Project no.
022674 (SSPI)

CAESAR

Computer Assisted Evaluation of industrial chemical Substances According to Regulations

Instrument: SPECIFIC TARGETED RESEARCH OR INNOVATION PROJECT

Thematic Priority: 8.1 - RESEARCH TO SUPPORT POLICIES

Publishable final activity report

Period covered: *from 1st April 2008 to 31st March 2009*

Date of preparation: *14th May 2009*

Start date of project: *1st April 2006*

Duration: *36 months*

Project coordinator name:
Emilio Benfenati
Project coordinator organisation name:
Istituto di Ricerche Farmacologiche "Mario Negri"

Revision [FINAL]

Table of contest

1. Project execution

1.1. Summary

CAESAR developed and made freely available quantitative structure-activity relationship (QSAR) models for five endpoints, specifically designed for REACH. The endpoints are:

- Bioconcentration factor
- Skin sensitisation
- Mutagenicity
- Carcinogenicity
- Developmental toxicity (teratogenicity).

The results in prediction were superior or analogue to those obtained with other commercial or public models. Thus, new independent models are now available for toxicity and property predictions of industrial chemicals.

The models are based on experimental data. We used data of quality and number as high as possible, verifying individually all structures, in order to have a robust basis for the models. The models have been validated with external test sets and a series of statistical methods.

The models predict a continuous value or a category (such as toxic or not) depending on the endpoint and REACH requirements. The models are based on a series of fragments or molecular descriptors, and use several algorithms.

The models have been optimised to be reproducible, transparent, freely available, validated, tailored for REACH (input data of high quality, output data suitable for classification and labelling, prioritisation and risk assessment).

We activated a series of collaborations with regulators, industry and scientists, such as US Environmental Protection Agency, Federchimica, UE funded projects OSIRIS, CHEMPREDICT, CHEMOMENTUM, CASCADE.

1.2. The Consortium

The contractors of CAESAR project are summarized in the figure below.

Participants				
No.	**Name**	**Short name**	**Logo**	**Country**
1	Istituto di Ricerche Farmacologiche "Mario Negri" *	IRFMN *		Italy 
2	Central Science Laboratory	CSL DEFRA		U.K. 
3	BioChemics Consulting SAS	BCX		France 
4	Dipartimento di Elettronica e Informazione, Politecnico di Milano	POLIMI		Italy 
5	KnowledgeMiner Software	KM		Germany 
6	Liverpool John Moores University	LJMU		U.K. 
7	UFZ Centre for Environmental Research	UFZ		Germany 
8	National Institute of Chemistry	NIC-LJ		Slovenia 
9	TNO Quality of Life	TNO		The Netherlands 

* coordinator

Figure 1. The CAESAR Consortium.

In addition there are two subcontractors:

- Adriana Gómez Delgado, agdesigner - Industrial & Visual Designer
- Nick R. Price, Technology for Growth.

1.3. Introduction

REACH (Registration, Evaluation, Authorisation and Restriction of Chemical substances) is the new European Community on chemicals and their safe use. The aim of REACH is to protect the human health and the environment. To do this, it requires industry (manufacturers and importers) to provide detailed information (in function of the tonnage) about the chemicals they produce/import: physicochemical, toxicological and ecotoxicological information. It will not be possible to commercialise chemical compounds without these data (as the principle: no data, no market).

To supply all this information, the existing data are not enough. In order to reduce the increasing of animal tests, the REACH legislation establishes the use of alternative methods, like in-vitro or in-silico methods (see annexes: III, VI, VII-X). In this project in-silico models were developed to predict the biological activity of the chemicals for five endpoint chosen between the most “animal-consuming”.

The so-called “non-testing methods” may be subdivided into two categories: read-across and Quantitative Structure Activity Relationship (QSAR). The first one is based on the interpolation of the existing data of similar compounds. For the second one, the missing property (physicochemical, toxicological or ecotoxicological properties) is predicted on the basis of the chemical structure. Using non-testing methods it is possible to reduce the number of animal tests and the costs of the tests. Moreover, with QSAR it is possible to proceed in batch, testing many compounds at the same time.

1.4. State-of-the-art

QSAR methods are applied in the USA or Canada since decades for regulatory purposes, and specific tools have been developed there. In the last years, thanks to new techniques and advanced algorithms (such as multilinear regression, artificial neural networks, genetic algorithms, fuzzy logic, expert systems), and to new molecular descriptors, many new in-silico models have been developed, which in principle can be used to satisfy the requirements of REACH. However, models for regulatory purposes require special attention and check.

1. In Europe, recently, some projects addressed regulatory needs, like DEMETRA (http://www.demetra-tox.net/), which developed a battery of QSAR models for the pesticide legislation.
2. At this moment some EC projects about non-testing approaches are active or have been just finished. Among these there are: OSIRIS (that use read-across, chemical category, QSAR, in-vitro and in-chemico methods), CASCADE (that use in-silico approaches such as docking models), CHEMOMENTUM (which developed a new general platform and some specific models), CADASTER (which is addressing specific chemical classes, http://www.cascadenet.org/).

Furthermore, many freely and commercial tools are available. For instance, the OECD QSAR Toolbox provides data, similarity analysis, read-across and QSAR models, and EPI Suite (from the US EPA) offers a battery of models, widely used in USA. The Joint Research Centre, Ispra, made available a few predictive models. The commercial initiative Insilicofirst mades available many predictive models.

.Although the REACH legislation promotes the use of alternative test, there are still problems about the acceptability of the in-silico models. One of these is the data quality; to build a good model good experimental data are necessary, but only for some endpoints databases with reliable data exist. A second problem is the applicability domain of the models; each model had to establish for which chemical it can give good predictions.

In order to protect human health and environment, the REACH legislation requests many data about different endpoints. To gather the missing data it is possible to use alternative methods. The CAESAR project has the aim to cover the lack of data using in-silico models for five endpoints. These have been chosen using two criteria:

- addressing those endpoints that need the highest number of animal testing (in vivo testing)
- checking the availability of high quality experimental data to be used as training set for building the predictive models.

For this purpose we used the list of endpoints requiring *In Vivo* experiments within REACH compiled by Van der Jagt et al, 2004 and ordered by the number of animals used in the toxicity experiments. Table 1 reports the endpoints sorted by the number of animal tests required. The coloured arrows indicate the chosen endpoints.

Table 1. List of “animal-consuming” endpoints for REACH and chosen endpoints.

Endpoints					
1	Two generation reprotoxicity		10	Sub-chronic toxicity	
2	Developmental toxicity studies		11	Long-term repeated toxicity	
3	Further Mutagenicity study		12	Short-term fish toxicity	
4	Skin sensitisation		13	Acute inhalation toxicity	
5	Long-term fish toxicity		14	Acute dermal toxicity	
6	Developmental toxicity screening		15	Acute oral toxicity	
7	Accumulation		16	In-vivo eye irritation	
8	Short-term repeated dose		17	In-vivo skin irritation	
9	Carcinogenicity		18	Lon-term bird toxicity	

Below we describe the chosen endpoints and their use within RACH.

Accumulation - Bioconcentration

Three processes are related to the accumulation of a chemical into the living organism:

- Bioconcentration, which is the process where the chemical concentration in an organism achieves a level that exceeds that in the environment as a result of the exposure of an organism to a chemical in the water; this process does not include exposure via the diet. Bioconcentration refers to a situation, typically derived under controlled laboratory conditions using fish, wherein the chemical is absorbed from the water via the respiratory surface and/or the skin only. The extent of chemical Bioconcentration is usually expressed in the form of a Bioconcentration factor (BCF).
- Biomagnification, which is the process by which lipid normalized chemical concentrations increase with trophic level in a food-chain.
- Bioaccumulation, defined as the process where the chemical concentration in an organism achieves a level that exceeds that in the environment as a result of chemical uptake through all routes of chemical exposure (e.g. dietary absorption, transport across the respiratory surface, dermal absorption). It is the combination of bioconcentration and biomagnifications. The extent of chemical Bioaccumulation is usually expressed in the form of a Bioaccumulation factor (BAF).

Thus, both the BCF and the BAF are described as the ratio between the concentration of the test substance in/on the fish (the most studied species or another specified organism) and the concentration of the same chemical in the surrounding medium (typically water, but it can be related to other medium) at steady state conditions (that means that the organism is exposed for a sufficient length of time so that the ratio does not change substantially). BCF and BAF are described by the following formulas:

$$\textbf{BCF} = CB/CWD = k1/(k2 + kE + kM + kG)$$

$$\textbf{BAF} = CB/CWD = \{k1 + kD\ (CB/CWD)\} / (k2 + kE + kM + kG)$$

where CB is the chemical concentration in the organism (g/kg−1), k1 is the chemical uptake rate constant from the water at the respiratory surface (L·kg−1·d−1), CWD is the freely dissolved chemical concentration in the water (g·L−1), kD is the uptake rate constant for chemical in the diet (kg*kg-1*d-1) and k2, kE, kM, kG are rate constants (d−1) representing chemical elimination from the organism via the respiratory surface, fecal egestion, metabolic biotransformation, and growth dilution, respectively.

Bioconcentration factors (BCFs) are used to relate pollutant residues in aquatic organisms to the pollutant concentration in ambient waters. Many chemical compounds, especially those with a hydrophobic component, partition easily into the lipids and lipid membranes of organisms and bioaccumulate.

The REACH legislation requests different information in function of the tonnage of production and/or import (see Point 9.3.2 of the REACH - Bioaccumulation in aquatic species, preferably fish). The preferred experimental conditions for BCF test are those reported in the OECD 305 guideline. The number of likely fish recommended for the test is in the range 132 to 240, for a duration of 44-116 days and a cost for each experiment in the range of 50-100 k€.

According to the REACH framework the potential use of BCF information includes the following:

- Classification & Labelling (C&L): all substances should be assessed for environmental hazard classification. Bioaccumulation potential is one aspect that needs to be considered in relation to long-term effects.
- Prioritization (PBT, vPvB): bioaccumulation is one of the criteria used for the PBT/vPvB assessment. For a definitive conclusion, reliable measured BCF data are generally necessary (for fish or an invertebrate such as molluscs). However, a provisional assessment can be made against screening criteria.
  To define if a chemical is PBT or vPvB the thresholds are: for B BCF > 2000 L/kg (whole organism weight) = 3.3 in Log unit; vB BCF > 5000 L/kg = 3.7 in Log unit.
- Chemical Safety Assessment (CSA): fish BCF and BMF (Biomagnifications Factor) values are used to calculate concentrations in fish as part of the secondary poisoning assessment for wildlife, as well as for human dietary exposure. An invertebrate BCF may also be used to model a food chain based on consumption of sediment worms or shellfish. An assessment of secondary poisoning or human exposure via the environment will not always be necessary for every substance. In the first instance, a predicted BCF may be used for first tier risk assessment.

As reported in Table 2 the potential use of BCF information in REACH satisfies C&L, B assessment and CSA (up to 100 ton/year), while for production above 100 ton/year a specific continuous value becomes necessary.

Table 2. Scheme of the requests of REACH for BCF.

Tonn/year	C&L	B and vB	CSA	BCF value
> 1	X	X		
> 10	X	X	X	
> 100	X	X	X	X

Thus, both quantitative and qualitative (classification) evaluation might be requested.

The PBT and vPvB assessment of a substance shall be based on all relevant information available, which is normally the information submitted as part of the technical dossier, including the physicochemical, hazard and exposure information generated in the context of the CSA.

Other properties or estimations particularly in relation to Log Kow may be used to infer bioconcentration properties of chemical compounds; thus, beside the general stimulus to QSAR in the REACH legislation, some QSAR-based estimation methods are already mentioned for deducing BCF properties.

Moreover, to properly assess the reliability of a QSAR model prediction a very useful criterion for comparison is represented by the experimental data variability. In the case of BCF the reported experimental uncertainty

can be up to 0.75 in Log unit (Dimitrov *et al*., 2005). In other databases we assessed within CAESAR, the observed experimental variability was in the range of 0.3-0.4 Log units.

Skin sensitisation

A sensitizer is an agent that is able to cause an allergic response in susceptible individuals. The consequence of this is that following subsequent exposure via the skin the characteristic adverse health effects of allergic contact dermatitis or atopic dermatitis may be provoked. After inhalation exposure, adverse health effects include asthma (and related respiratory symptoms such as rhinitis) or extrinsic allergic alveolitis.

Dermal allergy is based on an immunological mechanism. Protein binding, metabolic activation (if required), internalization and processing by Langerhans cells (LC), transport of antigen by LC to draining lymph nodes, and presentation to and recognition by T lymphocytes are gaining access among the key steps required for a chemical to induce sensitization via skin contact to the viable epidermis.

The skin sensitisation data potential for substances produced or imported in quantities of 1 tons/year should be evaluated.

In vivo testing does not need to be conducted if:

- the available information indicates that the substance should be classified for skin sensitisation or corrosivity; or
- the substance is a strong acid (pH<2.0) or base (pH>11.5); or
- the substance is flammable in air at room temperature.

The golden standard test under REACH for sensitization is the Murine Local Lymph Node Assay (LLNA) (OECD 429), which is the first-choice method for in vivo testing. The guideline recommends the use of 25 animals (mice) per assay. Only in exceptional circumstances another test should be used.

Sensitization potential may be identified from epidemiological studies, from animal experiments and/or other appropriate means that may include (Quantitative) Structure-Activity Relationships ((Q)SAR) analyses and/or extrapolation from structurally similar substances (read-across).

Under REACH, usually a quantitative hazard assessment, resulting in classification of the substance considered as a sensitizer or not, is sufficient. Based on the LLNA test, quantitative hazard assessment of induction of sensitisation is possible and may be required, but there is no clear guidance yet on the procedure.

When hazard identification is the primary objective and no potency data are required, a reduced LLNA test, using only one (high) dose group and a negative control group, may be sufficient. If an OECD 406 study (Guinea Pig Maximisation Test or Buehler test) is available, it may satisfy the data requirement as well, with the same restriction as for the reduced LLNA test, as it does not allow potency estimates.

Mutagenicity

Chemical substances, mixtures of substances and physical agents (e.g., radiation) can induce alterations in the genome of either somatic or germinal cells. There are several mutagenic endpoints of concern; these include point mutations (i.e., submicroscopic changes in the base sequence of DNA) and structural or numerical chromosome aberrations. Structural aberrations include deficiencies, duplications, insertions, inversions, and translocations, whereas numerical aberrations are gains or losses of whole chromosomes (e.g., trisomy, monosomy) or sets of chromosomes (haploidy, polyploidy).

Certain mutagens, such as alkylating agents, can directly induce alterations in the DNA. Mutagenic effects may also come about through mechanisms other than chemical alterations of DNA. Among these there are: interference with normal DNA synthesis (as caused by some metal mutagens), interference with DNA repair, abnormal DNA methylation, abnormal nuclear division processes, or lesions in non-DNA targets. Evidence that an agent induces heritable mutations in human beings could be derived from epidemiologic data indicating a strong association between chemical exposure and heritable effects. It is difficult to obtain such data because any specific mutation is a rare event, and only a small fraction of the estimated thousands of human genes and conditions are currently useful as markers in estimating mutation rates.

Mutagenicity studies are required for all the tonnage bands. Unlike most other endpoints, a negative mutagenicity result in vitro can be considered sufficient evidence for non-mutagenic potential but positive results must be confirmed in vivo. At 1-10 tonnage level (Annex VII) the in vitro gene mutation study in bacteria (Ames test) is required. At Annex VIII level (10-100 ton), two additional in vitro studies are required: a cytogenicity study and a gene mutation study in mammalian cells. If there is any positive result within these in vitro tests, in vivo mutagenicity studies are required. At higher tonnage in vivo studies are needed.

Carcinogenicity

The process of carcinogenesis involves the transition of normal cells into cancer cells via a sequence of stages that entail both genetic alterations (i.e. mutations) and non-genetic events. Genotoxic modes of action involve genetic alterations caused by the chemical interacting directly with DNA to result in a change in the primary sequence of DNA.

Non-genotoxic modes of action include epigenetic changes, i.e., effects that do not involve alterations in DNA but that may influence gene expression, altered cell-cell communication, or other factors involved in the carcinogenic process.

Chemicals are defined as carcinogenic if they induce tumours, increase tumour incidence and/or malignancy or shorten the time to tumour occurrence. Chemicals can induce cancer by any route of exposure (e.g., when inhaled, ingested, applied to the skin or injected), and carcinogenic potential and potency may depend on the conditions of exposure (e.g., route, level, pattern and duration of exposure). Carcinogenic chemicals have conventionally been divided into two categories according to the presumed mode of action: genotoxic or non-genotoxic.

For the endpoint of carcinogenicity, standard information requirements are specifically described for substances produced or imported in quantities of ≥1000 tons/year. The precise information requirements will

differ from substance to substance, according to the toxicity information already available and details of use and human exposure for the substance in question.

A carcinogenicity study may be proposed by the registrant or may be required by the Agency if:

- the substance has a widespread dispersive use or there is evidence of frequent or long-term human exposure; and
- the substance is classified as mutagen category 3 or there is evidence from the repeated dose study(ies) that the substance is able to induce hyperplasia and/or pre-neoplastic lesions.

If the substance is classified as mutagen category 1 or 2, the default presumption would be that a genotoxic mechanism for carcinogenicity is likely. In these cases, a carcinogenicity test will normally not be required.

The objective of investigating the carcinogenicity of chemicals is to identify potential human carcinogens, their mode(s) of action, and their potency. The golden standard test under REACH for carcinogenicity is OECD 451. OECD guideline no. 451 recommends the use of 400 animals (rats and mice) for the studies.

Carcinogenic potential may be identified from epidemiological studies, from animal experiments and/or other appropriate means that may include (Quantitative) Structure-Activity Relationships ((Q)SAR) analyses and/or extrapolation from structurally similar substances (read-across).

Once a chemical has been identified as a carcinogen, there is a need to elucidate the underlying mode of action (genotoxic or non-genotoxic). Human studies are generally not available for making a distinction between the modes of action. A conclusion on this depends on the outcome of mutagenicity/genotoxicity testing and other mechanistic studies. In addition to this, animal studies (e.g. the carcinogenicity study, repeated dose studies, and experimental studies with initiation-promotion protocols) may also inform on the underlying mode of carcinogenic action.

For genotoxic carcinogens exhibiting direct interaction with DNA, it is not generally possible to infer the position of the threshold from the no-observed-effect level on a dose-response curve, even though a biological threshold below which cancer is not induced may exist. For non-genotoxic carcinogens, no-effect-thresholds are assumed to exist and to be discernable.

Developmental toxicity

Developmental toxicity includes any effect interfering with normal development, both before and after birth. It includes effects induced or manifested either pre- or postnatally. This includes embryotoxic/foetotoxic effects such as reduced body weight, growth and developmental retardation, organ toxicity, death, abortion, structural defects (teratogenic effects), functional effects, peri- and postnatal defects, and impaired postnatal mental or physical development up to and including normal pubertal development.

The standard data requirements for developmental toxicity under the REACH Regulations are as follows:

- A reproduction/developmental toxicity screening test (OECD TGs 421 or 422), usually required for substances produced or imported in quantities between 10 and 100 tons/year (according to Annex VIII of the REACH legislation).

- A prenatal developmental toxicity study (EU B.31, OECD TG 414) in one species, usually required for substances produced or imported in quantities of ≥100 tons/year (according to Annex IX and X of the REACH legislation). A study in a second species may be considered necessary at these levels of production/importation.

However, these tonnage-related standard data requirements may be adapted, either reduced (data waiving), deferred or extended. Factors that can influence the testing requirements include structural relationships with other chemicals, the results of other toxicity studies, presence of mutagenic and carcinogenic properties, available data from humans exposed to the substance, concerns for endocrine disruption and the patterns of use and human exposure.

The golden standard test under REACH for developmental toxicity is OECD 414, which recommends the use of 80 adult animals (rats or rabbits) for this study. Studies according to OECD 421 or 422 are used for screening only. If positive, they may not be considered sufficient for classification & labeling and quantitative risk assessment. If negative for a substance produced or imported in quantities between 10 and 100 tons/year, no further testing is needed. A negative outcome of these screening studies for substances produced/imported in quantities over 100 tons/year implies an OECD 414 study needs to be performed.

Developmental toxicity may be identified from epidemiological studies, from animal experiments and/or other appropriate means that may include (Quantitative) Structure-Activity Relationships ((Q)SAR) analyses and/or extrapolation from structurally similar substances (read-across). Developmental effects should be considered in relation to adverse effects occurring in the parents. Since adverse effects in pregnancy or postnatally may result as a secondary consequence of maternal toxicity, reduced food or water intake, maternal stress, lack of maternal care, specific dietary deficiencies, poor animal husbandry, intercurrent infections etc., it is important that the effects observed should be interpreted in conjunction with possible concomitant maternal toxicity. The nature, severity and dose-response of all effects observed in progeny and parental animals should be considered and compared together to achieve a balanced integrated assessment of available data on all endpoints relevant for developmental toxicity.

1.5. Project objectives

One of the principal aims of the CAESAR project is to build models to predict toxicity of chemical substances for REACH. A second goal is to reduce the animal test and, in consequence, the cost of these tests. The third objective is to obtain a wide applicability and acceptability of these models by stakeholders for decision support.

In order to identify the endpoints which may in principle get the highest benefit from the use of QSAR models, we started from the document from the European Commission, Joint Research Centre, Institute for Health and Consumer Protection entitled “Assessment of additional testing needs under REACH. Effects of (Q)SARs, risk based testing and voluntary industry initiatives” (Pedersen et al, 2003), and its addendum (Van der Jagt et al, 2004. This document was used to choose the five endpoints for which develop a new QSAR models. Each model, accompanied by a protocol reporting information of quality control step and a

user manual, has been validated by a second partner taking into account particularly accuracy, sensitivity and applicability of the model.

The work and the results are reported on a freely accessible website (http://www.caesar-project.eu) developed to incorporate the models and presented into a dedicated workshop.

Moreover a network of stakeholders has been created to correctly use, promote, disseminate and exploit CAESAR activities.

1.6. Activity

1.6.1. Methodology and approaches

1.6.1.1. Sources of experimental data

1.6.1.1.1. Accumulation - BCF

The CAESAR model for BCF is based on a dataset including 473compounds with experimentally determined BCF values extracted from the publication of Dimitrov et al., 2005. The original Dimitrov dataset contains 511 compounds with biological values of high quality. Indeed, values have been obtained in according to official guidelines. Furthermore, for this, as for all the dataset, at least two partners checked the chemical structures, to identify possible errors or critical issues, such as mixtures, not clear definition of substituents likely transformation during the experiment (Zhao et al., 2008). For BCF, the final set of compounds was 473.

1.6.1.1.2. Skin sensitization

To address the skin sensitization issue two complementary approaches have been investigated:

- A global approach aimed at developing a classifier to discriminate sensitizers vs. non sensitizers
- A local approach that investigated a mechanistic based category formation coupled with a read-across approach within each category

This CAESAR model for skin sensitization is based on a data set that includes 209 compounds extracted from the paper by Gerberick et al., 2005 Chemical were checked as explained in 1.2.1.1.1. For developing classification models, the chemicals of this data set were labelled in two classes, sensitizer (S) and non-sensitizers (N), which gave a good distribution of the numbers of compounds in each class. The class S merges the first four ranges established by ECETOC: Extreme (EC3<0.1%), Strong (0.1%<1%), and Moderate (1%<10%) and Weak (EC3>10%) ranges; the class N groups all compounds belonging to the non-sensitizers (see Table 3). The data set was subdivided into a training set (about 80% of the compounds)

Table 3. - LLNA classification

EC3 (%)	LLNA Class
NC	Non Sensitiser
≥10	Weak Sensitiser
1-10	Moderate Sensitiser
0.1-1	Strong Sensitiser
<0.1	Extreme Sensitiser

used to develop the model and a test set (including the remaining 20% of the data) used to assess the performance of the model in prediction.

1.6.1.1.3. Mutagenicity

The CAESAR model for mutagenicity is based on a data set that includes 4225 compounds. As described for BCF, at least two partners checked all structures.

1.6.1.1.4. Carcinogenicity

To address the carcinogenicity issue two complementary approaches were investigated:

- a regression approach
- a continuous approach

The original dataset used for the development of CAESAR model for carcinogenicity contains 806 chemicals taken from CPDBAS with associated TD50 values for rat. For one compound we found an unusually high value, as compared with data on literature for this and related compounds. Since we have not been able to solve the conflicting data between the CPDBAS and the literature, we eliminated this compound from the data set, obtaining the 805 chemicals. As described for BCF, at least two partners checked all structures.

In classification any compound with a finite TD50 dose was associated with the toxic class, while non positive compounds were assigned to the non toxic class. In regression, only compounds with a TD50 dose were used.

1.6.1.1.5. Developmental toxicology

The dataset for developmental toxicity has been taken from Arena et al., 2004 and includes 292 compounds.

Chemical compounds were categorized into toxicant or non toxicant according to FDA risk factors (see Table 4). Thus, for some compounds a different classification has been obtained, compared to the data as in Arena et al., 2004.

1.6.1.1.6. Chemical structures and descriptors

QSAR models, predict the activity of a compound based on its chemical structure. Thus the use of a correct chemical structure is fundamental. To prune the original datasets form the compounds with a structure not

well identified or not in agreement with the structures used in the QSAR models (such as isomer mixture or salt), it is necessary to pre-process the chemical structure of all the compounds in the datasets. Thus, as anticipated above, we checked chemical structures individually using the five data sources in Box 1. The check was repeated by a second partner, and this was done for all chemicals for all datasets. This time-consuming activity (about 7000 chemicals have been checked) is necessary to achieve data sets of high quality, for the particular delicate application: models for regulatory purposes.

Table 4. - CAESAR binary classification on the basis of the FDA classes.

FDA classes	**Definition**	**CAESAR Binary class**
Category A	Negative human studies	Non developmental toxicant
Category B	Negative animal studies & No human studies executed OR Positive animal studies & Negative human studies	
Category C	Postive animal studies & No human studies executed OR No studies at all	Developmental toxicant
Category D	Postive human studies	
Category X	Animal OR human studies show abnormalities AND/OR Evidence of foetal risk based on human experience	

To build the model and to test their performance we used two different series of compounds: respectively a training and a test set. For this reason the compounds were split into a training set (of about the 80% of the compounds) and a test set (with the remaining 20%). Table 5 shows the splitting of the five databases.

Table 5 - splitting of the five databases.

DATABASE	**TOTAL (n)**	**TRAINIG SET (n)**	**%**	**TEST SET (n)**	**%**
BCF	473	387	80	95	20
Skin sensitization	209	167	80	42	20
Mutagenicity	4225	3380	80	845	20
Carcinogenicity	805	644	80	161	20
Developmental toxicity	292	234	80	58	20

To represent the chemical structure of a compound we used chemical descriptors. Only 2-dimensional descriptors were finally used in this project for their simplicity and because no improvement in model results was found when using 3D-descriptors, alone or combined with 2D-descriptors. Indeed we also calculated 3D-descriptors and developed models with them. 2D-descriptors are numerical values insensitive to a change in the molecular conformation, which occurs when the spatial arrangements of atoms in the molecule are modified through the free rotation of the atoms about single chemical bonds.

Box 1

The chemical information of these compounds were checked, using molecule name and/or CAS number, at five online databases:

- ChemFinder (http://chemfinder.cambridgesoft.com),
- ChemIDPlus (http://chem.sis.nlm.nih.gov/chemidplus/),
- Safe Nite (http://www.safe.nite.go.jp/english/kizon/KIZON_start_hazkizon.html),
- Biodegradability Database and Estimation (http://qsar.cerij.or.jp/cgi-bin/QSAR/index)
- PubChem Compound (http://www.ncbi.nlm.nih.gov/sites/).

Some compounds were omitted according to the following criteria:

- too little information to find the structure (i.e. lack of information to identify the structure, compound not found…);
- duplicates (compounds repeated in the original paper);
- mixtures (isomeric mixture);
- diastereo-isomers (CIS-TRANS mixture);
- metal complexes;
- salts (it was taken into consideration only the neutral form).

There are two main advantages in using 2D-descriptors in this kind of application (model for regulatory purposes):

1. They do not require optimization of the conformation, which most typically is done manually, or using stochastic approaches, and thus they produce more reproducible results than using 3D-descriptors, and reproducibility is a highly desirable feature in case of models for regulatory purposes
2. They are faster to be calculated.

Several main classes of bi-dimensional descriptors have been exploited for this project. These classes, together with the software package used for computing them in CAESAR, are summarized in the Table 6. More details can be found in the box 2.

Box 2

Constitutional descriptors can be defined as descriptors reflecting the molecular composition of a compound without connectivity and geometry information. Among these there are:

- total number of atoms in the molecule
- absolute and relative numbers of atoms of certain chemical identity (C, H, O, N, F, etc.) in the molecule
- absolute and relative numbers of certain chemical groups and functionalities in the molecule
- total number of bonds in the molecule
- absolute and relative numbers of single, double, triple, aromatic or other bonds in the molecule
- total number of rings, number of rings divided by the total number of atoms
- total and relative number of 6-atoms aromatic rings
- molecular weight and average atomic weight

They introduce a key argument in building QSAR models, the definition of molecular aromatic character.

Topological descriptors are descriptors reflecting the molecular connectivity without geometry information. The main typologies are:

- molecular connectivity indices
- molecular valence indices
- information content indices
- molecular shape indices
- topological electronic indices
- weighted distance matrix indices
- sum of distances between different elements
- indices based on distance/detour paths

Physico-chemical descriptors refer to physico-chemical molecular properties, (such as the boiling, melting points, water solubility, lipophilicity, solvation free energy, etc.). Among them, the octanol-water partition coefficient, LogP, when given as logarithm, has been the attribute used by most investigators to correlate structure and toxic effects for ecotoxicological endpoints. The octanol-water partition coefficient is a measure of the hydrophobicity and hydrophilicity of a substance. Cell membranes are composed of phospholipids, which have hydrophobic tails that produce a very hydrophobic environment in the middle of the membrane bilayer. In the absence of active membrane transport, more hydrophobic compounds have an easier time getting through a membrane. Compounds hydrophobicity plays a fundamental role in the modulation of many key ADME (absorption, distribution, metabolism, and excretion) processes, such as drug-membrane interactions, drug transport, biotransformation, distribution, accumulation, protein and receptor binding. For this reason, LogP has been regarded as one of the most important properties in QSAR studies. Despite its efficacy, however, it is important to mention that LogP cannot describe chemical features such as the compound reactivity, meaning the possibility for a substance to dissociate or react, giving a number of by-products, usually referred as metabolites. For this reason, in the present project, LogP is used in combination with other descriptors accounting for chemical reactivity, such as HOMO-LUMO ones.

Table 6. Classes of 2D descriptors used for modeling purposes with the software package used for computing them.

2D-Descriptor Class	Software
Constitutional	Codessa, Dragon, MDL QSAR
topological	Codessa, Dragon, MDL QSAR
Physico-chemical	ACD, Pallas, KowWin

CAESAR calculated thousands of chemicals descriptors and fragments, and developed hundreds of models. However, we selected a few models which resulted preferable. The user of the CAESAR model will not be requested to calculate all these descriptors and to run complicated algorithm. The final models use a small series of chemical and the algorithms have been translated and fixed with a much simpler series of algorithms. In other words, while in the modelling developmental phase a large series of simple and complex tools have been used, the final models incorporate the optimal solutions into simpler tools. More than one model for the same endpoint gave good results. However, we implemented one, also considering the convenience for the practical process. Thus, for instance, we preferred the use of descriptors of easier implementation for us, such as those from Dragon or from EPA.

1.6.1.2. Modelling approaches

Accumulation - BCF

Heuristic (HM) and genetic algorithm (GA) methods were used to select optimal descriptors. The software CODESSA version 2.21 was used, to give a complete search for the best multilinear correlations in the ordinary least squares regression (OLS) method. MobyDigs version 1.0 (http://www.talete.mi.it) was used for genetic algorithm-variable subset selection (GA-VSS).

Multiple linear regression (MLR) was used to develop the linear model and radial basis function neural network was used to build the models. We used R code for support vector machines (SVM) (Burges, 1998) models (http://www.R-project.org). SVM algorithms yield prediction functions that are expanded on a subset of training vectors, or support vectors.

Finally, the final hybrid model was obtained using a hybrid model approach: the outputs of the individual models were used as input of the hybrid model and combined into a simple linear equation:

$$logBCF = k_n[Min, Mean, Max(value\ given\ by\ models\ to\ combine)] + a_n$$

where n is the number of areas chosen to build the hybrid models, three in our case.

Skin sensitisation

The CAESAR model for skin sensitization was developed using Adaptive Fuzzy Partition (AFP). AFP was used to develop classification models implementing a fuzzy partition algorithm. It models relations between molecular descriptors and chemical activities by dynamically dividing the descriptor space into a set of fuzzy partitioned subspaces. The aim of this algorithm is to select the descriptor and the cut position that allow to

get the maximal difference between the two fuzzy rule scores generated by the new subspaces. The score is determined by the weighted average of the chemical activity values in an active subspace A and in its neighbouring subspaces.

This model uses 8 descriptors from Dragon software (nN; GNar; MDDD; X2v; EEig10r; GGI8; nCconj; O-058). For the feature selection of the chemical descriptors, a hybrid selection algorithm (HSA), which combines the genetic algorithm (GA) concepts and a stepwise approach, was used to select the best descriptors for classifying developmental toxicity dataset.

A complementary approach was developed to enable potential mechanisms of toxic action to be assigned to chemicals thought to be capable of skin sensitisation. This approach was based on the work of Aptula et al., 2005 who had previously suggested that for a chemical to be a skin sensitizer it must be capable (either directly or after some abiotic or metabolic transformation) of one of five electrophilic-nucleophilic reactions. The approach undertaken was to devise SMARTS patterns capable of identifying the electrophilic mechanisms previously assigned to the 209 Gerberick LLNA dataset. The 44 chemical TIMES-SS LLNA data, in which each chemical has also had a mechanism of action assigned to it by the same authors was then used to validate the applicability of the SMARTS patterns.

The ability to assign a chemical to one of the suggested electrophilic mechanisms allowed the development of a novel quantitative read-across approach for the prediction of skin sensitising potential of chemicals assigned to the Michael mechanistic domain (see figure 2).

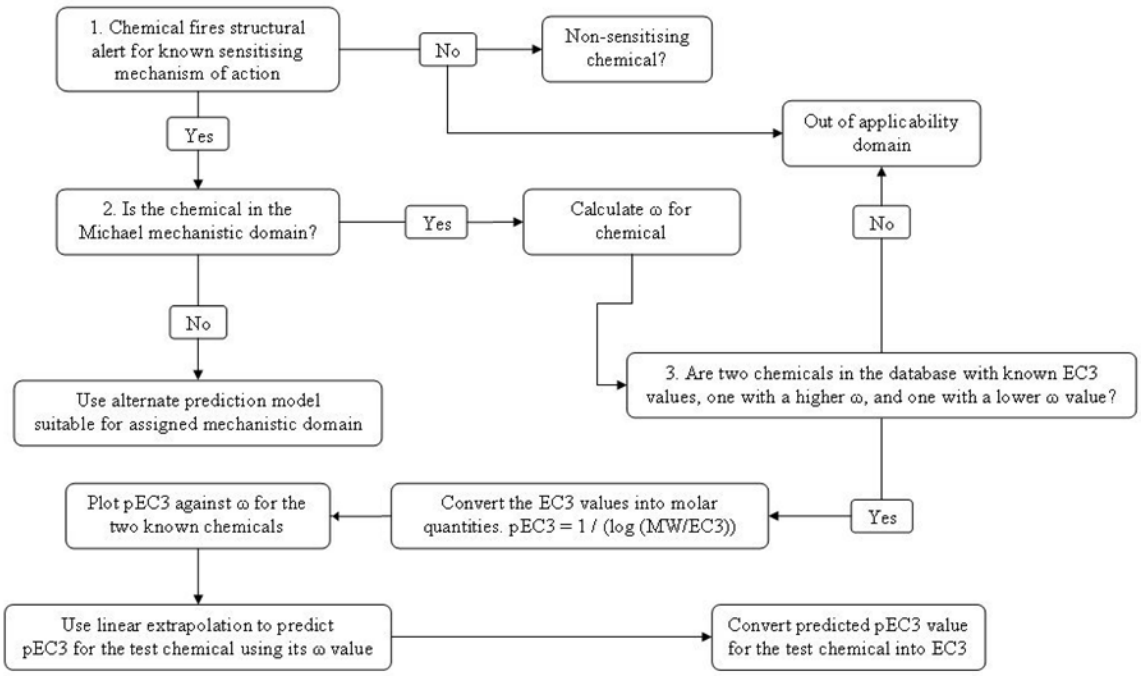


Figure 2. - Read across procedure utilising the electrophilic index (ω) as a quantitative measure of a chemical's electrophilicity

This approach utilised the electrophilic index (ω) which has been suggested to account for a chemical's inherent electrophilicity. This descriptor was calculated from the frontier molecular orbitals, these are the orbitals considered to be directly involved in the formation of the covalent bond between the nucleophilic sulphur containing protein side chains and the electrophilic chemical thought to be the initial event required for skin sensitisation.

Mutagenicity

For regulatory purposes, an integrated model was arranged cascading two complementary techniques: a machine learning algorithm (SVM), to build an early model with the best statistical accuracy, followed by an expert tool for False Negative (FN) removal, based on known structural alerts; this allowed refining predictions given by the first module (SVM).

The expert filter, to be applied only on compounds presumed safe by SVM, wraps two sets of structural alerts (selected from the Benigni/Bossa rulebase, Benigni et al., 2008) with different distinguishing features: the first has the aim to enhance the prediction accuracy attempting a precise identification of misclassified FN; the second one goes on with the FN removal as much as this does not noticeably downgrade the original prediction accuracy (by generating too many FP as well). To point out this distinction, compounds picked out by the first checkpoint are classified as 'mutagenic', and those picked out by the second one are classified as 'suspicious'. Unaffected compounds are finally classified as 'non-mutagenic'. Figure 3 shows the overall scheme.

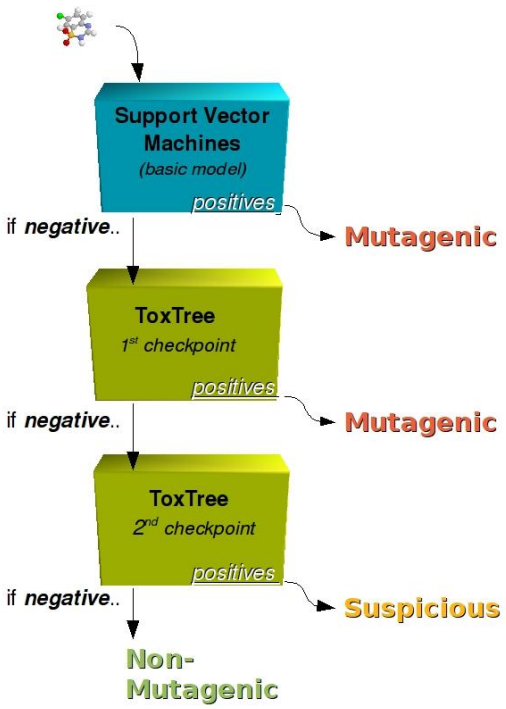


Figure 3. - The combined approach for Mutagenicity.

Carcinogenicity

Regression Model

In collaboration with the EC project CHEMPREDICT we developed quantitative structure - activity relationships (QSAR) models based on SMILES. Simplified molecular input line entry system (SMILES) has been used as elucidation of the molecular structure for QSAR to predict carcinogenicity. Using the Monte Carlo method we constructed optimal descriptors, which are a mathematical function of composition of the SMILES elements together with special codes of cycles present in molecules.

Classification Model

A classification model has been developed adopting the *Counter-Propagation Artificial Neural Network* (CP-ANN) method and a set of MDL Chemical Descriptors.

Developmental toxicology

The chemical descriptors have been calculated in collaboration with Dr Todd Martin, US EPA with a public US EPA program.

Several models have been developed: in one case (model A) the software used for the model is WEKA (Waikato Environment for Knowledge Analysis, http://www.cs.waikato.ac.nz/ml/weka/), an open source workbench. In this case we used 13 chemical descriptors.

The algorithm used for modeling is Random Forest that constructs a "forest" of random "trees". A “tree” is a sequence of rules, which uses chemical descriptors to split the chemicals into toxic or not.

A second model (model B) was developed using Adaptive Fuzzy Partition (AFP) – AFP was used to develop classification models implementing a fuzzy partition algorithm. It models relations between molecular descriptors and chemical activities by dynamically dividing the descriptor space into a set of fuzzy partitioned subspaces. The aim of this algorithm is to select the descriptor and the cut position that allow to get the maximal difference between the two fuzzy rule scores generated by the new subspaces. The score is determined by the weighted average of the chemical activity values in an active subspace A and in its neighboring subspaces.

In this case we used 6 chemical descriptors. For the feature selection, a hybrid selection algorithm (HSA), which combines the genetic algorithm (GA) concepts and a stepwise regression, was used to select the best descriptors for classifying developmental toxicity dataset.

Implementation of the models

As an activity outside the Description of work, we decided to implement as many as possible QSAR models, as developed within CAESAR, and at least one per endpoint. The models (which means the algorithms which calculate the property, based on chemical descriptors) developed within CAESAR are freely available, and as soon as the papers describing the models will be published, they be also widely disseminated. For instance, the code of the BCF model has been published in the open literature, and thus can be seen and used by anybody. The use of the CAESAR models requires, as anticipate in the contract, the calculation of chemical descriptors, which, when we started the project, were calculated using commercial software. This is still by far the most used practice also today, but recently a number of freely available software appeared,

which can calculate chemical descriptors. The availability of free software for chemical descriptors will facilitate the use of QSAR models, not only for the reason of the cost, but also because new solutions can be implemented. Indeed, the use of commercial software for chemical descriptors means that if the software changes and this occurs quite frequently, the previously developed model, based on the previous descriptors, may not work. This of course applies with any kind of software, also publicly available, but the possibility to solve this issue is facilitated if the software is open source. As an alternative, special agreement has to be done with the developer of the software.

CAESAR, also in this, showed a pioneer approach, anticipating the need in a pro-active way. In the last months of the project, when we were almost sure that models were good, we started a new activity for the full implementation of the CAESAR models.

This activity required the development of special algorithm for the descriptor calculation of some chemical descriptors. In other cases, we explored the possibility of agreements or collaborations with developers of software for chemical descriptor calculation. Thus, an agreement was achieved with the developers of Dragon, which is used for some models. In another case, we started a collaboration with the US EPA (see below, 1.6.3).

In this way, it has been possible to develop a special web site, which can be used to calculate the property of interest starting from the simple chemical structure, or, in some cases, introducing the list of the chemical descriptor values. However, as a further additional activity, outside CAESAR, we will implement soon all models in a way that the user will only need the chemical structure as starting point. A special software calculate the chemical descriptors in the background, and provide the predicted property (see below, 1.6.5).

Tools for similarity and applicability domain

The safe use of QSAR models for regulatory purposes may take advantage of the availability of tools to assess if a QSAR model can be used for the particular chemical of interest. Indeed, the QSAR model for regulatory purposes as to be a good model on “technical” point of view (general condition for the quality of the model), has to be good for the specific application for REACH (and thus has to be optimised for that purpose), but this is not yet enough, because this refers to a general assessment of the model. But the user is interested in a single compound, or anyhow in a series of chemicals. Thus, the general assessment should be focussed on the specific instances from the user.

For this purpose, we developed an additional tool, capable to show the chemicals present in the training set of the model more similar to the compound of interest, along with their predicted and experimental values, and a value of a similarity index. In this way, the user can see and evaluate in a very transparent way if the predictions for similar compounds were good or not.

Producing models of high quality, beyond the original plans

The models above-described have been based on high quality data. The models have been tailored to be used for REACH, dedicating attention to the reproducibility and easiness of use.

As additional activities, compared to what originally planned in the Description of work, we developed more than one model, for the same endpoint. In some cases, like for carcinogenicity, models address different outputs: carcinogenicity potency (a continuous value) or a carcinogenic class (binary classification). In other

cases the output was the same. Nevertheless, multiple models are useful, to confirm prediction. In some cases, as further additional activity, we integrated more models into a unique modelling approach, which optimizes the use of multiple models. This is the case of BCF, where different models are used in parallel and then results combined (not with a single statistical operator, like average or minimum value, but according to an intelligent system). In another case, the models are used in a sequence, like for mutagenicity.

Furthermore, as further activity, we added the calculation of the chemical descriptors in the overall model, achieving a maximum of easiness of the model. This was done for the BCF and skin sensitisation models, thanks to an agreement with the Dragon developers, and for developmental toxicity in a collaboration with US EPA. Since the software from US EPA became available only at the end of 2008, we had to repeat and retune models which have been previously done, in order to use the new chemical descriptors. This requested in practice to build up completely new models, and thus further, unplanned activities. The advantage of the solution of adopting descriptors from the US EPA is that they are publicly available, and a collaboration with the US EPA allowed us to implement some models into a model simpler for the user, since chemical descriptors have not to be calculated by the user. The same solution has been implemented with Dragon's developer, for the use of Dragon (which is a commercial package).

Finally, we developed the tool for similarity and applicability domain, as above described, which resulted to be an additional fundamental element for the benefit of the user.

1.6.2. The quality protocol

The main aims of the quality protocol are:

- Conformity: conformance to requirements (Objectives)
- Reliability: the models produce reliable prediction, according to specified criteria
- Reproducibility: the models have to give the same results when made by a second Partner
- Transparency: the deliverables and all model components must be well documented; all steps, processes, algorithms and theoretical description of the system have been accessible to the interested potential user.
- Testability: rigorous tests for validation have been performed
- Efficiency: the models must use resources in the best and most effective manner
- Portability: the software can be easily ported on different environments and computers

We verified all these conditions. We already described that we individually checked all structures, and that data have been taken in order to use values of high quality. This activity took about one year. Typically, this kind of work is not done, because it is time consuming, and because it does not allow an immediate reward for the efforts it takes. However, this activity is fundamental for the special application, related to the regulatory context. We notice that the US EPA, for instance, since years dedicate a special activity in this direction, which resulted in the development of the DSSTox database

(http://www.epa.gov/dsstox_structurebrowser/). We collaborated with the scientist responsible for this database (Dr Ann Rcihard, see par 1.6.3).

As w said, we implemented all algorithms developed within CAESAR. As described below, these implemented models have been done using Java. As a result, all CAESAR models have been rebuilt in this other language, and this required a complete check of all the equations and results. Thus, this activities also addressed the reproducibility of the results, repeating the model building into another situation. Furthermore, for the case of the model for BCF, it was also rebuilt by Partner Politecnico di Milan in Python, and thus this model has been "checked" three times: the original model, developed as described above, the model implemented in Java, and the model implemented in Python. Results were consistent.

The quality control of the final model involved also the common information technology testing, with debugging activities, and checks done with up to 900 compounds tested simultaneously.

The models have been tested on different browsers and operative systems.

1.6.3. Collaboration with US EPA, and other EC projects

CAESAR promoted collaborations, in order to improve the quality of the results.

In particular, we collaborated with the US EPA. In one case we collaborated with **Dr Ann Richard, Research Triangle Park, US EPA**, for the check of the data of the carcinogenicity data set, the CPDB. This data base has been developed by Lois Gold and co-workers. It includes experimental values on more than one thousand chemicals. Dr Richard, developing the DSSTox database of the US EPA, checked a series of datasets, including this one from Lois Gold. Thus, we used the data which are freely available at the DSSTox website. However, we found some possible errors, and for this we contacted Dr Ann Richard. She verified that for some structures there were errors in the chemical structures, and she corrected the wrong structures (acknowledgements to the coordinator of CAESAR is present in the DSSTox database). For another likely error, related to the experimental value of one compound, she has not been capable to solve the issue, because Dr Lois Gold retired. For this reason we eliminated this compound from our dataset (see above).

The collaboration with **Dr Martin Todd and Dough Young, Cincinnati, US EPA**, is more recent. This group made available a software to calculate chemical descriptors at the end of 2008. Thus, within a few months (CAESAR finished in March 2009), we rebuilt some of the CAESAR models using their software for chemical descriptors. They also gave us their software to calculate chemical descriptors, and this collaboration made possible the implementation of the model on developmental toxicity. We notice that the model has been developed by CAESAR (actually two models at least have been developed using the US EPA models, in addition to the many models developed before, which however used descriptors calculated with many different chemical descriptors, and thus of more difficult implementation and use), and thus this complies with our contractual duties. The collaboration activity with US EPA involves an additional activity, the implementation of the model with included chemical descriptors, which is not present in the contract, as described above.

We notice that the software to calculate chemical descriptor from the US EPA site is public, but not open source. The common user can get the prediction of chemical descriptors, but not the code of the software. Within our collaboration, we obtained the software (code) for the descriptors necessary for us.

CAESAR made available the data sets to other EC projects: **OSIRIS** (http://www.osiris-reach.eu/), **CHEMOMENTUM** (http://www.chemomentum.org/c9m), **CHEMPREDICT** and **CASCADE** (http://www.cascadenet.org/). We already discussed that the careful check of the data is very time consuming. The availability of high value data is precious, and in order to avoid repetition of work, we donated these data sets to the other projects.

We notice that all data are freely available since the end of the project, March 2009, and that they can be obtained from the published papers and from the CAESAR web site.

Furthermore, we collaborated in the development of QSAR models with **CHEMPREDICT**. This project is dedicated to the development of new QSAR models based on SMILES and InChI formats. Thus the purpose of CHEMPREDICT is on new techniques, not in new applications. CHEMPREDICT developed a model on carcinogenicity for continuous values (carcinogenicity potency) in collaboration with CAESAR and using the data set from CAESAR. Since CAESAR also developed other models for carcinogenicity addressing carcinogenicity class, these last models formally satisfy the requested activities according to the contract, and the model on carcinogenicity in collaboration with CHEMPREDICT is an additional activity, both for CAESAR and CHEMPREDICT.

Furthermore CAESAR made available the developed models to **OSIRIS**. These models have been evaluated for their use within REACH.

1.6.4. Dissemination, national events and final workshop

An aim of CAESAR project is to achieve a wider dissemination the project results. To reach the larger number of stakeholders (Regulatory Bodies and users such as chemical industry or citizens) the web site is nowadays a primary way. This policy allows dissemination of the scientific basis of modelling as well as practical results. Indeed, the web site is a user oriented portal to access the models, the user manual, the online documentation on the models and their results, and a series of papers and presentations, including the presentations given at the final workshop of CAESAR.

The scientific and technical information obtained during the CAESAR project are published or will be published in peer-reviewed international scientific journals and presented at national and international conferences. At least one paper for each endpoint with the description of the relative new model will be published on a scientific journal. On the web site can be found all the list of the scientific publications related to the project. A total of 18 publications and 64 communications have been acquired during the 36 months of the project.

In order to facilitate the access of the results of the models on scientific papers, CAESAR decided to publish a collection of them on an open journal (Central Chemistry Journal), even if this will increase the costs for the partners (the papers will appear after the end of the project). This is another example of the strong determination to provide maximum benefit for the users, not for the consortium.

Central Safety Laboratories (CSL) has organised a workshop in order to discuss the state of the art of QSARs in relation to the demands set by REACH. It has been held in Liverpool (UK) on May 16 2008, directly following the fifth CAESAR meeting. The workshop was attended by approximately 15 delegates from the scientific community, regulatory agencies, and industry and some 30 representatives of all CAESAR partners. A series of presentations were made at the workshop and have also been placed on the public CAESAR website. Briefly, the presentation discussed the use of QSARs, in relation to REACH, and the contribution made by CAESAR project in that direction, with a highlight of other complementary projects, such as OSIRIS, CHEMOMENTUM and CHEMPREDICT.

This workshop was not planned in the contract, but was decided in order to improve the dissemination of the CAESAR project and to get feed-back from stakeholders.

To present the final results of the project to the stakeholders, it was organised in international workshop on March 10-11, 2009 in Milan. Many Regulatory Bodies and industries were invited.

About 150 participants were present, mainly from industry. Regulators from almost all European countries, including new Member Stated and Associated Countries and Turkey, were present. Three representatives of ECHA were present, and they gave a presentation. Another presentation was given by JRC, which send participants. The approach of CAESAR, the models, their results, and an on line demo of the models was given. Wide space to discussion was given. A questionnaire was also distributed, which received feed-backs, comments and suggestions. The workshop and the models were appreciated as innovative.

1.6.5. The web site

3. A public website (http://www.caesar-project.eu) has been developed and officially opened at month 5 as explained in Deliverable 7. The initial global structure consists of these sections:

   - **Home** (Figure 4): gives a general overview of the project and the latest news related to it. There are two subsections, *approaches* explain the general approach that have been chosen in CAESAR, whether in the *about reach* section is possible to find information about the REACH legislation and how CAESAR project deal with it.
   - **Participants:** this section, in the final version, is divided in 4 subsections: *partners, subcontractors, collaborations* and *advisory board.*
   - **Links:** in this page there is a list of links to other EC Projects with which CAESAR has established some kind of collaborations.
   - **Contacts:** contains all the information needed to contact the project coordinator.
   - **Results:** is the most important section of the website and contains all the results of the CAESAR project, regarding the *publications and presentations* and the models developed for each *endpoint*. The endpoint subsection is furthermore sub divided in other sections, each one dedicated to one endpoint. These endpoint sections contains brief introduction and report the results obtained by the developed models.

- **Login:** this section is accessible from the Home page and from here is possible, for the registered users, to access to the CAESAR Intranet.

Most of these sections have been updated many times during the project, especially the results one.

During the last months the main website has been divided in three main so called sub-websites. Beside the sections dedicated to the CAESAR project, 2 other “root” sections have been added:

- the **workshop** sub-website, dedicated to the CAESAR final workshop that took place on March 10th - 11th 2009 in Milan. From this page is possible to download all the presentation files (in PDF version) from the workshop;
- the **software** sub-website gives the access to the on line application which contains the implementation of the models developed in the CAESAR project and is freely accessible and usable.

Figure 4. - The home page of the Caesar web site.

1.7. Results

1.7.1. The applet

1.7.1.1. BCF

1.7.1.1.1. Results

The model (Figure 5) is constituted by a combination of 2 Radial Basis Function Neural Network (RBF-NN) models developed with 5 descriptors each, for a total of 8 descriptors (2 are in common between the models).

In other words:

$$If\ \text{Mean} > 2.41\ then\ \text{logBCF} = 1.052 * \text{Mean} - 0.065\ else\ if\ 1.355 < \text{Mean} \leq 2.41\ then\ \text{logBCF} = 0.996 * \text{Min} + 0.042\ else\ \text{logBCF} = 0.936 * \text{Mean} - 0.123$$

Where Min, Mean and Max are calculated with the values given by the individual models.

The model reached an R2 = 0.83 on the training set, and R2 = 0.80 on the test set.

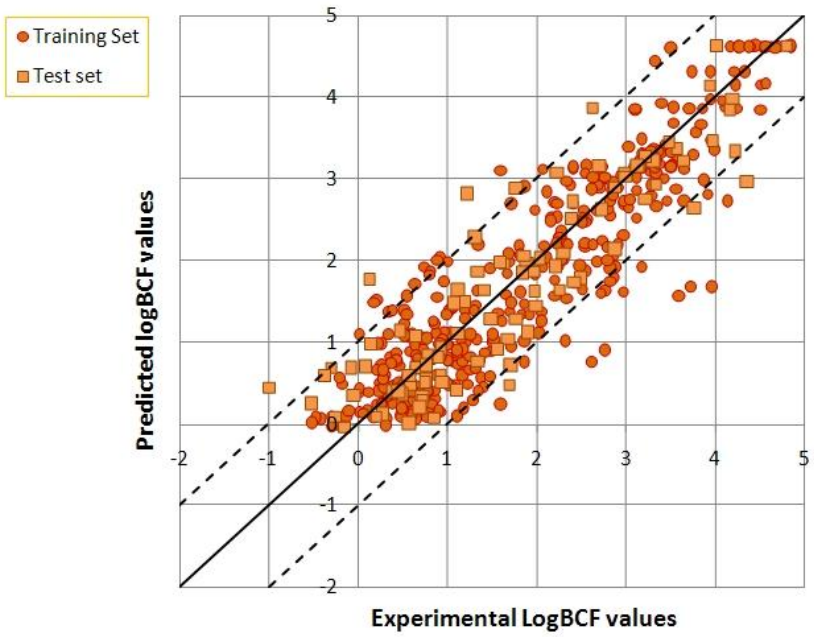


Figure 5. - The predicted and experimental LogBCF values obtained with the CAESAR model.

The error of the prediction was about 0.5 Log unit (***S**tandard **D**eviation of the **E**rror of **P**rediction*), which is of the same range of the experimental variability. The performances of this model were better than those from the EPISuite model for the same compounds.

The results of the CAESAR model do not change considering the different tautomers of the chemical structures.

1.7.1.1.2. Achievements to the state-of-the-art

The CAESAR BCF model, freely available on the web site of the project, has performances better than other software (see Table 7). The robustness of the model was tested with an external set of compounds, collected by the EURAS and the Arnot databases (see Table 8).

To obtain the predictions, the user needs only the SMILES or sdf of the molecule, because the molecular descriptors calculation is implemented into the software. In the applet we also implemented the rules that identify the chemical domain of the model. If a compound is outside the chemical domain, a remark advises the user that there is a greater uncertainty in the prediction.

Table 7 - Comparison of the performance between EPI Suite v4.0 and the CAESAR BCF model. The CAESAR BCF model R^2 is always greater than those of EPI Suite v4.0. The number of compounds predicted for each set is reported between parentheses.

	Epi Suite v4.0	**CAESAR**
Training set	R^2 = 0.72 (451)	R^2 = 0.83 (368)
Test set	R^2 = 0.67 (112)	R^2 = 0.82 (91)
External compounds	R^2 = 0.56 (71)	R^2 = 0.61 (184)

Table. 8 - The table show the statistical analysis for both the entire datasets and for the dataset pruned using the rules.

	Entire dataset		**Pruned dataset**	
	R^2	**SDEP**	**R^2**	**SDEP**
Arnot	0.58	0.93	0.70	0.69
EURAS	0.62	0.81	0.69	0.70

1.7.1.2. Skin sensitization

1.7.1.2.1. Results

As we explained, for this endpoint two approaches have been adopted. In the first case the model addressed a general population of industrial chemicals (global model), while in the second case a specific model has been developed for compounds of a particular class. The model in this case is than coupled to a read across approach, which serves to predict the toxicity.

Results of the binary classification global model are shown in figure 6. It shows a very limited number of errors. As in all CAESAR models, particular attention has been dedicated to minimise the occurrence of false negatives. Thus, errors which appear in figure 5 are mainly false positives.

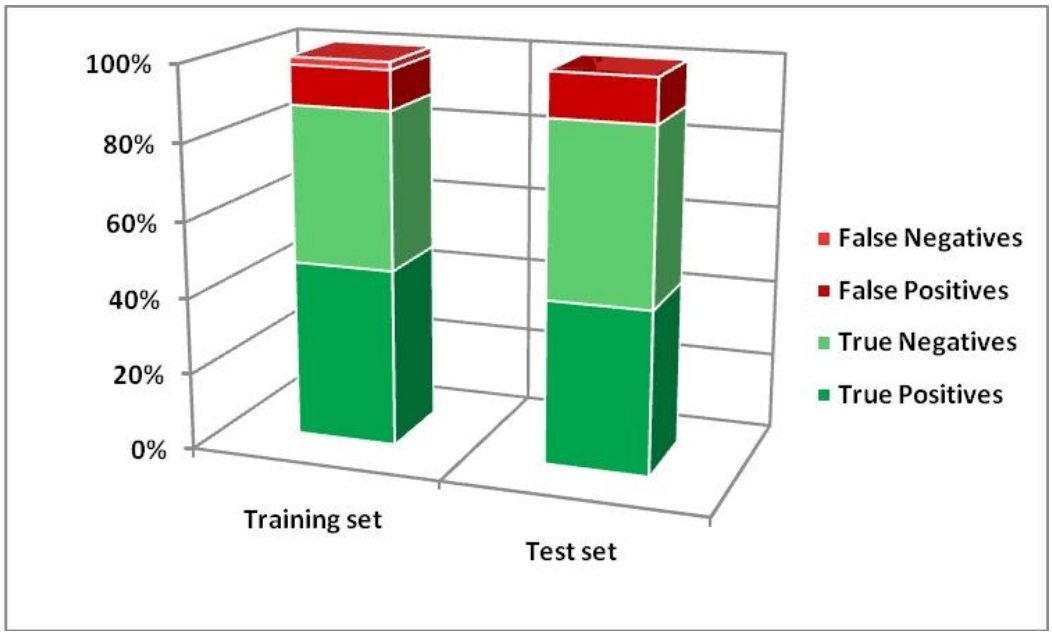


Figure 6. - The results of the CAESAR classification model for skin sensitization

The approach for the mode-of-action specific model is based on the assignment, done by human experts, of a particular kind of activity. A model has been developed which predict this kind of activity. Table 9 shows the confusion matrix for the set of chemicals, as predicted. On the columns there are the predicted classes, while on the rows there the chemical assigned by human experts. Expert assignment is from Roberts et al, 2007a. A limited number of errors are present.

Table 9. – The results for validation set of chemicals.

	Ac	**pro-MA**	**SB**	**SN2**	**non**
Ac	2				
pro-MA		5			4
SB			4		
SN2				6	
non					23

Table 10 shows the confusion matrix showing the classification statistics for the training data. The columns indicate the expert assignment from reference Roberts et al., 2007b; the rows report the class from the SMARTS pattern.

The results of this approach are the subject of a publication (Enoch et al., 2008a).

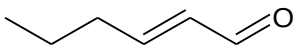
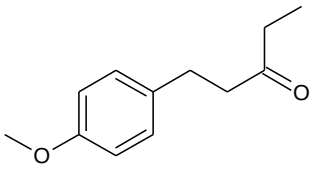
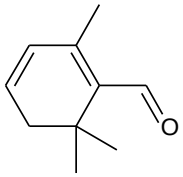
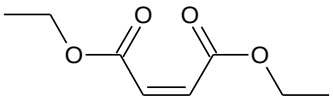
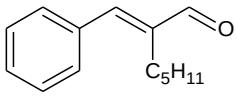
Once a specific activity has been assigned, the property value can be predicted. Examples of the predictions obtained by the method can be seen in the table 11 and full details of how to carry out the calculations and

predictions for the remaining chemicals assigned to the Michael domain can be found in (Enoch et al.,2008b).

Table 10. - classification statistics for the training data

	Ac	**MA**	**pro-MA**	**SB**	**pro-SB**	**SN2**	**pro-SN2**	**SNAr**	**SN1**	**non**
Ac	24			1		1				
MA		29								2
pro-MA			25							
SB				36						3
pro-SB					4					
SN2	2	2				45				
pro-SN2							2			
SNAr								3		
SN1										1
non										22

Table 11. - examples of predictions. NP means a prediction has not been made is not a chemical more electrophilic (larger ω) or less electrophilic (smaller ω) in this small.

Name	**Structure**	**Exp. EC3**	**Pred. EC3**	**ω**
trans-2-hexenal		5.5	NP	1.608
1-(4-methoxyphenyl)-1-penten-3-one		9.3	9.87	1.734
safranal (1,1,3-trimethyl-2-formylcyclohexa-2,4-diene)		7.5	5.29	1.796
diethyl maleate		5.8	8.74	1.804
α-amyl cinnamic aldehyde		11	NP	1.839

1.7.1.2.2. Achievements to the state-of-the-art

The so developed models represent innovative solutions. The global model, which predicts a class, gave very good results. It is important to notice that the model is conservative: errors as false negatives are very low.

The local model combines an expert system which can be coupled to read across. It is very transparent and reflects human knowledge.

1.7.1.3. Mutagenicity

1.7.1.3.1. Results

A global overlook of the performances of the combined model on the test set is illustrated in figure 7: compounds are classified as mutagens, but 86 are non-mutagens (false positive). The remaining compounds, classified as non mutagens, can be further processed with the second set of rules. In this case 12 further compounds, which are mutagens, are indentified. Thus, the overall model can extract from the presumed safe compounds the most suspect ones with an impressive specificity, if related to the very low number of real mutagens still present.

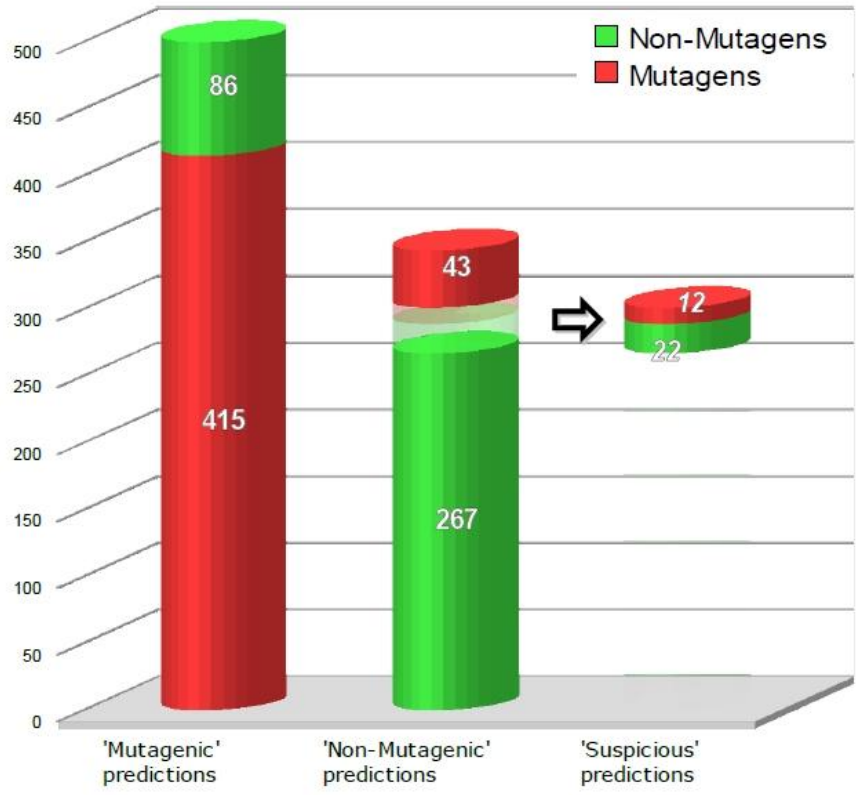


Figure 7. - The results of the combined model on the test set

1.7.1.3.2. Achievements to the state-of-the-art

The model on mutagenicity gave good results (Figure 8). It has to be remembered that he variability of the experimental methods is about 85%. The accuracy (or concordance) of our model is similar to this figure. Other models exist, and we tested them. The best one, on this large set of compounds (overall, almost 5000 compounds) was the model ToxTree (Benigni et al., 2008). The results of ToxTree and CAESAR in terms of accuracy were quite similar; however, CAESAR had a lower number of false negatives, and thus it is preferable for regulatory purposes. Commercial programs, like TOPKAT (TOPKAT, version 6.1), MultiCASE (http://www.multicase.com/) or DEREK (Marchant et al., 2008), gave worst results.

Figure 8. - The results of the CAESAR classification model for mutagenicity with SVM

1.7.1.4. Carcinogenicity

1.7.1.4.1. Results

Good results have been obtained with the regression model, for continuous values (developed in collaboration with CHEMPREDICT) on both the training and test set, as shown below. Figure 9 show the correlation between the experimental TD50 (x-axis) and that calculated by the model (y-axis).

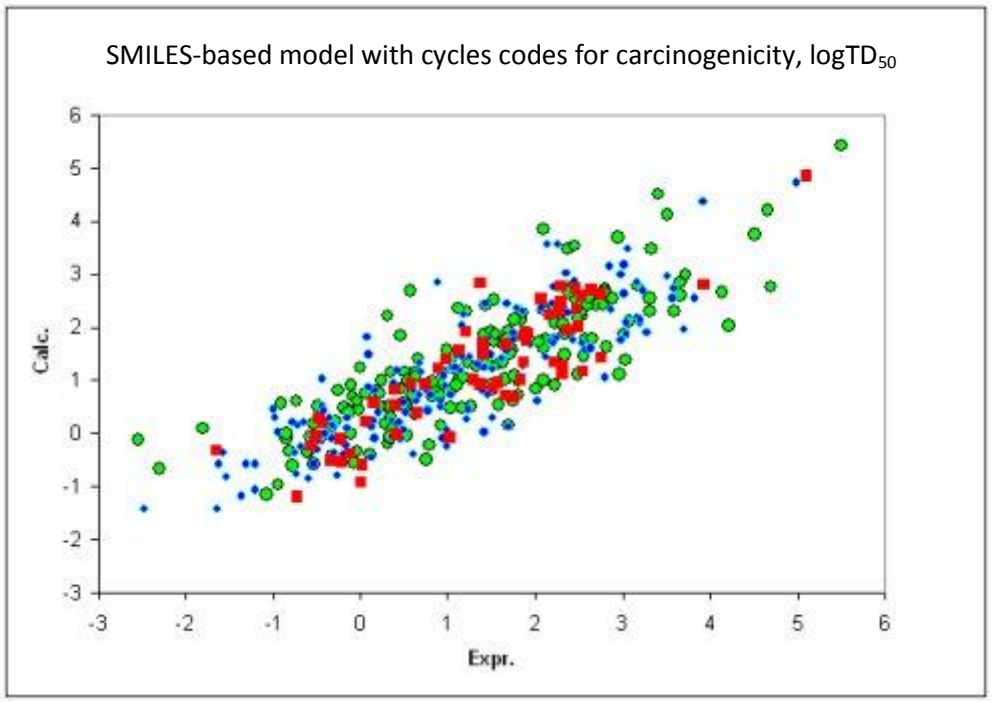


Figure 9. - The results of the CAESAR regression model; (○) Training set (n = 170, r^2 = 0.68, s = 0.80), (•) Calibration set (n = 170, r^2 = 0.75, s = 0.74), (■) Test set (n = 61, r^2 = 0.76, s = 0.62).

Results of the binary classification model are shown in figure 10.

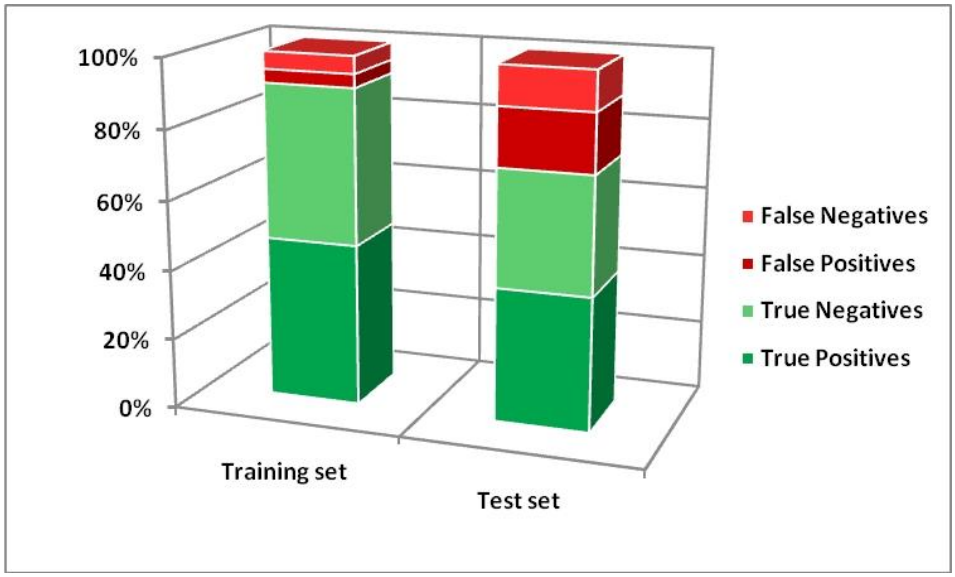


Figure 10. - The results of the CAESAR classification model for carcinogenicity

1.7.1.4.2. Achievements to the state-of-the-art

Carcinogenicity is a complex endpoint. Regulators evaluate a number of data to assess it. It is not expected that the QSAR model for carcinogenicity will replace animal models. Also the reproducibility of the animal models are quite low, and for this reason a series of tests on different animals are used, with animals also of

different sex. The experimental variability of the mutagenicity method, Ames test, is 85%, as we said above. It is expected that the variability of the animal test is higher, due to the complexity of the organism. A further issue is the extension of the results from the animal model to the human beings.

Keeping into account all these factors, we believe that CAESAR model can contribute to the evaluation of the carcinogenicity activity, as additional source of information. Results using other models and those presented in the literature are in general not superior to those of CAESAR.

1.7.1.5. Developmental toxicity

1.7.1.5.1. Results

Figures 11 and 12 show the results for the two models on developmental toxicity dataset.

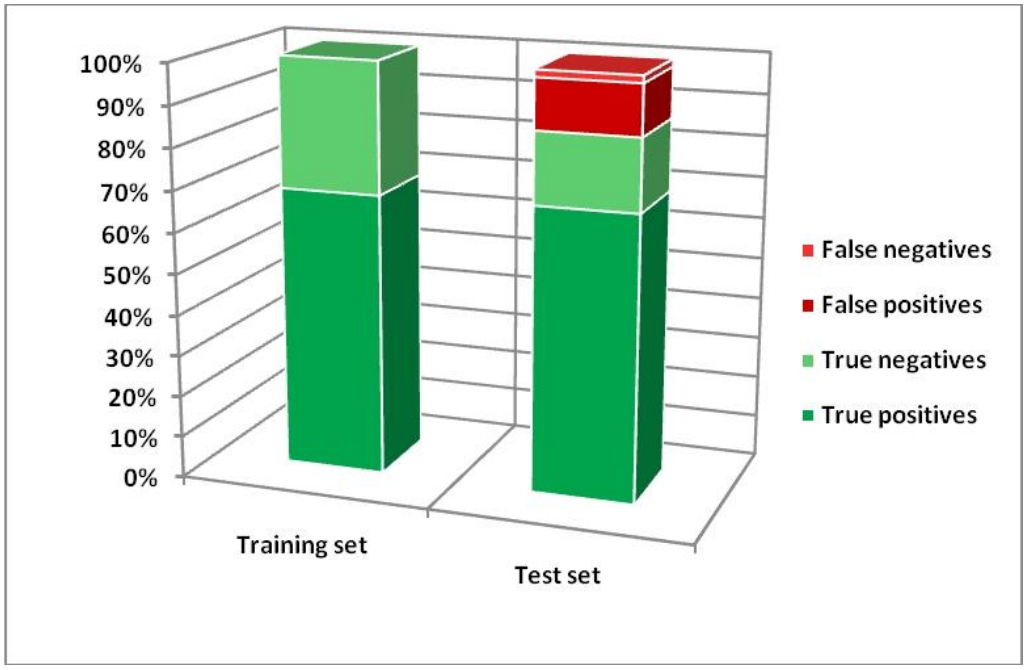


Figure 11. - Results of the model A for developmental toxicity

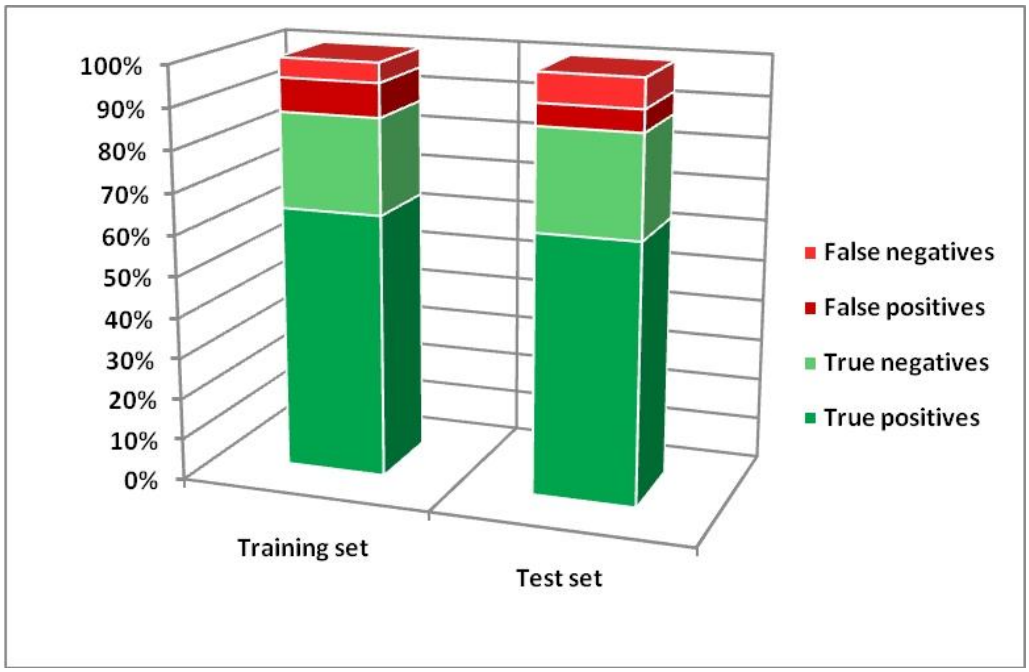


Figure 12. - Results of the model B for developmental toxicity.

1.7.1.5.2. Achievements to the state-of-the-art

The developmental toxicity endpoint is a complex one. There is a lack of QSAR models for this endpoint. Thus any model is welcome. As in the case of the carcinogenicity model the role of QSAR is as support for a general assessment, which is based on a series of data and information.

The positive aspect of the CAESAR model is that it has a low number of false negatives. However, the quite limited number of the compounds available to build up the model requires that the model is used only as support information.

1.8. Impact on the industry and research sector

The main impacts on industry of the CAESAR project are:

- the possibility to use new and validated QSAR models to gathering data for REACH fulfilment, replacing the animal tests, or in any case to provide valid data as support for the assessment of chemicals;
- advancement in Information Technology (IT) specific modelling tools, identifying new targets, whilst promoting novel solutions for some of the (Q)SAR issues. This includes a general platform for QSAR modelling (which has been developed for the implementation of the CAESAR models), including tools for similarity;
- a general approach to develop QSAR models for regulatory purposes, identifying the specific issues and how to cope with them;

- high quality data sets with toxicity data and chemical structures of for five endpoints.

There are thousands of predictive models published in recent years, but typically they are not suitable for regulatory purposes, because they have not taken into account one or more factors that are essential for validation or quality assurance, or for the specific application which is very delicate.

The typical approach is to develop a model and *then* propose its use. However, it is clear that this is a generic approach, not addressing the specific application it should be dedicated to.

The particular application of REACH needs to keep into account the stakeholders. This means that the issues to be evaluated are not only theoretical but also practical. Furthermore, other aspects, such as the cultural barriers and easiness of use should be investigated.

CAESAR addressed all these critical factors. We already discussed the scientific aspects, in the previous sections. Furthermore, we dedicated special attention to the users and the access of the models. For this reason we preferred to develop model simple to be used, and which give the same results. We underline that other approaches preferred to develop flexible models, capable to give different results for the same compound, depending on the user. This approach, in our opinion, has some drawbacks in the specific case of models for regulatory purposes, because different results can be obtained, for instance, by the industrial or regulatory user, or in different countries. This complicates the assessment of the results, and needs a successive, separate, evaluation of the results.

We preferred to follow an approach, such as that used by the US EPA. The same model is available to all actors: industry and regulators. Thus, the same rules are known, the model is transparent and reproducible.

This provides the maximum possibility of exploitation, because the model is directly usable for the regulatory application. The models can be used to process a large number of chemical in batch. This fact also increases the possibility of use, and thus the impact of the models.

Industry can use the models to provide data for REACH. The same model is available to regulators, which can verify the correctness of the values. The models have been optimised to reduce false negatives, and this also increases the acceptability from the regulatory bodies.

Furthermore, models can have an impact on industry, because they can use the model planning new compounds, and assessing in advance if a certain structure may present int he future critical aspects.

Thus, the impact for industry is high. The models are freely available, and this makes their use highly competitive. Indeed, some of the addressed endpoints are expensive to be done experimentally. It has been criticised that REACH will impose a high const to industry, but the use of CAESAR's model will solve this issue, when models will be used.

Furthermore, as we explained, the use of the models can be pro-active, to anticipate future potential problems related to certain chemical structures. As a result, industry will have a powerful tool to produce safer chemicals, which is the final target of REACH.

Regulators will get benefit from CAESAR's models too. Indeed, the models are easy to be used, can be adopted for check of all chemical structures (with restrictions as defined), and, in addition, they can be used to screen large number of compounds.

Recently, it was found that the number of chemicals potentially to be evaluated, on the basis of the preregistration of the compounds, the number of compounds to be evaluated is much higher than that which was expected. Originally, the number of dossiers which would be evaluated by regulators was quite low as percentage of the dossier submitted by industry. This higher number of preregistration may further reduce the actual number of checked dossier. In this perspective, the availability of tools, such as those developed by CAESAR, is highly beneficial for regulators, which can easily and rapidly process a high number of compounds, even all those preregistered.

Thus the potential impact of CAESAR's model, if adopted by regulators for screening purposes, is that regulators could easily process all compounds, to identify critical issues with some of them. This can improve the evaluation of the chemicals in the European market, and eventually minimise the possibility to have toxic compounds on the market.

The models developed within CAESAR project can be used to fulfil the request of REACH because they include all those high quality factors that are needed to make the use of (Q)SARs acceptable for regulatory purposes and all components of the models was carefully assessed and described. Moreover, the CAESAR models covers the most important endpoints for REACH for which there are not suitable models exist.

The models use an innovative approach: the automatic calculation of a wide range of physicochemical properties for each chemical from a 2D structure. The procedure is reproducible and fast and avoids any error associated with manual operations. However, for comparison purposes, we will also carry out manual optimisation and calculation of conformation and 3D structure based physicochemical descriptors.

Besides the specific models, the impact of CAESAR is also on the proposition of a general easy platform to offer the QSAR models. The use of this platform is very simple, because calculation is done on the background. Results are clearly presented, indicating similar compounds, and assessing the applicability domain.

This general platform and approach can positively improve the clarity and evaluation of the results, improving the models acceptance, in general.

Another important impact of CAESAR is the generation of guidelines for selecting the quality of the properties that are used as input in the models. The variability of the property data and the physicochemical descriptors was analysed.

A further impact of CAESAR, both for industry and regulators, is the availability of data sets of very high quality.

The information provided within the CAESAR approach is suitable for future use within the hypothesis of using the probabilistic risk assessment (currently evaluated within the EC as an alternative to current procedure to achieve a more realistic assessment), since uncertainty and variability will be carefully addressed.

CAESAR will have also a positive impact on the information technology (IT) community interested in (Q)SARs. IT scientists have already developed new tools for other applications, and are flexible to tackle practical problems. For the IT community, the major benefit will be an advancement in their specific modelling tools, identifying new targets, whilst promoting novel solutions for some of the (Q)SAR issues.

CAESAR project will have a positive impact merging different scientific communities to work with regulators and stakeholders. The gathering of scientists and other stakeholders under CAESAR and increasing the use of a common forum will further establish networks, and will define common targets between communities with diverse interests.

In the end, CAESAR project has contribute to the four problems identified in “Towards a European Research Area”, since it will reduce a) fragmentation of its research base, b) the brain drain to USA, c) under-financing of research, d) and it will convert research into useful societal benefits.

1.9. Conclusions and future developments

CAESAR developed new QSAR models specific for REACH for these endpoints: bioconcentration in fish, skin sensitisation, mutagenicity, carcinogenicity, developmental toxicity.

These models have been planned to allow an easy access and use: industry and regulators can freely use the models through the Internet. Three simple steps allow the user introducing the chemical(s), getting the prediction, and saving or printing it. This simplicity does not mean that the models are weak. Vice versa, this shows that the effective use of the models has been one of our main concerns.

The CAESAR models are sound and scientifically based. Globally data on about 7,000 compounds have been used to build up the models. The chemical structures have been individually double-checked, even though their sources were of high quality. We took into consideration the regulatory purposes of the models, not only in the quality check of the input data, but also assessing the intended use of the models, for instance verifying that the number of false negatives was low, and a similar approach was used for the continuous values.

Many tens of different models have been developed, and then we chose the best ones. Validation of the models was done according to sound statistical criteria, evaluating internal and external validation (external test set).

The models provide guidance on their applicability. For this, different approaches have been used. We defined a series of chemical groups/classes, for which the model does not offer optimal performances. Furthermore, the models automatically highlight if some reasons of concern appear, for the ideal use of the model, e.g. boundaries of the descriptor space. Finally, the model shows the predicted value/category, the experimental one, and the error of the six most similar compounds, in relation to the target compound for prediction. Thus, the user has a clear, immediate appreciation whether similar compounds were included developing the models and how accurate were the predictions for them. A similarity score between compounds is also provided, to guide the user.

The models can run hundreds of compounds in batch within a few seconds starting from very simple and easily available chemical formats such as SMILEs and sdf.

1.10. References

Aptula, A.O., Patlewicz G., and Roberts D.W., (2005) Skin sensitization: Reaction mechanistic applicability domains for structure-activity relationships. *Chemical Research in Toxicology, 18:1420*-1426.

Arena V.C., Sussman N.B., Mazumdar S., Macina S. Yu, O.T., (2004) The utility of structure-activity relationship (SAR) models for prediction and covariate selection in developmental toxicity: comparative analysis of logistic regression and decision tree models. *SAR QSAR Env. Res.,15, 1-18*.

Benigni R., Bossa C., Jeliazkova N., Netzeva T. & Worth A. (2008). The Benigni / Bossa rulebase for mutagenicity and carcinogenicity – a module of Toxtree. *European Commission report EUR 23241 EN*. http://ecb.jrc.ec.europa.eu/DOCUMENTS/QSAR/EUR_23241_EN.pdf.

Burges, C. J. C. (1998). *A Tutorial on Support Vector Machines for Pattern Recognition*. Kluwer Academic Publishers, Boston.

Dimitrov S., Dimitrova N., Parkerton T., Comber M., Bonnell M., Mekenyan O. (2005) Base-line model for identifying the bioaccumulation potential of chemicals. SAR QSAR Environ Res, 16, 531-554.

Enoch, S. J., Madden J. C., and Cronin, M. T. D., (2008a) Identification of mechanisms of toxic action for skin sensitisation using a SMARTS pattern based approach. *SAR and QSAR in Environmental Research, in-press.*

Enoch, S.J., Cronin M.T.D., Schultz T.W., et al., (2008b) Quantitative and Mechanistic Read Across for Predicting the Skin Sensitization Potential of Alkenes Acting via Michael Addition. *Chemical Research in Toxicology, 21:513-520.*

Gerberick, G.F., Ryan C.A., Kern P.S., et al., (2005) Compilation of historical local lymph node data for evaluation of skin sensitization alternative methods. *Dermatitis, 16:157-202*.

Pedersen F., de Bruijn J., Munn S. & van Leeuwen K. (2003). Assessment of additional testing needs under REACH. Effects of (Q)SARS, risk based testing and voluntary industry initiatives. EUROPEAN COMMISSION DIRECTORATE GENERAL JRC JOINT RESEARCH CENTRE Institute for Health and Consumer Protection

Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. *OJ L 396,* 30.12.2006, p. 1–849

Roberts, D.W., Patlewicz G., Dimitrov S.D., et al., (2007a) TIMES-SS - A mechanistic evaluation of an external validation study using reaction chemistry principles. *Chemical Research in Toxicology, 20:1321-1330.*

Roberts, D.W., Patlewicz G., Kern P.S., et al., (2007b) Mechanistic Applicability Domain Classification of a Local Lymph Node Assay Dataset for Skin Sensitization. *Chemical Research in Toxicology., 20:1019-1030.*

Van der Jagt et al. (2004). Alternative approaches can reduce the use of test animals under REACH. Addendum to the report “Assessment of additional testing needs under REACH. Effects of (Q)SARS, risk based testing and voluntary industry initiatives”

Zhao C., Boriani E., Chana A., Roncaglioni A., Benfenati E., (2008) A New Hybrid QSAR Model for Predicting Bioconcentration Factor (BCF). *Chemosphere, 73, 1701-1707*.

OECD Technical Guidelines

OECD Guideline No. 305. Bioconcentration: Flow-through Fish Test

OECD Guideline No. 406. Skin sensitization

OECD Guideline No. 414. Prenatal development toxicity study

OECD Guideline No. 421. Reproduction/developmental toxicity screening test

OECD Guideline No. 422. Combined repeated dose toxicity study with the reproduction/developmental toxicity screening test

OECD Guideline No. 429. Skin sensitization: Local Lymph Node Assay

OECD Guideline No. 451. Carcinogenicity studies

2. Dissemination and use

This section explains the publishable results as indicated in the final dissemination plan (se Deliverable D30 for further information about the dissemination plan).

Overview table

Exploitable Knowledge (description)	Exploitable product(s) or measure(s)	Sector(s) of application	Timetable for commercial use	Patents or other IPR protection	Owner & Other Partner(s) involved
The strategies, methodologies, and quality parameters used for developing QSAR models for *in silico* prediction of the toxicity of industrial chemicals.	Tens of QSAR models for *in silico* prediction of the following 5 toxicity endpoints: 1. Bioconcentration factor 2. Carcinogenicity 3. Mutagenicity 4. Skin sensitisation 5. Developmental toxicity	For use in chemical risk assessment by: 1. Regulators 2. Industry producing, importing or using the chemicals 3. Researchers using the methodologies for developing new regulatory models for other toxicological endpoints	Background knowledge published to enable researchers to follow the strategies and develop other validated in silico tools for risk assessment. Any potential commercial exploitation will have to ensure that it does not jeopardise the free use of the models and datasets on the internet.	The background IPR used to develop the models belongs to CAESAR consortium. Each partner can exploit the models developed within CAESAR project, provided that CAESAR is acknowledged and that it does not conflict with the free use of the java-based web tool and the data sets available through the internet.	The overall ownership of the models belongs to the CAESAR project through the Coordinator P1 (IRFMN). For calculation of certain chemical descriptors the owners are outside the consortium. The models have been jointly developed by the CAESAR partners (IRFMN, BCX, POLIMI, CSL, LJMU, NIC, KM). Different partners also carried out joint quality checks and descriptor calculations (BCX, IRFMN, CSL, UFZ, NIC).
5 java-based QSAR models (implemented as automated tool) for prediction of: BCF, carcinogenicity, mutagenicity, skin sensitisation, developmental toxicity	Software to automatically obtain predicted value/class.	For use in chemical risk assessment by: 1. Regulators 2. Industry producing, importing or using the chemicals 3. Researchers using the methodologies for developing new regulatory models for other toxicological endpoints	The models are publicly, freely usable and available, for the non-commercial use, through the internet. Any possible commercial exploitation has to warrantee the free use on the internet of the existing java-based models and datasets.	The software is proprietary, and covered by licences of different groups, also outside the consortium (for instance for the Dragon and US EPA descriptors).	IRFMN for the models implementation. For the calculation of several chemical descriptors the owners are outside the consortium.

Exploitable Knowledge (description)	Exploitable product(s) or measure(s)	Sector(s) of application	Timetable for commercial use	Patents or other IPR protection	Owner & Other Partner(s) involved
Quality assessed datasets on 5 toxicological endpoints	Quality assessed data, chemical information (structures and descriptors) for the following 5 toxicity endpoints: 1. Bioconcentration factor 2. Carcinogenicity 3. Mutagenicity 4. Skin sensitisation 5. Developmental toxicity	For use in chemical risk assessment by: 1. Regulators 2. Industry producing, importing or using the chemicals 3. Researchers using the methodologies for developing new regulatory models for other toxicological endpoints	As per agreed terms of the CAESAR project, the information and data will be made accessible for free access, and as such commercial exploitation has not been planned.	The files are freely downloadable from CAESAR website.	Different partners participated in quality assurance of data and chemical information (IRFMN, UFZ, BCX, CSL).

The main exploitable outcomes of CAESAR project comprise 5 QSAR java-based models for *in silico* prediction of the following 5 toxicity endpoints for industrial chemicals:

1. Bioconcentration factor
2. Carcinogenicity
3. Mutagenicity
4. Skin sensitisation
5. Developmental toxicity

The models have been developed following strict quality control procedure in relation to toxicity data, chemical structures, descriptor calculations, and statistical algorithms used. The models have also been tested using datasets not used in modelling process, and applicability domains have been established. Each model has been particularly tested for false negative predictions and appropriately adjusted to remove the main outliers. Each model thus represents a robust and reliable tool for the assessment of chemical toxicity in the absence of experimental data. Additionally, the models are presented through java-based tools to automatically obtain predicted toxicity values/ class. As per agreed terms of the CAESAR project, the information and data will be made accessible for free access at the CAESAR website: http://www.caesar-project.eu. The user can obtain the prediction values starting from the simple chemical structure, as SMILES or sdf format.

The models are also backed up by toxicity datasets and chemical information (structures and descriptors) that have been thoroughly checked and quality-controlled by CAESAR partners. Also provided is the detailed background information in relation to strategies adopted by CAESAR for quality assurance and development and testing of the models. This information will provide another valuable resource for the development of new models by other researchers in the future.

The data sets represent a very important output, because many thousands of compounds have been checked. High quality data are useful to build up QSAR models, but in general they are useful for REACH, because any assessment of the chemical compounds relies on data. Thus, the data sets can be exploited in several ways. The data sets contain the properties for the endpoints addressed within CAESAR (bioconcentration factors, mutagenicity, carcinogenicity, skin sensitization, developmental toxicity), the name and CAS registry number of the compounds, the SMILES structure, and a series of calculated chemical descriptors.

Many conceptual QSAR models have been developed within CAESAR. There are tens of models, developed for the five endpoints addressed within CAESAR: bioconcentration factors, mutagenicity, carcinogenicity, skin sensitization, developmental toxicity. All these models use chemical descriptors which have been calculated with proprietary software, generated outside the consortium. Thus, the user should calculate these descriptors independently. This is the typical procedure for QSAR models.

Several partners have been involved in the development, check and optimization of the models. Each partner can exploit the models developed within CAESAR, provided that CAESAR is acknowledged and that this does not conflict with the free use of the java-based CAESAR models and data sets available through the internet.

In brief, the CAESAR models provide a valuable resource to industry, regulators and research community for use as a reliable and free alternative tool that will enable a rapid assessment of chemical substances to fulfil some of the data needs under REACH without the use of test animals. Besides the REACH-related applications, the results of CAESAR can be also used for other regulatory purposes and sectors (cosmetics, and plant protection products, for instance).