.0**	**28.1**	**18.8**	**6.6**	**0.1**	**3.3**
	Phase-In 100 to 1000	**1 333**	**40.4**	**2.4**	**26.6**	**19.7**	**7.6**	**0.0**	**3.4**
	Non-Phase-In>100	**359**	**29.2**	**2.2**	**8.4**	**45.1**	**0.8**	**0.0**	**14.2**
Bioaccumulation(Fish)	**Phase-In>1000**	**798**	**42.1**	**1.5**	**24.7**	**0.0**	**25.6**	**3.1**	**3.0**
	Phase-In 100 to 1000	**278**	**21.2**	**1.8**	**37.1**	**0.0**	**38.5**	**0.0**	**1.1**
	Non-Phase-In>100	**20**	**70.0**	**0.0**	**15.0**	**0.0**	**5.0**	**0.0**	**10.0**
Short term toxicity to fish	**Phase-In>1000**	**6 942**	**52.6**	**0.0**	**20.2**	**1.8**	**14.2**	**2.1**	**9.1**
	Phase-In 100 to 1000	**1 405**	**48.7**	**0.0**	**27.3**	**0.9**	**16.2**	**1.3**	**5.7**
	Non-Phase-In>100	**143**	**53.1**	**0.0**	**8.4**	**4.2**	**4.2**	**2.1**	**28.0**
Long term toxicity to fish	**Phase-In>1000**	**3 281**	**27.4**	**0.8**	**21.2**	**33.9**	**9.0**	**4.3**	**3.3**
	Phase-In 100 to 1000	**812**	**35.5**	**1.2**	**34.7**	**17.1**	**8.3**	**1.2**	**2.0**
	Non-Phase-In>100	**101**	**13.9**	**0.0**	**3.0**	**65.3**	**5.9**	**2.0**	**9.9**
Long term toxicity to bird	**Phase-In>1000**	**2 007**	**10.8**	**0.2**	**6.4**	**72.7**	**9.9**	**0.0**	**0.0**
	Phase-In 100 to 1000	**350**	**16.3**	**0.0**	**10.3**	**41.4**	**28.9**	**0.0**	**3.1**
	Non-Phase-In>100	**36**	**0.0**	**0.0**	**0.0**	**91.7**	**2.8**	**0.0**	**5.6**
Toxicity to other above-ground organism	**Phase-In>1000**	**495**	**26.1**	**0.0**	**3.2**	**17.0**	**42.8**	**0.0**	**10.9**
	Phase-In 100 to 1000	**254**	**33.1**	**0.0**	**0.8**	**2.8**	**51.6**	**0.0**	**11.8**
	Non-Phase-In>100	**0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
Additional ecotoxicological information	**Phase-In>1000**	**644**	**37.9**	**0.0**	**22.2**	**1.1**	**4.2**	**0.5**	**34.2**
	Phase-In 100 to 1000	**129**	**14.0**	**0.0**	**20.2**	**0.8**	**5.4**	**2.3**	**57.4**
	Non-Phase-In>100	**12**	**16.7**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**83.3**

Appendix 2

Check list of 10 key questions proposed by JRC for risk assessors when evaluating QSAR models in the legal frame work.

No	Question	Interpretation
1	Is the predicted endpoint clearly defined?	If the endpoint is not clearly defined, the use of the prediction will be open to different interpretations, and thus of questionable value.
2	If the predicted endpoint is clearly defined ("yes" to Q1), does it represent a direct information requirement under the legislation of interest, or is it related to one of the information requirements?	If the predicted endpoint corresponds directly with an information requirement, it may be possible to use the prediction instead of experimental data. Alternatively, if the predicted endpoint is indirectly related to an information requirement, it may be useful as supporting information.
3	If the model is statistically based (as opposed to knowledge-based), is the model training set fully available?	If the model training set of a statistically-based model is not fully available (e.g. because the data are proprietary), it will be impossible for another practitioner to independently reproduce the model, which may reduce confidence in the model estimates. However, this may not be an issue if the model is coded into a software tool. This does not apply to knowledge-based models, which are based on human knowledge and do not have a clearly identified training set.
4	Is the method used to develop the model documented or referenced (e.g. in a scientific paper or QMRF)	If the details of model development are not documented, it will be impossible for another practitioner to independently develop and confirm the model, which may reduce confidence in the model estimates. Even if the method is documented, it will require a QSAR specialist to determine whether the documentation is sufficiently detailed to reproduce the model.
5	Is information available (in terms of statistical properties) concerning the performance of the model, including its goodness-of-fit, predictivity, robustness and error of prediction (uncertainty)?	The statistical properties of a model can provide evidence of its usefulness in a given context (e.g. need to minimise false negatives) and can also be used to assess whether the model has been overfitted (see question 7).
6	If the model is statistically based (as opposed to knowledge-based), does	The overfitting of statistically based models is undesirable because it can result in unpredictable errors. This consideration does not apply to

	examination of the available statistics indicate that the model may have been overfitted?	knowledge-based models. Overfitted statistical models typically show worse predictivity (outside their training sets) than their internal validation statistics imply. Several simple diagnostics exist, for example: a) the model estimation error (uncertainty of prediction) should not be significantly less than the known experimental error. b) the ratio of datapoints (chemicals) to variables (descriptors) should be at least 5:1.
7	Does the model training set contain the chemical of interest?	If the model training set contains the chemical of interest, then a prediction is not needed because some experimental data is available for direct use.
8	Does the model make reliable predictions for analogues of the chemical structure of interest?	The generation of reliable predictions for analogues of the chemical of interest increases confidence in the prediction. In the case of a software tool, it should be indicated whether the software automatically identifies analogues and their associated data within the model training set. In the case of a literature model, it should be considered whether suitable analogues can be identified in the training set (if available).
9	Is the model prediction substantiated with argumentation based on the applicability domain of the model?	Confidence in a prediction is increased if information is available concerning the applicability domain of the model, and thus whether the model is applicable to the chemical of interest. The applicability domain can include physicochemical and structural space, as well as mechanistic and metabolic considerations.
10	Can the model prediction be easily reproduced?	Not all model predictions can be easily reproduced, depending on the complexity and transparency of the model development process, and the availability of a user-friendly software tool implementing the model. If the model is a simple SAR (structural alert) it should be possible to apply it by visual inspection. However, some differences of expert interpretation may arise. If the model is a QSAR in the form of a transparent mathematical formula, it will be possible to apply it in a spreadsheet (e.g. Excel). If the model is implemented in the form of a freely or commercially available software tool, it is possible for different users to verify the same result (even if the model development process is not transparent), thereby increasing confidence in the prediction.