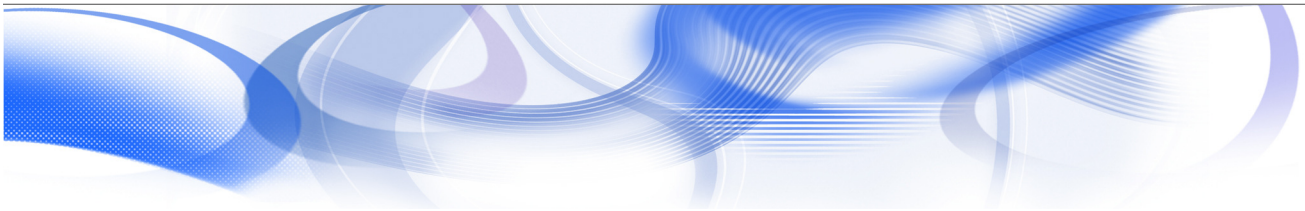
JRC Scientific and Technical Reports



Characterisation of the REACH Pre-Registered Substances List by Chemical Structure and Physicochemical Properties

Klaus Daginnus

EUR 24138 EN 2010

JRC
EUROPEAN COMMISSION

ihcp
Institute for Health and Consumer Protection

The mission of the JRC-IHCP is to protect the interests and health of the consumer in the framework of EU legislation on chemicals, food, and consumer products by providing scientific and technical support including risk-benefit assessment and analysis of traceability.

European Commission
Joint Research Centre
Institute for Health and Consumer Protection

Contact information
Address: Via Enrico Fermi 2749, Ispra 21027 (VA), Italy
E-mail: Andrew.WORTH@ec.europa.eu
Tel.: +39 0332 789566 Fax: +39 0332 786717

Klaus Daginnus (current contact details)
E-mail: klaus.daginnnus@bsu.hamburg.de

http://ihcp.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

JRC 55818

EUR 24138 EN
ISBN 978-92-79-14835-4
ISSN 1018-5593
DOI 10.2788/54409

Luxembourg: Office for Official Publications of the European Communities

Abbreviations

ACD	Advanced Chemistry development
ACD NTS	ACD/Name to Structure Batch
ADMET	Adsorption, Distribution, Metabolism, Excretion and Toxicology
CAS	Chemical Abstracts Service
DSSTOX	Distributed Structure Searchable Toxicity database (US EPA)
EChA	European Chemicals Agency
EINECS	European Inventory of Existing Commercial chemical Substances
EPA	Environmental Protection Agency (US)
EPISUITE	Estimation Program Interface Suite
EU	European Union
EUSES	European Union System for the Evaluation of Substances
FOOTPRINT	Functional Tools for pesticides risk assessment and management
InChI	IUPAC International Chemical Identifier
IUPAC	International Union of Pure and Applied Chemistry
Log D	Logarithm of the (octanol/water) distribution coefficient
Log P	Logarithm of the (octanol/water) partition coefficient
MCS	Multi-constituent substance
NCI	National Cancer Institute (US)
NIST	National Institute of Standards and Technology (US)
OECD	Organisation for Economic Co-operation and Development
PEC	Predicted Environmental Concentration
PHYSPROP	Physical Properties database
PNEC	Predicted No Effect Concentration
PRS	Pre-registered Substance (REACH)
QSAR	Quantitative Structure Activity Relationship
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals
SAR	Structure Activity Relationship
Sdf	MDL structure data file
SID	Substance Identification Data: IUPAC name, InChI, InChI key, SMILES
SMILES	Simplified Molecular Input Line Entry Specification
TSCA	Toxic Substances Control Act (US)
UVCB	Unknown or of Variable Composition or of Biological Origin
Xml	Extensible Markup Language

Abstract

In the European Union, the registrants of chemical substances under the REACH legislation are explicitly encouraged, and even required, to use non-testing methods as a means of identifying the presence or absence of hazardous properties of substances in order to meet the information requirements of REACH while at the same time minimising testing on vertebrate animals. The need to use non-testing methods or other alternative (non-animal) methods such as in vitro tests, has led to the development and implementation of Integrated Testing Strategies based as far as possible on the integrated use of non-animal data. The use of non-testing methods within such strategies implies the need for computational tools and a structured workflow to facilitate their application.

The list of pre-registered substances (PRS) published by the European Chemicals Agency includes chemicals that industry may to register in accordance with the deadlines specified in the REACH legislation. The PRS list does not include information on chemical structures, which are a prerequisite for the development and application of non-testing methods. Therefore, in order to facilitate the implementation of non-testing methods for the regulatory assessment of REACH chemicals, the Computational Toxicology Group within the Joint Research Centre (JRC) has:

- generated structures for the PRS were by using the ACDLabs Name-to-Structure (NTS) software and validated them with a random sample
- processed the structures to generate substance identifiers such as IUPAC name, InChI codes and SMILES strings
- processed the structures to obtain information on chemical characteristics suitable for a preliminary assessment of the hazard and exposure
- indicated the availability of experimental toxicological data with DSSTOX and FOOTPRINT tags
- created a “QSAR-ready” data file to support the application of non-testing methods, such as QSARs

The application of ACD NTS resulted in a high rate of yield for the generation of structures (85%) for mono-constituent substances with a high reliability – in total, about 80,000 structures were generated. By comparing these results with inventories of structures available from other publicly available sources of information, additional high quality structures including precise information on stereochemistry were generated. A quality review resulted in the assignment of quality labels to the structures and in the further checking of about 5500 structures.

To support QSAR predictions and to estimate key physicochemical properties of the substances, these structures were processed to obtain an inventory of PRS parent substances, which serves as a standardised input for computational tools containing about 62,000 records. For these parent compounds the key features of the structures were calculated and key physicochemical properties were estimated using EPISUITE, Pipeline Pilot and ADMET Predictor.

This chemical characterisation of the parent substances can be used to support a preliminary assessment of hazard and exposure. The data highlight the importance of ionisation to predict hazard and exposure for about 40% of the substances in the inventory.

CONTENTS

1. Introduction

In the European Union, the registrants of chemical substances under the REACH legislation are explicitly encouraged, and even required, to use non-testing methods as a means of identifying the presence or absence of hazardous properties of substances in order to meet the information requirements of REACH while at the same time minimising testing on vertebrate animals. The need to use non-testing methods or other alternative (non-animal) methods such as in vitro tests,= has led to the development and implementation of Integrated Testing Strategies based as far as possible on the integrated use of non-animal data. The use of non-testing methods within such strategies implies the need for computational tools and a structured workflow to facilitate their application. An integrated use of non-testing methods by means of a structured workflow has been proposed (Bassan and Worth 2008). The workflow starts with the collection of existing information.

The list of pre-registered substances (PRS) includes around 143,000 records (ECHA 2009), including substances that industry will decide to submit to the European Chemicals Agency (ECHA) for registration according to deadlines established in the REACH legislation. The substances in the PRS list are only identified by their names, CAS number and EC number, but not by chemical composition and structural information.

Information on the chemical composition (identity of the main component, other components, purity, impurities) of the substance is necessary, because predictions from (Q)SAR methods are generated from a single well defined structure, generally the two-dimensional structural formula in a convenient format such as SMILES or InChI. The purity/impurity profile may be useful at a later stage to explain discrepancies between experimental and non-testing data.

Moreover the REACH definition for substances includes not only mono-constituent substances, but also multi-constituent substances (MCS) and poorly defined substances like UVCBs. (Q)SAR predictions for MCS and UVCBs can only focus on identified constituents and their assigned structures. For this reason the potential risk of UVCBs might be underestimated. However, some UVCBs like petroleum substances can be assessed using the hydrocarbon block method and filling data gaps by using non-testing methods (CONCAWE Petrotox).

It is essential to verify the information on substance identification, names, registry numbers and structures. If a substance is known by a CAS or EC Number or by its name, it is necessary to derive its structure, e.g. as SMILES, to be used in the prediction process. If the structure is known, it is important to verify that the structural information agrees with the CAS number or with the name. Systematic names like IUPAC have an unambiguous relationship to the corresponding structure, but not all the substances in the PRS are assigned to systematic names.

Commercial chemical substance databases like CAS or public ones like PubChem or ChemSpider may be used to generate and validate information on substance identification. Some software products generate chemical structures from substance names, e.g. ACD Labs Name to Structure for systematic and semi-systematic names of general organic compounds and many natural product derivatives. In addition substance identifiers IUPAC names, InChI and SMILES can also be used.

Having verified the structure, it is important to identify its key chemical features. These include the main functional groups present as well as relevant protonation states and isomeric (e.g. tautomeric) forms. Then, information on the compound is collected from a range of available databases. A profiling of the compound is carried out by targeting a number of sources.

The chemical characterisation of the substance includes physical and chemical properties that relate to the potential for exposure, bioaccumulation and toxicity and enables a preliminary assessment of fate and toxicity of the substance.

EPISUITE is the primary modelling system utilised within U.S. EPA for providing estimates of the common physicochemical properties necessary for predicting chemical fate and transport such as octanol/water partition coefficients, water solubility, hydrolysis rate constants, and Henry's law constant. For example, log P is the key property to estimate the bio-concentration factor (Pavan et al 2008) and baseline toxicity (van Leeuwen 2007) and the BIOWIN modules are used to estimate biodegradation.

The EPISUITE calculators are based primarily on a fragment constant approach that has been validated with an independent set of chemicals. In general EPISUITE predicts physicochemical data within an order of magnitude, which is normally sufficient for regulatory screening purposes.

EPISUITE is not able to calculate ionisation constants. The use of other computational methods is necessary to estimate pKa values for complex chemical structures that contain multiple ionisable functional groups and to calculate the octanol/water partition coefficients (octanol/water distribution coefficient) and water solubility as a function of pH (Kavlock 2008). For example, ADMET Predictor is able to predict pKa for multi-protic acids and bases. Its predictive models, e.g. log P, have been ranked as the most accurate in independent third party comparisons (Mannhold 2008).

This study deals with the chemical characterization of the PRS list. It aims to provide data for the preliminary assessment of substances according to the following steps:

- to generate the structures for organic mono-constituent substances,
- to validate the generated structures
- to generate substance identification data like the IUPAC name, InChI and SMILES
- to collect experimental data for the chemical characterisation
- to generate predictions of physicochemical properties for the chemical characterisation

2. Generation of structures for Pre-registered Substances

The PRS list [version March 2009] was downloaded as an xml-file from EChA (2009), imported into Pipeline Pilot and converted to MDL sd format. The list was analysed in terms of the completeness of substance identification information and the registration dates (see table 1).

Table 1: Availability of information on PRS list

Field	Number
Total	143835
EC Number	143835
CAS Number	118285
Name	143812
Registration Date	143835
Synonym (including name and language)	14431

The table shows that some substances can only be identified by their name or synonym or registry numbers.

The inventory was merged with the EINECS inventory (JRC, 2008) to add the EINECS name and the EINECS molecular formula. All EINECS substances were transferred to the PRS list.

Substance identification data can be searched and verified in large data repositories, such as SciFinder, PubChem and ChemSpider or expert systems for the regulatory assessment of substances such as the OECD Toolbox, EPISUITE or the Danish QSAR database.

SciFinder is commercial and is the reference for structures and their assigned CAS numbers. It includes more than 49,000,000 organic and inorganic substances (SciFinder, Count 30.07.2009).

PubChem is publicly available and includes about 70,000,000 records from 99 data suppliers, thereof 39,000,000 unique structures (PubChem Count 30.07.2009). PubChem offers computed substance identification data like IUPAC-name, InChI and SMILES and information on key features of the structure as well as links to the data providers for more chemical information. PubChem offers search by uploading a chemical structure file or a molecular formula file as a query and export of the data, but not for name or text search.

ChemSpider is publicly available and includes about 50,000,000 records from 205 data suppliers (ChemSpider Count 30.07.2009). ChemSpider offer computed substance identification data like IUPAC name, InChI and SMILES and computed chemical information, as well as links to the data providers for more chemical information. Downloading upto 1000 structures per day is possible, but the user is limited to assemble 5000 records or less to build an in-house database.

The OECD Toolbox includes around 60,000 structures, regulatory inventories and databases with experimental results.

According to license agreements and use limitations theses data repositories cannot be used to assemble a database for PRS. To generate structures for the PRS inventory, the following data repositories and software were examined:

The EPISUITE includes a CAS-SMILES converter for about 108,000 substances and access to the PHYSPROP database.

The Danish (Q)SAR database is a repository of estimates from over 70 (Q)SAR models for 166,072 chemicals. The (Q)SAR models encompass endpoints for physicochemical properties, fate, eco-toxicity, absorption, metabolism and toxicity.

The DSSTOX provides a public forum for publishing downloadable, structure-searchable, standardised chemical structure files associated with toxicity data. DSSTOX has its own published quality review procedures.

The Toxic Substances Control Act (TSCA) inventory covers existing substances in US commerce.

The National Cancer Institute (NCI) database makes available screening results and chemical structural data on compounds that are covered by the developmental therapeutics program of NCI and not covered by a confidentiality agreement (update September 2003).

The NIST chemistry web-book provides users with easy access to chemical and physical property data for chemical species through the internet. The data provided in the site are from collections maintained by the NIST Standard Reference Data Program and outside contributors. Chemical names were data-mined and NTS was used to generate structures. Only structures having a NTS-report "ok" were kept for further processing.

ACD NTS generates structures from systematic and non-systematic chemical names of general organic and selected biochemical and inorganic compounds. The ACD/Dictionary offers a short look-up path to verified and pre-drawn target molecules through an index of 150,000 common and systematic names supplemented by CAS number. The quality of the generated structure can be assessed from the NTS report. ACD NTS was used to generate structures for PRS, TCSA and NIST chemistry web-book from names in an automated mode. ACD/Name generates Substance Identification Data (SID) from structures.

Figure 1 displays the results using existing data repositories and automated procedures to generate structures for the PRS.

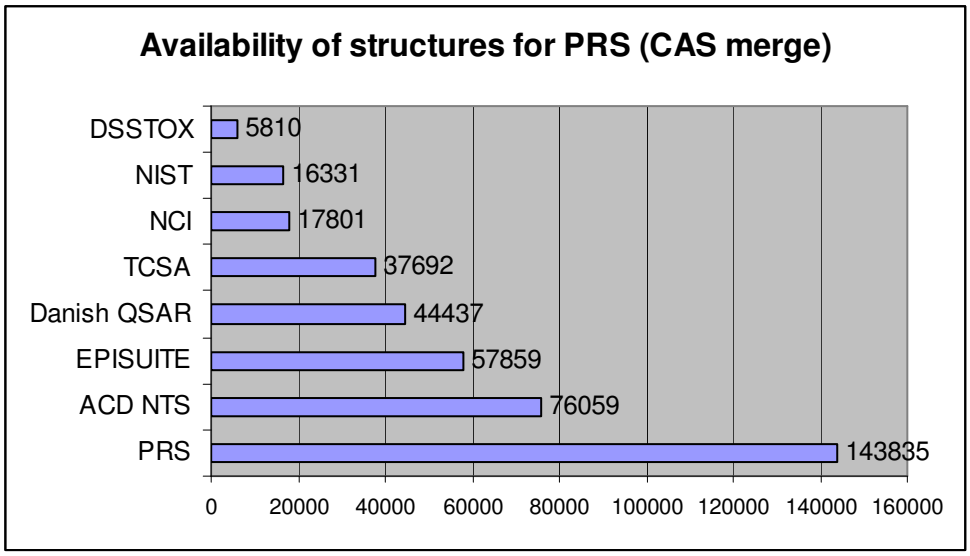


Figure 1: Availability of structures for PRS (CAS merge)

The figure shows that the ACD NTS generated most structures for the PRS inventory, followed by EPISUITE and Danish QSAR database.

The PRS inventory was run in several ACD NTS batches using the PRS name, the EINECS name, the CAS Number and the synonyms. The quality of NTS generated structures can be assessed from the NTS report:

Table 2: Generation of Quality label for generated structures

NTS report	Comment	Quality label	
Ok	substance unambiguously identified by NTS	Very high	60405
Warning: *Dictionary*	substance identified by ACD/Dictionary;	high	5787
Other Warnings	e.g. ambiguity, stereo-descriptors ignored	Warning	9687
Error*	no structure generated	-	67776

The NTS report was used to merge (CAS merge) the batch runs using Pipeline Pilot with priority rules according to Very high > High > Warnings.

These generated structures were manually checked and upgraded, if appropriate by the following measures:

1) SMILES indicated structures having bad valences [172] or structures having a formal charge other than zero [1732].
2) ACD NTS structures were upgraded (CAS merge) with structures coming from a pool of previously manually refined structures (EINECS list -processed file; JRC 2008). The list includes more than 4100 records.
3) If ACD NTS failed to generate structures for substances subject to classification and labelling (ClassLab), high production volume chemicals or low production volume chemicals (ESIS), they were generated manually, if appropriate;
4) If substances indicate a variable composition - mainly in the case of salts indicated by an x in the EINECS molecular formula — they were represented by the structures of the educts;
5) If duplicated structures (Canonical SMILES merge) were detected, they were checked and refined, if appropriate.
6) User messages[1]

To fill data gaps, structures coming from DSSTOX (February 2009 update) and NCI database, as well as structures generated from names of TSCA inventory and NIST chemistry web-book were used. A few substances were removed from the data stream[2].

The data were merged (CAS merge) using Pipeline Pilot with priority according to

Manually generated / refined structures > NTS structures > external chemical inventory

DSSTOX > TSCA > NCI database > NIST web-book

In total 80,413 structures were generated for the PRS inventory by this procedure.

A final check for structures having bad valences [12] or having a formal charge other than zero [694] lead to following results

- Some “substances?” are registered in the CAS-registry by the parent structure / large fragment of a salt. The counter-ion is not specified; therefore the structures must have an electrical charge, e.g. quaternary amines or azo-compounds.
- Some complex inorganic substances could not be adjusted without any formal charges.
- Some inorganic metals could not be handled by the properly by SMILES and the software (ACD Labs, Pipeline Pilot). Automated processing of these compounds resulted often in changes to the pre-defined oxidation state of the metal atom.

[1] User messages includes comments on the EINECS list processed file given by Andrey Yerin, ACD Labs January 2009 personal communication

[2] if substances are not assigned to EINECS formula (likely UVCBs) or substances include the name fragment “poly” (likely polymers)

- A few organic compounds can only be represented by violation of the valence model, such as coordination compounds

Additionally, the following descriptors for substance identification were added to the PRS:

- Pipeline Pilot: molecular formula, molecular weight and canonical SMILES
- ACD Labs: IUPAC name, InChI and InChI Key
- DSSTOX generic ID
- TSCA tag

The FOOTPRINT pesticides properties database as well as the DSSTOX indicate the availability of experimental data for (eco)-toxicological endpoints.

Figure 2 displays the main results of the generation of SID for PRS:

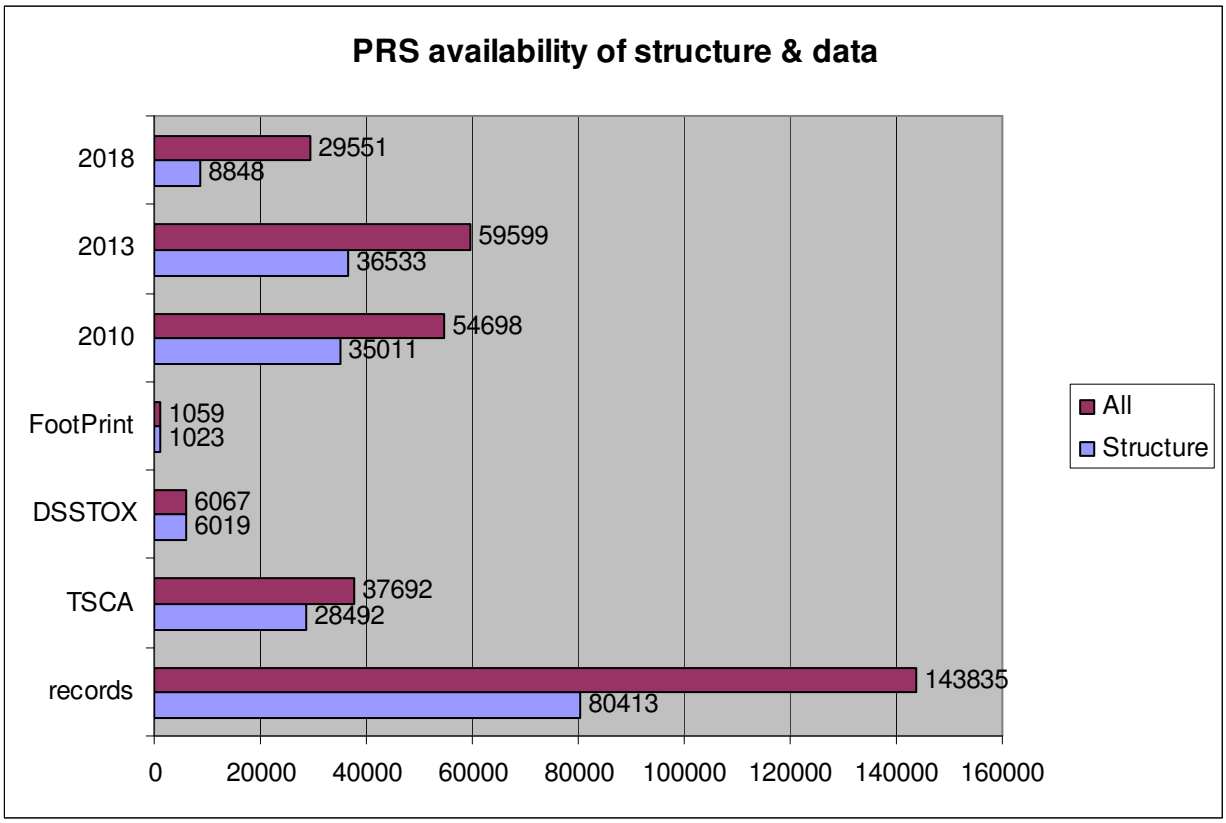


Figure 2: Availability of structures & experimental data for PRS

The data compilation is available on the JRC-IHCP Computational Toxicology webpage as an sdf file. (Annex I).

3. Validation of the NTS software

A random sample of the EINECS inventory was taken to check the performance of ACD NTS. NTS- generated data were compared with information on the structures by using SciFinder.

Figure 3 displays the main results of the validation of the NTS:

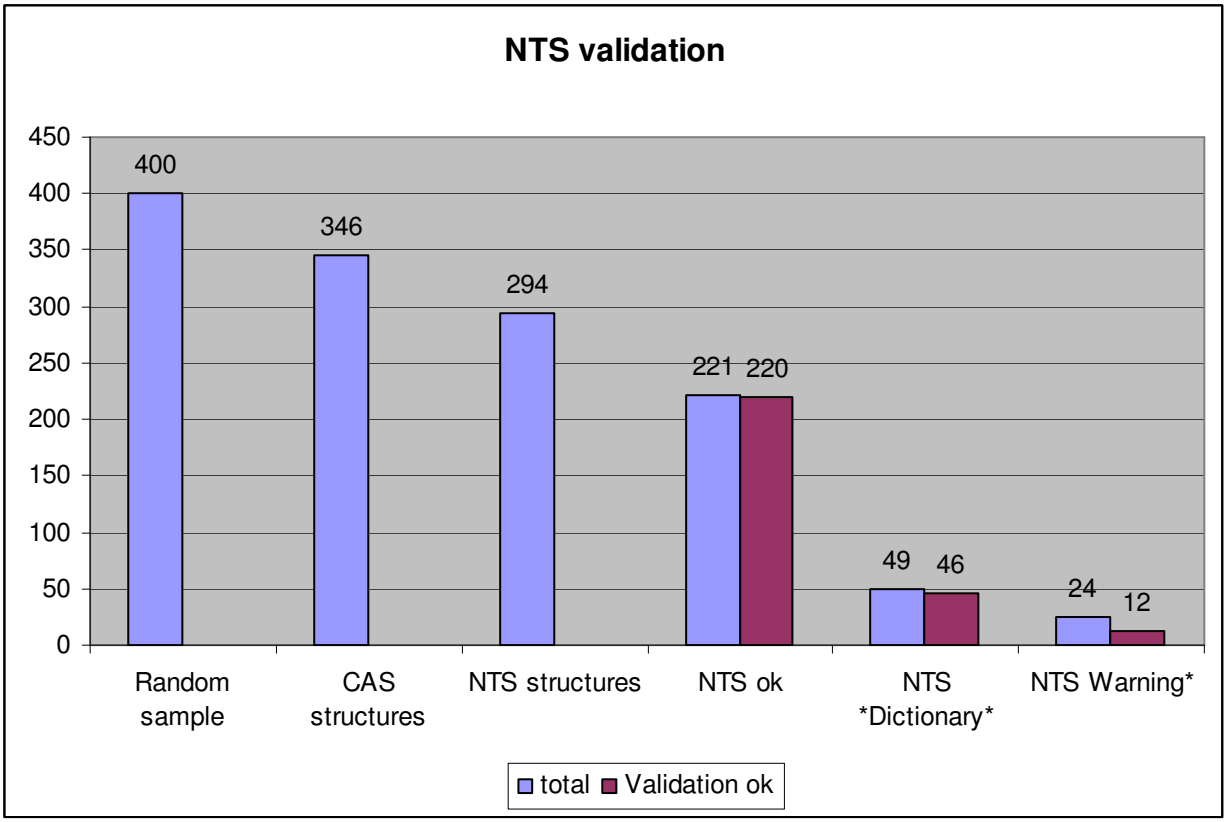


Figure 3: NTS validation

The random sample has a confidence level of 95% with a confidence interval of approximately 5%, (calculated by sample size calculator)

The random sample includes 54 substances of the type MCS or UVCB [13.5%], for which a structural representation can only be done on a component level. For 294 of 346 discrete chemicals a structure was generated [84%].

The data demonstrate the very high reliability of the ACD NTS if the structures were unambiguously identified by the algorithms or by the ACD dictionary. Further detailed analysis demonstrated that in the very few cases of deviations only one of many stereo-descriptors mismatched or in case of salts the ACD dictionary generated the structure for the parent substance, as indicated by the ACD NTS-report.

The data further demonstrate the usefulness of ACD NTS-reports for quality labelling of the generated structures including warnings:

- MCS and UVCBs are usually classified by SciFinder as incompletely defined substances. In some cases SciFinder presents the structure for MCS or UVCB with broken bonds, which indicates that the substance is composed of many constitutional isomers. These substances are often identified by the ACD NTS message: “Ambiguity is possible”
- PRS and EINECS names may have ambiguity in the names and more than one structure could be assigned to the name. These substances are often identified by the ACD NTS message: “Structure 1 of X”
- Metal ions of inorganic or metallo-organic substances may have multiple oxidation states. Automated structure generation may assign oxidation states that do not fit, e.g. for Pb+2

compounds. These substances are often identified by the ACD NTS message: “Invalid Charge”

Annex II includes the substance identification data of the CAS random sample and Annex III specifies information on deviating structures from ACD NTS compared with CAS.

The quality labelling system offers an assessment of reliability in the structural information. In particular in the case of warnings, the user is invited to check the structural information for the substance. The following table inform about the assigned quality label:

Quality label	
VeryHigh	PRS name unambiguously identified by ACD NTS
High	PRS name identified by ACD Dictionary
Checked by Computational Toxicology	structure checked and manually refined, if appropriate
Ambiguity	PRS name ambiguous, more than one structure can be assigned to a name. Ambiguity may indicate MCS or UVCB substances
Warning: invalid structure	Electrically charged structures or structures that violate valence rules
Warning	other warnings, mostly with regard to stereochemistry
External structure	reference to structure, not derived from PRS name

Table 1: Quality labels

Figure 4 shows the distribution of the assigned quality labels in the data file (Annex I)

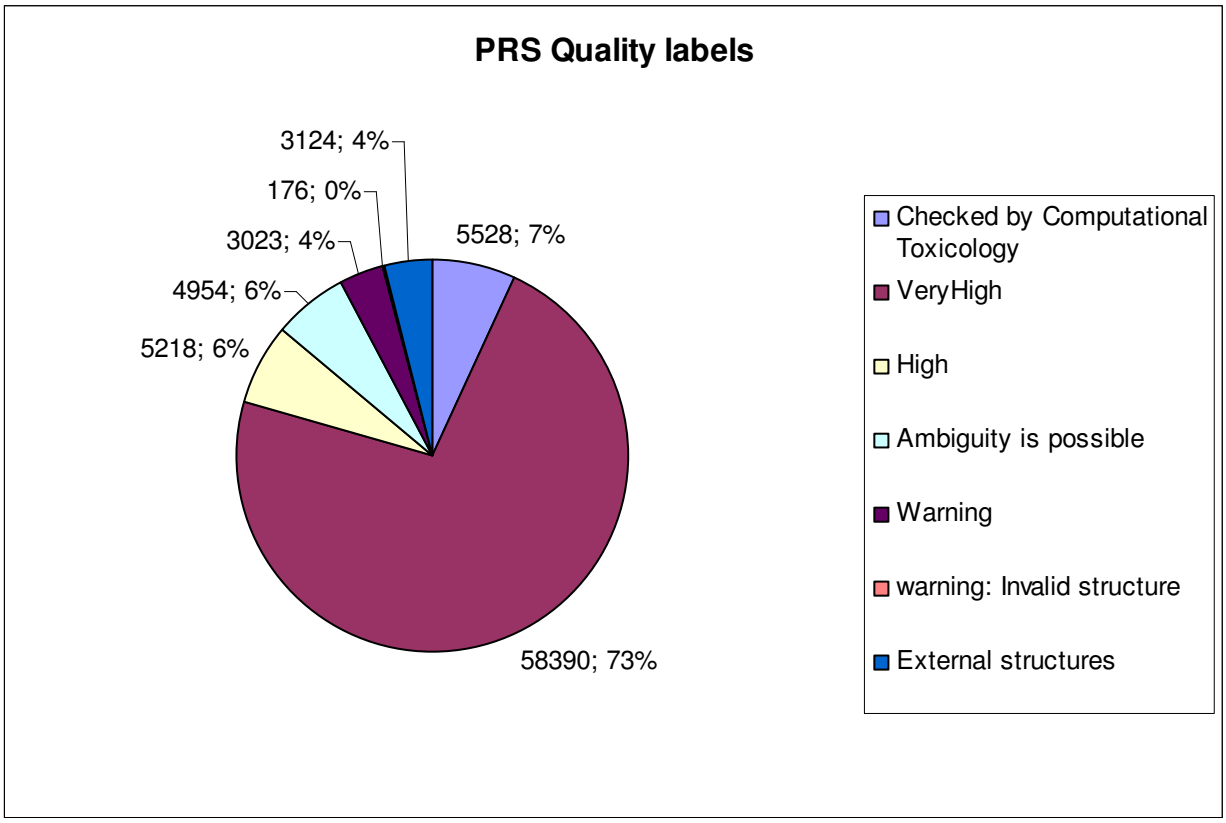


Figure 4: PRS Quality labels

Figure 4 shows in general high reliability (VeryHigh, High, Checked by the JRC Computational Toxicology group) for 90% of the generated structures. In cases of ambiguity the structural representation refers to one component of an MCS or UVCB.

4. Use of PRS inventory for QSAR and read-across applications

Non-testing methods are generally applicable for non-fragmented organic compounds. Inorganic and metallo-organic substances are usually out of the applicability domain. To obtain a standardised representation of the parent structure for organic PRS substances, the PRS inventory was processed in the following way:

1) Split and keep largest fragment, and keep smallest fragment
    a. Protonate acids, de-protonate bases
    b. Generate Canonical SMILES for parent structure
    c. Filter and remove parent structures for those having other atoms than C,N,O,P,S,F,Cl,Br,I,H
    d. Filter and remove parent structures not having C
    e. Merge of parent compounds by Canonical_Smiles
2) Merge of parent compounds from the largest / smallest fragments in step 1

All these steps led to a PRS parent inventory containing 62,016 records, which provides a standardised input for computational chemistry software. For each parent substance, the SIDs were calculated and the information on the PRS was assigned to the belonging parent substance. 6229 parent substances refer to more than one PRS. IUPAC names, SMILES and InChIs were generated for the parent substances.

The data compilation is available as an sdf file (Annex IV).

5. Chemical Characterisation of the PRS parent substances

EPISUITE was used to data-mine the PHYSPROP database and to predict physicochemical data for following the properties

> Boiling point, melting point, vapour pressure, water solubility, log P, HENRY, rate constants for atmospheric oxidation by ozone and hydroxyl radicals, hydrolysis rate constants (Kb + Ka)

Figure 5 shows the availability of experimental data compared with the inventory of structures matching PRS parent substances.

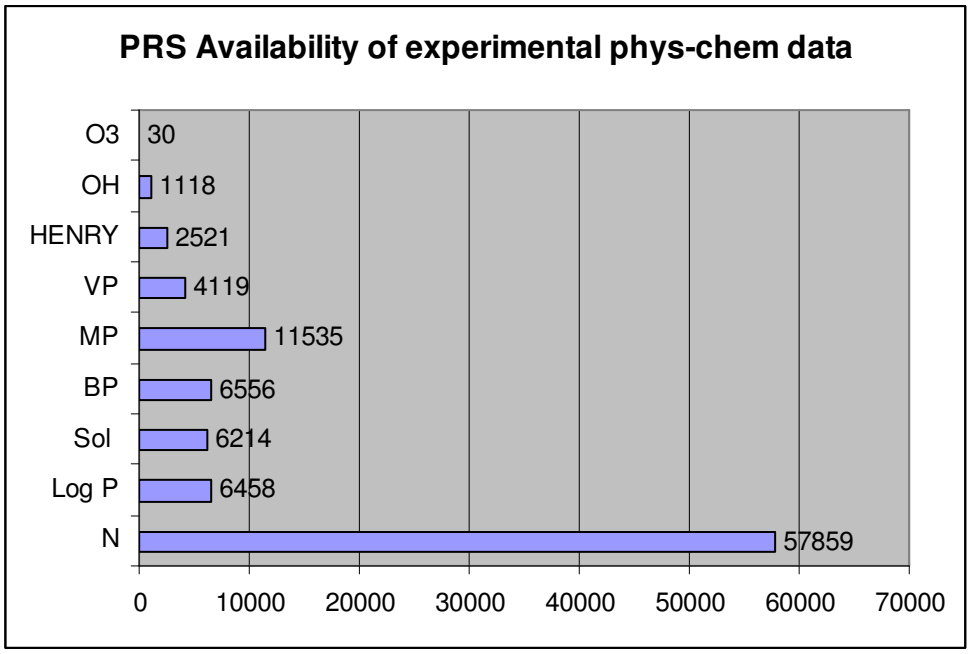


Figure 5: Availability of experimental physicochemical data in EPISUITE/PHYSPROP

The next figure shows the aggregate state of the parent substances estimated from EPISUITE predictions (BP, MP)

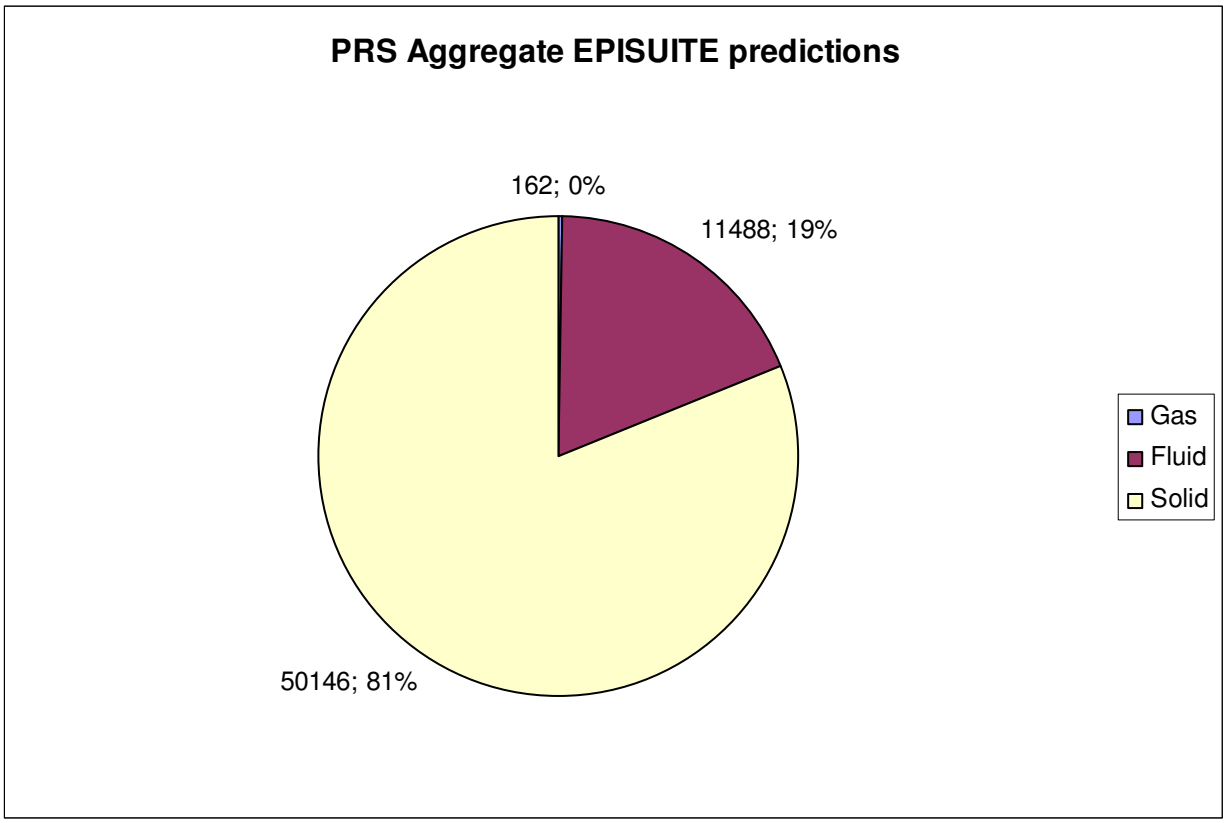


Figure 6: Aggregation state of PRS Parent substances

The next figure shows substance specific fate properties. EUSES sensitivity analysis demonstrated that, at a screening level, the fate component of the exposure potential is driven by the substance's hydrophobicity, biodegradability and volatility (Jager 1997, 2000: Jager 1998) Each of these parameters can be assigned to either a 'high' or 'low' class, within which the difference of the EUSES PEC/PNEC response is not large. Readily biodegradable has been estimated by using EPISUITE BIOWIN 3 and BIOWIN 5 results (ECHA 2008b). This leads to eight combinations of fate related properties, each associated with a specific 'Factor 3 for exposure potential.

The PRS_parents_EPISUITE.sdf data file (Annex IV) includes the presented information.

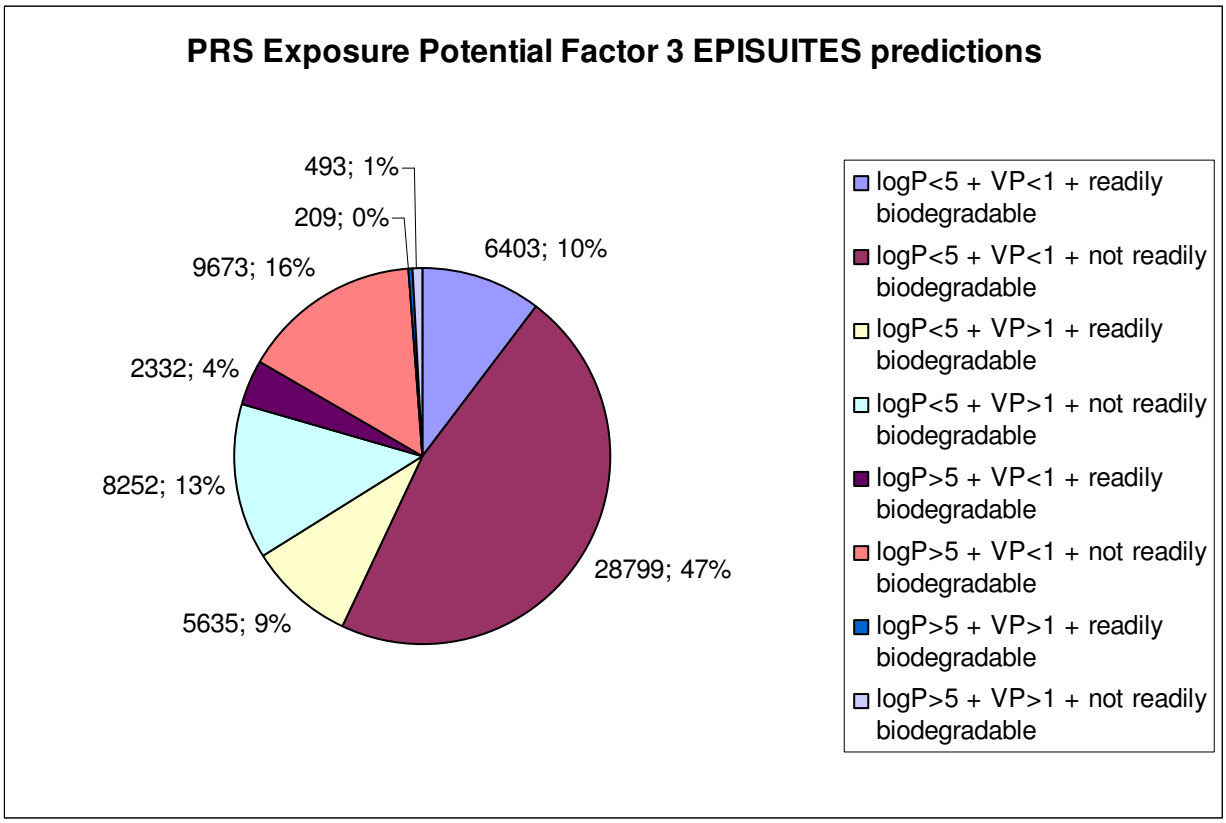


Figure 7: Exposure potential Factor 3 for PRS based on EPISUITE predictions

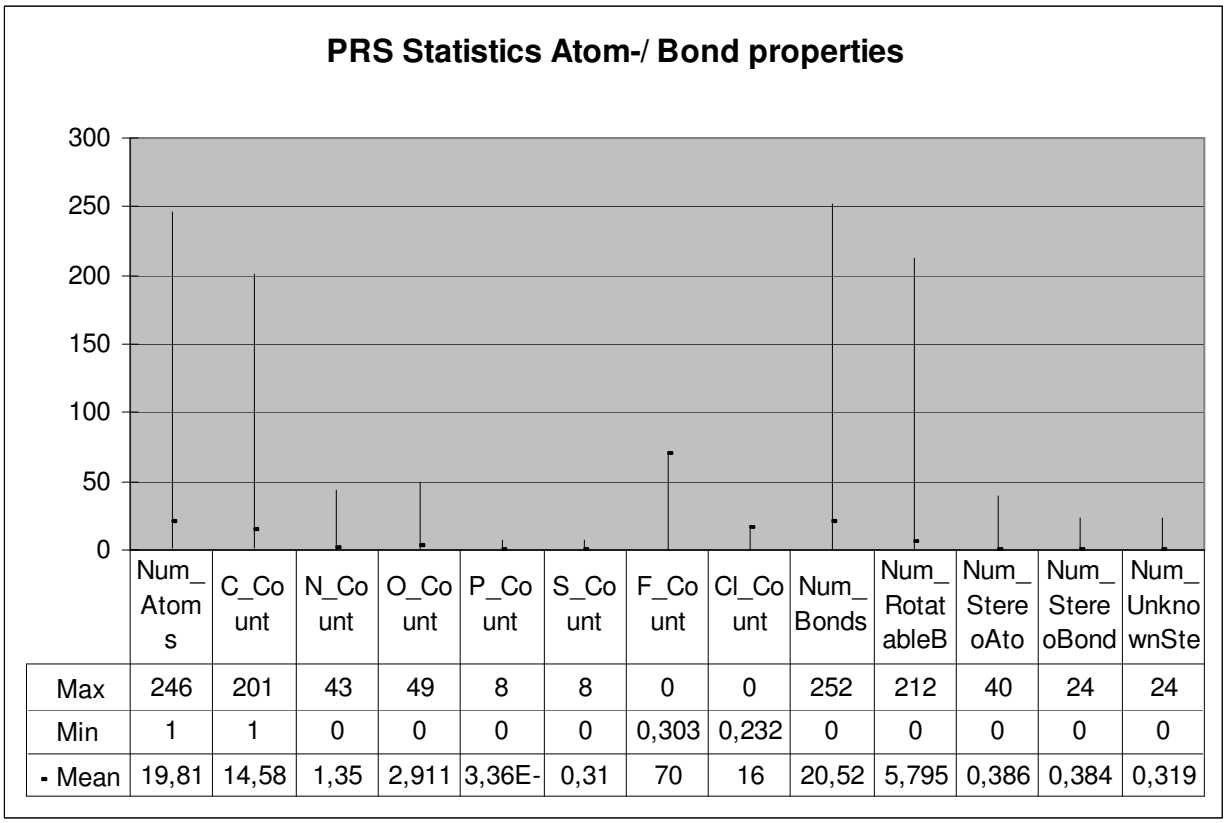
PIPELINE PILOT was used to describe the key features of the chemical structure and key physicochemical properties.

Atomic / bond properties: e.g. Element count (C, N, O, S, P, Halogens), Number of bonds

Stereochemistry: Tautomerism, known - / unknown stereobonds, is chiral

Physicochemical data: Molecular weight, molecular formula, logP, water solubility, ADME

Figure 8 shows the statistics for atom and bond properties for the PRS. The minimum / maximum values indicate the large range of substances and substance classes, the mean the average PRS substance.



	Num_ Atoms	C_Count	N_Count	O_Count	P_Count	S_Count	F_Count	Cl_Count	Num_ Bonds	Num_ RotatableB	Num_ StereoAto	Num_ StereoBond	Num_ UnknownSte
Max	246	201	43	49	8	8	0	0	252	212	40	24	24
Min	1	1	0	0	0	0	0,303	0,232	0	0	0	0	0
- Mean	19,81	14,58	1,35	2,911	3,36E-	0,31	70	16	20,52	5,795	0,386	0,384	0,319

Figure 8: Atom – and Bond properties statistics for PRS Parent Substances

The next figure shows information on stereochemistry for the PRS. The results demonstrate that for some substances the obtained structures include ambiguity:

In case of tautomeric substances, QSAR predictions may be affected by the position of the mobile hydrogen atoms.

In case of unknown stereo-bonds, molecular modelling (3D) may be affected by the chosen stereo configuration. When providing a substance dossier for REACH, data providers have to specify the stereo isomer of the substance, if possible.

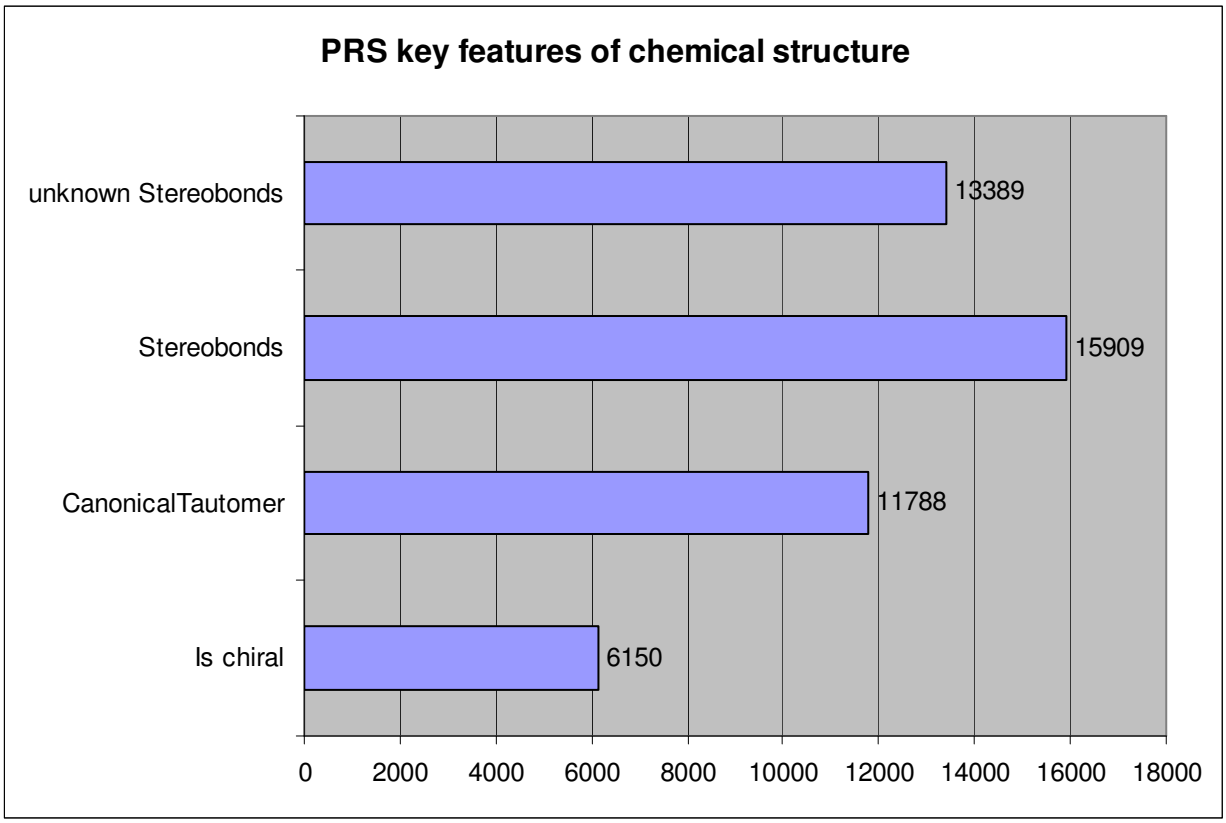


Figure 9: PRS key features of the chemical structure

Substances that obey Lipinski's rule of five were tagged. The ADMET predictor software was used to describe the ionisation of the PRS parent substances. Figure 10 illustrates that ionisation is relevant for about 40% of the PRS parent substances.

The program was run in two batches with following maximum settings:

- Limit prediction to 10 pKa per molecule
- Ignore aliphatic hydroxyl- groups
- Ignore aliphatic amides
- Skip molecules with more than 20 ionisable groups

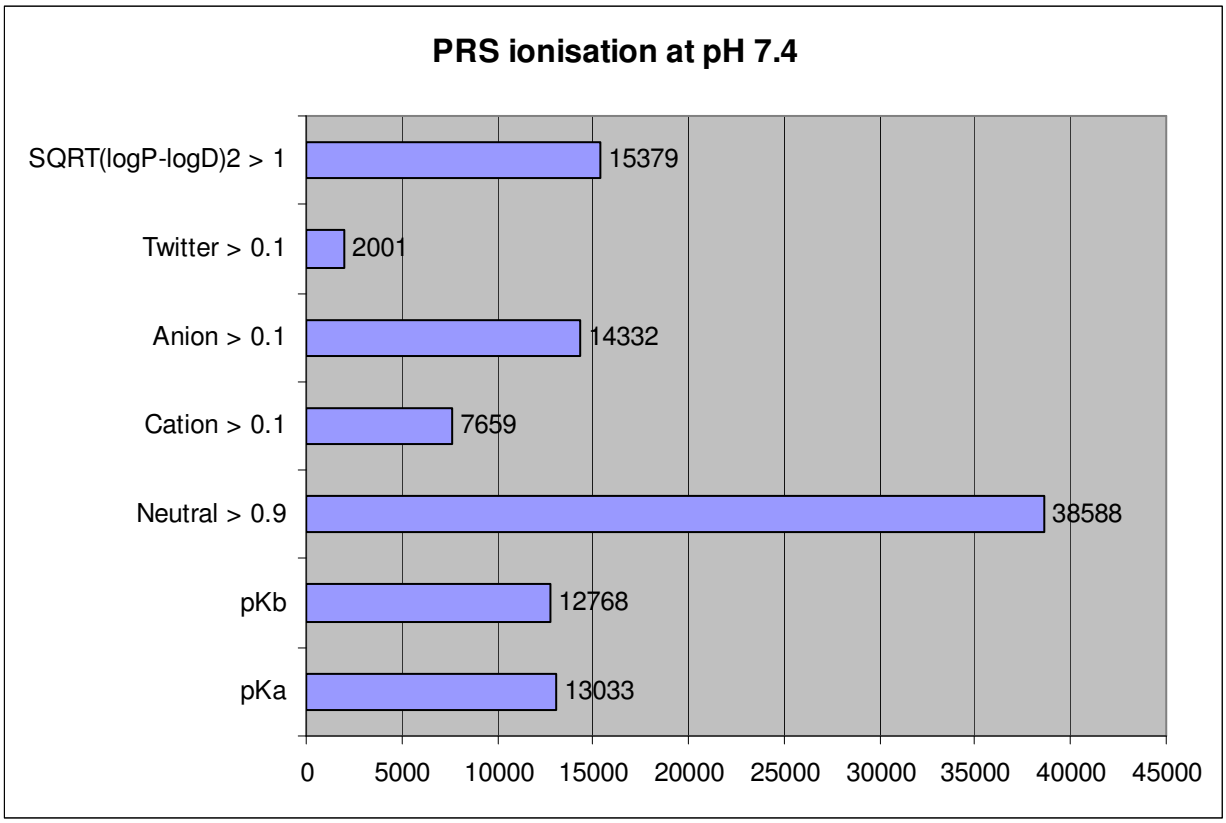


Figure 10: PRS ionisation state at pH 7.4

6. Summary and Conclusions

In the European Union, the registrants of chemical substances under the REACH legislation are explicitly encouraged, and even required, to use non-testing methods as a means of identifying the presence or absence of hazardous properties of substances in order to meet the information requirements of REACH while at the same time minimising testing on vertebrate animals. The need to use non-testing methods or other alternative (non-animal) methods such as in vitro tests,= has led to the development and implementation of Integrated Testing Strategies based as far as possible on the integrated use of non-animal data. The use of non-testing methods within such strategies implies the need for computational tools and a structured workflow to facilitate their application.

In this study, to support the workflow:

- structures for the PRS were generated with ACD NTS and validated with a CAS random sample
- the Generated structures were processed to generate substance identifiers such as IUPAC name, InChI codes and SMILES strings
- the generated structures were processed to obtain information on chemical characteristics suitable for the preliminary assessment of the hazard and exposure
- the availability of experimental toxicological data is indicated by DSSTOX and FOOTPRINT tags
- the generated information is available as a data-file to support in depth QSAR analysis

The application of ACD NTS resulted in a high rate of yield for the generation of structures (85%) for mono-constituent substances with a high reliability – in total, about 80,000 structures were generated. By comparing these results with inventories of structures available from other publicly available sources of information, additional high quality structures including precise information on stereochemistry were generated. A quality review resulted in the assignment of quality labels to the structures and in the further checking of about 5500 structures.

To support QSAR predictions and to estimate key physicochemical properties of the substances, these structures were processed to obtain an inventory of PRS parent substances, which serves as a standardized input for computational tools containing about 62.000 records. For these parent compounds the key features of the structures were calculated and key physicochemical properties were estimated using EPISUITE, Pipeline Pilot and ADMET Predictor:

- EPISUITE was used for a preliminary assessment of hazard and exposure .
- Pipeline Pilot was used to calculate key features of the structures as well as physicochemical and ADME data
- ADMET Predictor was used to calculate the ionisation of the substances, including multi-protic acids and bases, as well as physicochemical and ADMET properties

The availability of experimental data is indicated by the Distributed Structure Searchable Toxicity Database network (DSSTOX) indicating the DSSTOX-Generic-ID and the Footprint database (a tool for pesticide risk assessment and management in Europe).

7. Annexes

Annex I: PRS processed inventory

The following data compilation is available on the JRC-IHCP Computational Toxicology webpage as an sdf file (PRS_processed_file.sdf).

Field	Description
CAS Number	PRS CAS from ECHA
SMILES Parent	The simplified molecular input line entry specification or SMILES is a specification for describing the structure of chemical molecules using short ASCII strings. Automatically generated by Pipeline Pilot
Molecular Formula	Molecular formula generated by Pipeline Pilot
Molecular Weight	Molecular Weight generated by Pipeline
IUPAC Name	IUPAC names (English) automatically generated by ACD Name.
Molecular Formula	Molecular formula generated by Pipeline Pilot
Molecular Weight	Molecular Weight generated by Pipeline
InChI	The IUPAC International Chemical Identifier is a textual identifier for chemical substances, designed to provide a standard and human-readable way to encode molecular information and to facilitate the search for such information in databases and on the web. Automatically generated by ACD Name
InChI_key	The InChI key is a fixed length (25 characters) condensed digital representation of the InChI that is not human-readable. The InChI Key specification was released in September 2007 in order to facilitate web searches for chemical compounds, since these were problematical with the full-length InChI. Automatically generated by ACD Name

Annex II: Substance identification data for CAS random sample

CAS	CAS Index Name
96846-76-3	Butanoic acid, 3-methyl-, 1,1-dimethyl-5-methyleneheptyl ester
96508-01-9	Fatty acids, soya, Bu esters
65850-54-6	1,3,2-Dioxaphosphorinane, 2,2'-[(2,5-dimethyl-1,4-phenylene)bis(methylene)]bis[5,5-dimethyl-, 2,2'-dioxide
60664-15-5	4-Thiazolidinecarboxylic acid, ethyl ester, (4R)-
58861-53-3	Pyridine, 2-(4-fluorophenyl)-
51741-78-7	Ethane(dithioic) acid, 2-(3-methyl-2(3H)-benzothiazolylidene)-, methyl ester
50649-56-4	Benzoic acid, 4-(octyloxy)-, 4-pentylphenyl ester
36587-93-6	Guanidine, N,N'''-1,4-butanediylbis-, sulfate (9CI)
32208-02-9	Ethanaminium, N-methyl-2-[(1-oxooctadecyl)oxy]-N,N-bis[2-[(1-oxooctadecyl)oxy]ethyl]-, methyl sulfate (1:1)
31250-06-3	4,7,13,18-Tetraoxa-1,10-diazabicyclo[8.5.5]eicosane
31024-35-8	1-Propanamine, 3-(dimethoxymethylsilyl)-N-methyl-
16110-09-1	Pyridine, 2,5-dichloro-
16011-97-5	1,4-Butanediamine, N1,N4-dimethyl-
9054-63-1	Aminopeptidase, microsomal
6481-55-6	4H-1-Benzopyran-2-carboxylic acid, 5,7-dihydroxy-4-oxo-3-phenyl-, monosodium salt (8CI,9CI)
6270-55-9	Propanoic acid, 2-methyl-, 2-furanylmethyl ester
4320-30-3	L-Glutamic acid, compd. with L-arginine (1:1)
4008-41-7	1H-Imidazole-1-ethanamine, 2-(8-heptadecenyl)-N-[2-[2-(8-heptadecenyl)-4,5-dihydro-1H-imidazol-1-yl]ethyl]-4,5-dihydro- (9CI)
2627-95-4	Disiloxane, 1,3-diethenyl-1,1,3,3-tetramethyl-
2012-21-7	Phenol, 2,4-bis(2-phenylethenyl)-
1877-77-6	Benzenemethanol, 3-amino-
1841-19-6	1,3,8-Triazaspiro[4.5]decan-4-one, 8-[4,4-bis(4-fluorophenyl)butyl]-1-phenyl-
456-24-6	Pyridine, 2-fluoro-5-nitro-
441-61-2	3-Buten-2-amine, N-ethyl-N-methyl-4,4-di-2-thienyl-
315-37-7	Androst-4-en-3-one, 17-[(1-oxoheptyl)oxy]-, (17b)-
100063-55-6	Cuprate(1-), [9,16,23-tris(aminosulfonyl)-29H,31H-phthalocyanine-2-sulfonato(3-)-kN29,kN30,kN31,kN32]-, ammonium, (SP-4-2)- (9CI)
97171-76-1	Octanoic acid, oxybis(2,1-ethanediyloxy-2,1-ethanediyl) ester (9CI)
94406-05-0	Bicyclo[2.2.1]heptane-1-methanesulfonic acid, 7,7-dimethyl-2-oxo-, (1S)-, compd. with [7S-(7a,7aa,14a,14ab)]-dodecahydro-7,14-methano-2H,6H-dipyrido[1,2-a:1',2'-e][1,5]diazocine (2:1) (9CI)
94230-75-8	Phosphonic acid, [[(2-ethylhexyl)imino]bis(methylene)]bis-, dipotassium salt (9CI)
58505-58-1	Carbamimidothioic acid, 3-(trimethoxysilyl)propyl ester, hydrochloride (1:1)
52286-59-6	b-D-Glucopyranoside, (3b,6a,12b)-20-(b-D-glucopyranosyloxy)-3,12-dihydroxydammar-24-en-6-

CAS	CAS Index Name
	yl 2-O-(6-deoxy-a-L-mannopyranosyl)-
45296-12-6	9-Octadecenoic acid (9Z)-, 2-(diethylamino)ethyl ester
41981-72-0	Thiazole, 4,5-dimethyl-2-propyl-
39007-81-3	Pregna-1,4-diene-3,11,20-trione, 17-hydroxy-21-[(1-oxohexadecyl)oxy]- (9CI)
37672-83-6	Benzenesulfonic acid, 4-[4,5-dihydro-3-methyl-5-oxo-4-(2-phenyldiazenyl)-1H-pyrazol-1-yl]-, barium salt (2:1)
29122-68-7	Benzeneacetamide, 4-[2-hydroxy-3-[(1-methylethyl)amino]propoxy]-
19438-61-0	1,3-Isobenzofurandione, 5-methyl-
16245-97-9	Oxirane, 2-[(octadecyloxy)methyl]-
15219-34-8	Ethanedioyl dibromide
13973-45-0	Phosphonic acid, [(decylimino)bis(methylene)]bis-, trisodium salt (9CI)
6208-26-0	Distannoxane, 1,1,1,3,3,3-hexakis(2-methylpropyl)-
4498-67-3	1H-Indazole-3-carboxylic acid
3769-57-1	Ethanol, 2,2'-[[4-[2-(2-chloro-4-nitrophenyl)diazenyl]-3-methylphenyl]imino]bis-
3542-44-7	1-Propanesulfonic acid, 3-hydroxy-, sodium salt (1:1)
3034-42-2	1H-Imidazole, 1-methyl-5-nitro-
2816-43-5	Germane, triphenyl-
1583-83-1	Benzene, 1-(difluoromethoxy)-4-methyl-
603-72-5	2H-Naphth[1,8-cd]isothiazole, 1,1-dioxide
463-71-8	Carbonothioic dichloride
125-40-6	2,4,6(1H,3H,5H)-Pyrimidinetrione, 5-ethyl-5-(1-methylpropyl)-
95465-96-6	Spinels, chromium cobalt blue green
94201-16-8	Pyrazine, 2-(1-methylethyl)-3-[(2-methyl-3-thienyl)thio]-
93924-52-8	Hydrocarbons, C18
93918-56-0	9,12-Octadecadienamide, N,N'-[iminobis(2,1-ethanediylimino-2,1-ethanediyl)]bis- (9CI)
62546-04-7	Benzenesulfonic acid, 2-[[5-(aminocarbonyl)-1-ethyl-1,6-dihydro-2-hydroxy-4-methyl-6-oxo-3-pyridinyl]azo]-4-[(2,3-dibromo-1-oxopropyl)amino]- (9CI)
61129-40-6	Cadmate(2-), di-m-chlorotetrachlorodi-, barium (1:1)
60312-95-0	Ethanone, 1-(4'-bromo[1,1'-biphenyl]-4-yl)-2-phenyl-
51591-38-9	Butanedioic acid, 2,3-bis(acetyloxy)-, (2R,3R)-
39980-90-0	Cytidine, cyclic 2',3'-(hydrogen phosphate), ion(1-), N,N,N-triethylethanaminium (9CI)
27287-73-6	[1,2,4]Triazolo[1,5-a]pyrimidin-5(1H)-one, 2-amino-6-methyl-
15853-35-7	Phosphonium, triphenyl(phenylmethyl)-
13843-81-7	Chromic acid (H2Cr2O7), dilithium salt (8CI,9CI)
10597-89-4	Muramic acid, N-acetyl-
10428-19-0	Distannoxane, 1,1,3,3-tetrabutyl-1,3-dichloro-
7785-23-1	Silver bromide (AgBr)

CAS	CAS Index Name
7195-47-3	1,2,4,5-Benzenetetracarboxylic acid, tetrakis(oxiranylmethyl) ester (9CI)
6627-74-3	2-Cyclohexene-1-methanol, 2,6,6-trimethyl-
6417-51-2	Diindolo[3,2,1-de:3',2',1'-ij][1,5]naphthyridine-6,13-dione, 7,14-diphenyl-
4282-31-9	2,5-Thiophenedicarboxylic acid
3207-09-8	2-Butenedioic acid (2E)-, 1-butyl 4-(2-hydroxyethyl) ester
1324-29-4	Anthrazinesulfonic acid, 5,6,9,14,15,18-hexahydro-5,9,14,18-tetraoxo-, sodium salt (1:1)
403-41-8	Benzenemethanol, 4-fluoro-a-methyl-
122-51-0	Ethane, 1,1',1''-[methylidynetris(oxy)]tris-
79-37-8	Ethanedioyl dichloride
75-65-0	2-Propanol, 2-methyl-
94333-53-6	Piperazine, 1-[(7-bromo-3-hydroxy-2-naphthalenyl)carbonyl]-4-methyl-, monohydrochloride (9CI)
94248-22-3	2-Propenoic acid, 2-methyl-, 2-[[[[[3-[[[3-hydroxy-2,2-bis[[(2-methyl-1-oxo-2-propenyl)oxy]methyl]propoxy]carbonyl]amino]methylphenyl]amino]carbonyl]oxy]methyl]-2-(hydroxymethyl)-1,3-propanediyl ester (9CI)
94086-60-9	Butanedioic acid, (tetrapropenyl)-, disodium salt (9CI)
93894-07-6	Phenol, 2-nonyl-, cadmium salt (9CI)
64741-95-3	Residual oils (petroleum), solvent deasphalted
56378-59-7	Zinc, (2(3H)-benzothiazolethione)dichloro- (9CI)
41344-13-2	Butanedioic acid, sulfo-, 4-[2-[[3-[(2-hydroxydodecyl)methylamino]-1-oxopropyl](2-hydroxyethyl)amino]ethyl] ester, disodium salt (9CI)
37466-21-0	L-Arginine, 2-(acetyloxy)benzoate (1:1)
32643-00-8	Benzoic acid, 3-(aminosulfonyl)-5-(butylamino)-4-phenoxy-, butyl ester
32503-27-8	1-Butanaminium, N,N,N-tributyl-, sulfate (1:1)
18300-91-9	Pentadecanenitrile
14931-83-0	Cobaltate(2-), [[N,N'-1,2-ethanediylbis[N-[(carboxy-kO)methyl]glycinato-kN,kO]](4-)]-, (OC-6-21)-
14321-27-8	Benzenemethanamine, N-ethyl-
12777-45-6	Bismuth tin oxide
10296-76-1	1-Propanesulfonic acid, 2,3-dihydroxy-
5663-71-8	Benzamide, 2-(acetyloxy)-
3414-62-8	Inosine, oxime (9CI)
2784-89-6	1,4-Benzenediamine, 2-nitro-N1-phenyl-
2051-32-3	3-Octanol, 3-ethyl-
2043-43-8	Propanamide, 2-hydroxy-
557-27-7	Propanoic acid, magnesium salt (9CI)
272-14-0	Thieno[3,2-c]pyridine
136-47-0	Benzoic acid, 4-(butylamino)-, 2-(dimethylamino)ethyl ester, hydrochloride (1:1)

CAS	CAS Index Name
99-93-4	Ethanone, 1-(4-hydroxyphenyl)-
67-47-0	2-Furancarboxaldehyde, 5-(hydroxymethyl)-
98903-85-6	Pelargonium graveolens, ext., sapond., acetylated
97926-05-1	Alcohols, C12-14-secondary, b-(2-hydroxyethoxy)
94248-50-7	Benzene, dinitro-1,2-bis(2,4,6-trinitrophenoxy)- (9CI)
93925-44-1	Silicic acid (H4SiO4), tetrakis(2-methoxyethyl) ester, reaction products with bis(acetyloxy)dibutylstannane
65520-45-8	Butanedioic acid, 1,4-bis[2-[2-(2-butoxyethoxy)ethoxy]ethyl] ester
65152-19-4	2-Naphthalenesulfonic acid, 6-amino-4-hydroxy-5-[2-[2-(trifluoromethyl)phenyl]diazenyl]-, compd. with 2,2'-iminobis[ethanol] (1:1)
64742-36-5	Distillates (petroleum), clay-treated heavy paraffinic
63753-10-6	Benzaldehyde, 2-hydroxy-5-nonyl-
63216-88-6	Benzenesulfonic acid, 2,2'-thiobis[5-nitro-
59850-84-9	L-Proline, 5-oxo-, 2,4,6-trichlorophenyl ester
19150-71-1	Butanoic acid, 3-oxo-, octadecyl ester
14343-71-6	2-Oxazolidinone, 5-(4-morpholinylmethyl)-3-[[(5-nitro-2-furanyl)methylene]amino]-, (2R,3R)-2,3-dihydroxybutanedioate (1:1)
11092-86-7	6-Thiadecaborate(1-), dodecahydro-, cesium (1:1)
9025-64-3	Nuclease, deoxyribo-, II
9000-90-2	Amylase, a-
7787-49-7	Beryllium fluoride (BeF2)
7542-37-2	D-Streptamine, O-2,6-diamino-2,6-dideoxy-b-L-idopyranosyl-(1->3)-O-b-D-ribofuranosyl-(1->5)-O-[2-amino-2-deoxy-a-D-glucopyranosyl-(1->4)]-2-deoxy-
7392-72-5	1,3-Propanediaminium, N1,N1,N1,N3,N3-pentamethyl-N3-octadecyl-, chloride (1:2)
6375-27-5	Butanamide, N-(2,5-dimethoxyphenyl)-3-oxo-
4316-93-2	Pyrimidine, 4,6-dichloro-5-nitro-
1888-75-1	Lithium, (1-methylethyl)- (9CI)
661-54-1	1-Propyne, 3,3,3-trifluoro-
594-11-6	Cyclopropane, methyl-
137-06-4	Benzenethiol, 2-methyl-
80-65-9	2-Oxazolidinone, 3-amino-
100402-61-7	D-Glucose, ether with 1,6-hexanediol
94291-87-9	Undecane, 1-[(3,7-dimethyl-6-octenyl)oxy]-1-methoxy- (9CI)
94021-55-3	Chromate(5-), bis[3-hydroxy-4-[(2-hydroxy-6-sulfo-1-naphthalenyl)azo]-7-nitro-1-naphthalenesulfonato(4-)]-, pentahydrogen (9CI)
93857-21-7	L-Proline, 5-oxo-, compd. with 1,2-dihydro-1,5-dimethyl-2-phenyl-3H-pyrazol-3-one (1:1) (9CI)
63373-82-0	Bicyclo[2.2.1]heptane-2-methanol, 3,3-dimethyl-, (1S,2S,4R)-
51727-47-0	2-Propenoic acid, 2-methyl-, 2-[[(phenylamino)carbonyl]oxy]ethyl ester

CAS	CAS Index Name
51026-28-9	Carbamodithioic acid, (hydroxymethyl)methyl-, monopotassium salt (9CI)
40239-01-8	2-Heptanone, 4,6,6-trimethyl-
36323-28-1	Benzene, 1,3-bis(dibromomethyl)-
31351-20-9	Benzeneethanaminium, N-dodecyl-b-hydroxy-N,N,a-trimethyl-, bromide (9CI)
29349-67-5	Benzene, dimethyl[(methylphenyl)sulfonyl]-
21752-31-8	Butanoic acid, 2-amino-4-(S-methylsulfonimidoyl)-, [R-(R*,S*)]- (9CI)
19961-27-4	2-Propanamine, N-ethyl-
14815-59-9	1(3H)-Isobenzofuranone, 3-(4-hydroxyphenyl)-3-[4-(phosphonooxy)phenyl]-, compd. with cyclohexanamine (1:2) (9CI)
12222-60-5	2H-Naphtho[1,2-d]triazole-5-sulfonic acid, 2,2'-[1,1-diazenediylbis[(2-sulfo-4,1-phenylene)-2,1-ethenediyl(3-sulfo-4,1-phenylene)]]bis-, sodium salt (1:6)
10039-64-2	Benzenepropanoic acid, 3-amino-, ethyl ester
7439-69-2	Phosphonic acid, [(diethylamino)carbonyl]-, dibutyl ester (9CI)
4113-15-9	Peroxydicarbonic acid, bis(3,5,5-trimethylhexyl) ester (7CI,8CI,9CI)
3249-28-3	2,4-Hexadienedial
2583-25-7	Propanedioic acid, 2-(2-propen-1-yl)-
1613-66-7	Germane, dichlorodiphenyl-
615-42-9	Benzene, 1,2-diiodo-
157-06-2	D-Arginine
95-24-9	2-Benzothiazolamine, 6-chloro-
51-55-8	Benzeneacetic acid, a-(hydroxymethyl)- (3-endo)-8-methyl-8-azabicyclo[3.2.1]oct-3-yl ester
97692-41-6	Chromate(1-), [m-[4-[[5-[(5-chloro-2-hydroxyphenyl)azo]-2,4-dihydroxyphenyl]azo]-3-hydroxy-7-nitro-1-naphthalenesulfonato(5-)]]dihydroxydi-, sodium (9CI)
97553-38-3	Mentha sativa, ext.
93894-75-8	2,7-Naphthalenedisulfonic acid, 3,3'-[azoxybis[[3-(2-hydroxyethoxy)-4,1-phenylene]azo]]bis[4-hydroxy-5-[[(4-methylphenyl)sulfonyl]amino]-, tetrasodium salt (9CI)
62265-99-0	Benzene, 1,3-dibromo-2-methoxy-4-methyl-5-nitro-
58085-76-0	Phenol, 3,5-dinonyl-
58061-48-6	2-Pyridinemethanamine, N-butyl-
51810-70-9	Zinc phosphide
40596-46-1	Dodecanoic acid, 2-octyl-
32815-96-6	Acetic acid, bromo-, 2-nitrobutyl ester (8CI,9CI)
30273-11-1	Benzenamine, 4-(1-methylpropyl)-
29089-67-6	Isononaneperoxoic acid, 1,1'-(1,1,4,4-tetramethyl-1,4-butanediyl) ester
17676-33-4	Spirosta-5,25(27)-diene-1,3-diol, (1b,3b)-
17094-21-2	Butanoic acid, 2-methyl-3-oxo-, methyl ester
10342-06-0	L-Serine, N-[(1,1-dimethylethoxy)carbonyl]-, compd. with N-cyclohexylcyclohexanamine (1:1) (9CI)

CAS	CAS Index Name
7147-77-5	2-Furancarboxaldehyde, 5-(4-nitrophenyl)-
6117-91-5	2-Buten-1-ol
6107-56-8	Octanoic acid, calcium salt (2:1)
4351-70-6	Phosphonic acid, [1-[[(2-chloroethoxy)(2-chloroethyl)phosphinyl]oxy]ethyl]-, 1-[bis(2-chloroethoxy)phosphinyl]ethyl 2-chloroethyl ester
3301-90-4	2H-Pyran-2-one, 6-ethyltetrahydro-
2083-32-1	Thiophene, 2,5-dihydro-3-(4-methyl-3-penten-1-yl)-, 1,1-dioxide
1738-69-8	L-Leucine, phenylmethyl ester
1574-34-1	1-Pentyne, 3-methylene-
536-60-7	Benzenemethanol, 4-(1-methylethyl)-
512-56-1	Phosphoric acid, trimethyl ester
70-22-4	2-Pyrrolidinone, 1-[4-(1-pyrrolidinyl)-2-butyn-1-yl]-
94291-99-3	10H-Phenothiazine-2-carbonitrile, 10-[3-(4-hydroxy-1-piperidinyl)propyl]-, monomethanesulfonate (salt) (9CI)
94108-11-9	Benzoic acid, 2-(acetyloxy)-4-[[(dimethylamino)carbonyl]thio]-, methyl ester
94088-58-1	Benzoic acid, 2-[(phenylsulfonyl)amino]-, monosodium salt (9CI)
94042-72-5	2-Naphthalenesulfonic acid, 6-amino-4-hydroxy-5-[[4-[[(4-methyl-2-sulfophenyl)amino]sulfonyl]phenyl]azo]- (9CI)
93917-70-5	2-Propen-1-ol, 3-bicyclo[2.2.1]hept-5-en-2-yl-2-methyl-
64935-37-1	2-Heptenal, 2-propyl-, (2E)-
59686-69-0	Hexanedioic acid, diisohexadecyl ester (9CI)
37723-78-7	Butanoic acid, 2-[[2-[3-(acetylamino)-2,4,6-triiodophenoxy]ethoxy]methyl]-
32138-95-7	Aluminum, tris[1-(hydroxy-kO)-2(1H)-pyridinethionato-kS2]- (9CI)
22819-91-6	Heptanenitrile, 7-chloro-
14832-14-5	Copper, [1,2,3,4,8,9,10,11,15,16,17,18,22,23,24,25-hexadecachloro-29H,31H-phthalocyaninato(2-)-kN29,kN30,kN31,kN32]-, (SP-4-1)-
13296-76-9	9,11,13-Octadecatrienoic acid
13106-44-0	Ethanaminium, N,N,N-trimethyl-2-[(1-oxo-2-propen-1-yl)oxy]-, methyl sulfate (1:1)
12585-23-8	RNA (Saccharomyces cerevisiae valine-specific tRNA 1) (9CI)
10374-74-0	7-Tetradecene
7803-63-6	Sulfuric acid, ammonium salt (1:1)
6361-46-2	2-Naphthalenesulfonic acid, 7-[(4-amino-3-methylbenzoyl)amino]-4-hydroxy-
2836-04-6	1,3-Benzenediamine, N1,N1-dimethyl-
2006-05-5	2-Hexene, 6,6-dimethoxy-2-methyl-
771-56-2	Benzene, 1,2,3,4,5-pentafluoro-6-methyl-
189-96-8	Benzo[pqr]picene
105-48-6	Acetic acid, 2-chloro-, 1-methylethyl ester

CAS	CAS Index Name
102-19-2	Benzeneacetic acid, 3-methylbutyl ester
98-67-9	Benzenesulfonic acid, 4-hydroxy-
73-49-4	6-Quinazolinesulfonamide, 7-chloro-2-ethyl-1,2,3,4-tetrahydro-4-oxo-
95370-51-7	Carbamic acid, [2-(sulfothio)ethyl]-, C-(g-w-perfluoro-C6-9-alkyl) esters, monosodium salts
94248-98-3	Acetamide, N-[6-[2-(5-nitro-2-furanyl)ethenyl]-3-pyridazinyl]-, monohydrochloride (9CI)
94022-62-5	Bicyclo[4.2.0]octan-2-ol, 4,4,6-trimethyl-, acetate (9CI)
94021-14-4	Phenol, 2,2'-methylenebis[4-(1,1-dimethylethyl)-3,6-dimethyl-
93904-64-4	Benzaldehyde, 3-chloro-4-hydroxy-5-(1-methylethyl)-
63451-33-2	Glycine, N,N'-1,2-cyclohexanediylbis[N-(carboxymethyl)-, magnesium sodium salt (1:1:2)
50598-33-9	2,5-Furandione, 3-ethyldihydro-3-methyl-
38775-22-3	Benzenesulfonic acid, 2,2'-([1,1'-biphenyl]-4,4'-diyldi-2,1-ethenediyl)bis-
35203-93-1	1H-Indene-5-sulfonamide, 2,3-dihydro-
28753-07-3	Ethanesulfinic acid, 1-chloro-
15782-05-5	2-Naphthalenecarboxylic acid, 4-[2-(5-chloro-4-methyl-2-sulfophenyl)diazenyl]-3-hydroxy-, strontium salt (1:1)
14373-91-2	Sulfuric acid, dysprosium(3+) salt (3:2)
12262-58-7	Cyclohexanone, peroxide
12200-88-3	Vanadate (V10O286-), sodium (1:6)
10353-05-6	Pyridinium, 1-[[(6R,7R)-2-carboxy-8-oxo-7-[(2-thienylacetyl)amino]-5-thia-1-azabicyclo[4.2.0]oct-2-en-3-yl]methyl]-, thiocyanate (9CI)
7440-66-6	Zinc
3982-87-4	Phosphine sulfide, tris(2-methylpropyl)-
3003-38-1	Benzoic acid, 4-[(4,5-dihydro-3-methyl-5-oxo-1-phenyl-1H-pyrazol-4-yl)azo]-, phenylmethyl ester (9CI)
2432-91-9	Butanethioic acid, 3-methyl-, S-(1-methylpropyl) ester
1519-46-6	Phosphonium, 1,1'-[1,2-phenylenebis(methylene)]bis[1,1,1-triphenyl-, bromide (1:2)
387-79-1	Pregn-5-en-20-one, 3,17-dihydroxy-, (3b)-
138-41-0	Benzoic acid, 4-(aminosulfonyl)-
96-53-7	2-Thiazolidinethione
94-75-7	Acetic acid, 2-(2,4-dichlorophenoxy)-
71-55-6	Ethane, 1,1,1-trichloro-
94293-51-3	7,14-Methano-2H,6H-dipyrido[1,2-a:1',2'-e][1,5]diazocine, dodecahydro-, dihydriodide, [7S-(7a,7aa,14a,14ab)]- (9CI)
94108-93-7	2-Naphthalenesulfonic acid, 7-[[3-(acetylamino)benzoyl]amino]-4-hydroxy-3-[[2-methyl-4-[(2-methyl-4-sulfophenyl)azo]phenyl]azo]-, sodium salt (9CI)
93924-61-9	Hydrocarbons, C20-50, residual oil hydrogenation vacuum distillate
64741-39-5	Tallow, oil, lithium salts
63468-98-4	Benzenediazonium, 4-(ethoxycarbonyl)-, sulfate (2:1)

CAS	CAS Index Name
57448-83-6	L-Ascorbic acid, compd. with N-(2,5-dioxo-4-imidazolidinyl)urea (1:1)
54618-06-3	Uridine, 5'-benzoate (9CI)
54043-73-1	1-Cyclopentene-1-propanol, b,2-dimethyl-5-(1-methylethenyl)-, 1-acetate
29952-87-2	Propanoic acid, 2-(4-chlorophenoxy)-2-methyl-, compd. with 5-hydroxy-6-methyl-3,4-pyridinedimethanol (1:1)
28600-36-4	Sulfurous acid, gold potassium salt (8CI,9CI)
15782-04-4	Benzenesulfonic acid, 4-[2-(2-hydroxy-1-naphthalenyl)diazenyl]-, barium salt (2:1)
14850-23-8	4-Octene, (4E)-
13348-41-9	1,6-Hexanediamine, N,N'-dicyclohexyl- (7CI,8CI,9CI)
10602-34-3	2-Butene, 1,1-diethoxy-
7783-66-6	Iodine fluoride (IF5)
7459-75-8	3,6-Acridinediamine, hydrochloride (9CI)
5953-76-4	2-Butenoic acid, 2-methyl-, methyl ester, (2Z)-
5332-24-1	Quinoline, 3-bromo-
5144-42-3	L-Ornithine, 2-oxopentanedioate (2:1) (9CI)
1707-92-2	Phosphoric acid, tris(phenylmethyl) ester
1492-18-8	L-Glutamic acid, N-[4-[[(2-amino-5-formyl-1,4,5,6,7,8-hexahydro-4-oxo-6-pteridinyl)methyl]amino]benzoyl]-, calcium salt (1:1)
387-59-7	Pregn-4-ene-3,20-dione, 6-fluoro-11,17,21-trihydroxy-16-methyl-, (6a,11b,16a)- (9CI)
194-59-2	7H-Dibenzo[c,g]carbazole
106-29-6	Butanoic acid, (2E)-3,7-dimethyl-2,6-octadien-1-yl ester
79-50-5	2(3H)-Furanone, dihydro-3-hydroxy-4,4-dimethyl-
95370-50-6	2-Butenedioic acid (2Z)-, di-C13-18-alkyl esters
93982-15-1	Benzenemethanaminium, ar-dodecyl-N,N-dimethyl-N-[2-[(1-oxooctadecyl)oxy]ethyl]-, chloride (9CI)
93859-04-2	Benzene, 1-dodecyl-2,4-diisocyanato-
64491-92-5	Mercurate(1-), [metasilicato(2-)-kO](2-methoxyethyl)-, hydrogen (9CI)
62005-65-6	1H-Inden-5-ol, 2,3-dihydro-6-isononyl-1,1,3,3-tetramethyl- (9CI)
61919-18-4	Benzenediazonium, 2-methoxy-5-nitro-, (T-4)-tetrachlorozincate(2-) (2:1)
61788-36-1	Octadecanoic acid, epoxy derivs., Bu esters
60029-23-4	Formamide, N-[2,2,2-trichloro-1-(4-morpholinyl)ethyl]-
37312-62-2	Proteinase, Serratia extracellular
31635-99-1	Propanedioic acid, dihydroxy-, disodium salt (9CI)
25751-57-9	Antimony, tris(2,4-pentanedionato-kO,kO')-, (OC-6-11)- (9CI)
20543-04-8	Octanoic acid, copper salt (1:?)
12115-66-1	Magnesium carbonate hydroxide (Mg4(CO3)3(OH)2)
10112-15-9	Benzenamine, N-ethyl-2-nitro-

CAS	CAS Index Name
7778-18-9	Sulfuric acid, calcium salt (1:1)
7059-23-6	Hydrazinecarboximidamide, 2,2'-(1-methyl-1,2-ethanediylidene)bis-, hydrochloride (1:2)
7023-01-0	1,9,10-Octadecanetriol
02.08.6856	Benzo[a]phenazinium, 9-[(3-methoxyphenyl)amino]-7-phenyl-5-(phenylamino)-4,10-disulfo-, inner salt, sodium salt (1:1)
6396-93-6	Benzenepropanamine, a-(2-methylpropyl)-
2765-18-6	Naphthalene, 1-propyl-
1941-24-8	Methanaminium, N,N,N-trimethyl-, nitrate (1:1)
1115-59-9	L-Alanine, ethyl ester, hydrochloride (1:1)
624-49-7	2-Butenedioic acid (2E)-, 1,4-dimethyl ester
92-48-8	2H-1-Benzopyran-2-one, 6-methyl-
56-84-8	L-Aspartic acid
101357-25-9	Phenol, 2,4,6-trinitro-, reaction products with 2,4-dinitrophenol and sodium sulfide (Na2(Sx))
94465-73-3	Equisetum debile, ext.
65996-65-8	Iron ores, agglomerates
65036-95-5	Molybdate (Mo8O264-), hydrogen, compd. with 1,3,5-triazine-2,4,6-triamine (1:4:4)
61211-10-7	Butanedioic acid, sulfo-, 1,4-bis(trimethylhexyl) ester, sodium salt (9CI)
58962-45-1	Phenol, 4-chloro-3,5-dimethyl-, potassium salt (9CI)
36475-52-2	Phosphonic acid, P,P',P",P"'-[1,2-ethanediylbis[[(phosphonomethyl)imino]-2,1-ethanediylnitrilobis(methylene)]]tetrakis-
28925-00-0	Ethanone, 1-cyclododecyl-
27496-76-0	Benzene, ethenyldimethyl-
25574-69-0	Difuro[3,4-b:3',4'-d]furan-1,3,5,7-tetrone, tetrahydro-
24032-65-3	Acetamide, N-[2-[[4-[3-(4-chlorophenyl)-4,5-dihydro-1H-pyrazol-1-yl]phenyl]sulfonyl]ethyl]-
23580-54-3	Pyridine, 5-ethyl-2-propyl-
15284-51-2	Tetradecanoic acid, calcium salt (2:1)
9080-55-1	Proteinase, Aspergillus oryzae neutral
8002-29-7	Tar oils
6638-02-4	Benzenemethanaminium, N-ethyl-N-[4-[[4-[ethyl[(3-sulfophenyl)methyl]amino]phenyl]phenylmethylene]-2,5-cyclohexadien-1-ylidene]-3-sulfo-, inner salt (9CI)
6283-67-6	2-Butenedioic acid (2E)-, nickel(2+) salt (1:1)
3089-24-5	1,6-Hexanediol, 2,2,4-trimethyl-
2611-61-2	Glycine, (2R,3R)-2-[(2,2-dichloroacetyl)amino]-3-hydroxy-3-[4-(methylsulfonyl)phenyl]propyl ester, hydrochloride (1:1)
1553-34-0	Piperidine, 1-methyl-3-(9H-thioxanthen-9-ylmethyl)-, hydrochloride (1:1)
1327-44-2	Silicic acid, aluminum potassium salt
1080-06-4	L-Tyrosine, methyl ester

CAS	CAS Index Name
897-18-7	1,3,4-Oxadiazole, 2-(1-naphthalenyl)-5-phenyl-
865-31-6	Methanol, aluminum salt (8CI,9CI)
326-56-7	1,3-Benzodioxole-5-carboxylic acid, methyl ester
99948-91-1	Tannins, reaction products with ethylenediamine and formaldehyde, acetates (salts)
93919-87-0	Glycine, N-[2-[(carboxymethyl)(2-hydroxyethyl)amino]ethyl]-N-(1-oxo-9-octadecenyl)-, disodium salt, (Z)- (9CI)
65581-12-6	1-Aziridineacetic acid, a-ethyl-, (1-methylethylidene)bis[(2-bromo-4,1-phenylene)oxy-2,1-ethanediyl] ester (9CI)
65156-97-0	1,5-Dioxaspiro[5.5]undecane, 9-(1,1-dimethylethyl)-
54229-13-9	1-Propanaminium, N-[2-[[4-[2-(2-chloro-4-nitrophenyl)diazenyl]phenyl]ethylamino]ethyl]-2-hydroxy-N,N-dimethyl-, chloride (1:1)
39431-98-6	Aluminum, m-bromo-m-hydroxytetrahydroxydi-
37187-22-7	2,4-Pentanedione, peroxide
37087-96-0	Benzoic acid, 2-chloro-5-[(3,5-dimethyl-1-piperidinyl)sulfonyl]-, trans- (9CI)
32353-63-2	Benzoxazolium, 3-(2-carboxyethyl)-2,5-dimethyl-, bromide (8CI,9CI)
26968-58-1	Benzene, (chloromethyl)ethyl-
19430-93-4	1-Hexene, 3,3,4,4,5,5,6,6,6-nonafluoro-
17746-05-3	Undecanoyl chloride
10507-50-3	13-Docosenoic acid, 1-methylethyl ester, (13Z)-
9028-26-6	Dehydrogenase, uridine diphosphoglucose
9000-97-9	Aminotransferase, aspartate
6469-93-8	1-Propanamine, 3-(2-chloro-9H-thioxanthen-9-ylidene)-N,N-dimethyl-, hydrochloride, (3Z)- (9CI)
5219-61-4	Acetonitrile, 2-(phenylthio)-
2740-83-2	Benzenemethanamine, 3-(trifluoromethyl)-
2719-64-4	Benzene, (1-propylnonyl)-
2032-65-7	Phenol, 3,5-dimethyl-4-(methylthio)-, 1-(N-methylcarbamate)
1304-54-7	Beryllium nitride (Be3N2)
603-35-0	Phosphine, triphenyl-
547-60-4	Bicyclo[3.1.1]heptan-3-one, 2,6,6-trimethyl-, (1R,2S,5S)-rel-
539-66-2	Butanoic acid, 3-methyl-, sodium salt (9CI)
77-39-4	1-Piperidinepropanol, a-cyclopentyl-a-phenyl-
97808-58-7	Sulfonic acids, C13-18-alkane, chloro
97752-26-6	1,2-Propanediol, 3-[2-hydroxy-3-(9-octadecenyloxy)propoxy]-, (Z)-
97722-24-2	Tar bases, coal, crude, chlorotetradecane-quaternized
97489-45-7	Phosphoric acid, mixed Bu and iso-Bu esters, compds. with ethanolamine
94552-21-3	Cyclohexane, oxidized, dehydrogenated, distn. residues
94166-45-7	Naphtho[1,2-d]thiazolium, 2-[(3-ethyl-5-methoxy-2(3H)-benzothiazolylidene)methyl]-1-(3-

CAS	CAS Index Name
	sulfopropyl)-, inner salt
63870-30-4	Benzoxazolium, 3-ethyl-2-[2-(phenylamino)ethenyl]-, iodide (1:1)
60580-61-2	1,3-Benzenedicarboxylic acid, 5-nitro-, zinc salt (1:1)
59970-93-3	1,5-Naphthalenedisulfonic acid, 2-[[6-[(4-amino-6-chloro-1,3,5-triazin-2-yl)methylamino]-1-hydroxy-3-sulfo-2-naphthalenyl]azo]-, disodium salt (9CI)
56744-60-6	2-Propenoic acid, 2-methyl-, 1,1'-[(1-methylethylidene)bis(4,1-phenyleneoxy-2,1-ethanediyloxy-2,1-ethanediyl)] ester
54503-10-5	1-Pyrrolidinecarboxylic acid, 2-(aminocarbonyl)-, 1,1-dimethylethyl ester
18029-54-4	5H-Dibenzo[a,d]cyclohepten-5-ol, 5-[3-(dimethylamino)propyl]-
17773-10-3	Ethanaminium, 2-hydroxy-N,N,N-trimethyl-, iodide (1:1)
16974-11-1	9-Dodecen-1-ol, 1-acetate, (9Z)-
15429-36-4	1,2-Hydrazinedicarboxylic acid, 1,2-dimethyl-, diethyl ester (9CI)
13736-22-6	Benzenesulfonic acid, 4-formyl-, sodium salt (9CI)
8011-61-8	Tyrocidine
7598-61-0	Phosphonic acid, P-(2,2-diethoxyethyl)-, diethyl ester
6385-60-0	7-Oxabicyclo[2.2.1]hept-5-ene-2,3-dicarboxylic acid, disodium salt, (endo,endo)- (9CI)
5796-17-8	D-Tyrosine, 3-hydroxy-
5311-88-6	Propanal, 2-phenylhydrazone
540-10-3	Hexadecanoic acid, hexadecyl ester
506-42-3	9-Octadecen-1-ol, (9E)-
338-83-0	1-Propanamine, 1,1,2,2,3,3,3-heptafluoro-N,N-bis(1,1,2,2,3,3,3-heptafluoropropyl)-
119-04-0	D-Streptamine, O-2,6-diamino-2,6-dideoxy-b-L-idopyranosyl-(1->3)-O-b-D-ribofuranosyl-(1->5)-O-[2,6-diamino-2,6-dideoxy-a-D-glucopyranosyl-(1->4)]-2-deoxy-
97659-42-2	4-Pyridinecarboxylic acid, 2-ethyldihydro-6-oxo- (9CI)
97375-13-8	2-Anthracenesulfonic acid, 1-amino-4-[[3-[[[(2,3-dichloro-6-quinoxalinyl)sulfonyl]amino]sulfonyl]-2,6-dimethylphenyl]amino]-9,10-dihydro-9,10-dioxo-, monosodium salt (9CI)
94139-24-9	9-Octadecenoic acid (9Z)-, diamide deriv. with N-(2-aminoethyl)-N'-[2-[(2-aminoethyl)amino]ethyl]-1,2-ethanediamine (9CI)
94042-77-0	Bicyclo[4.1.0]hept-3-ene-2-carboxaldehyde, 3,7,7-trimethyl-
93924-69-7	Quaternary ammonium compounds, benzyl-C12-14-alkyldimethyl, salts with monolauryl sulfobutanedioate, sodium salts
65087-04-9	Disulfide, 2,4-dimethylphenyl 2,6-dimethylphenyl
58031-81-5	1H-Imidazole, 4-nitro-, sodium salt (9CI)
55373-89-2	Pentacosanoic acid, methyl ester
51052-02-9	1,1,5,5-Pentanetetrol
25151-00-2	Dodecanoic acid, 1,1',1''-(butylstannylidyne) ester
17373-93-2	Acetic acid, (2-methylphenyl)methyl ester
16576-95-7	Benzenemethanaminium, N,N-dimethyl-N-undecyl-, chloride (9CI)

CAS	CAS Index Name
13575-60-5	Carbamodithioic acid, N,N-diethyl-, C,C'-[carbonylbis(iminomethylene)] ester
12592-70-0	Gallium strontium sulfide (Ga2SrS4)
10605-21-7	Carbamic acid, N-1H-benzimidazol-2-yl-, methyl ester
6104-56-9	1,3-Naphthalenedisulfonic acid, 7-[2-[4-[[4-[2-[6-[(4-aminobenzoyl)amino]-1-hydroxy-3-sulfo-2-naphthalenyl]diazenyl]benzoyl]amino]-2-methylphenyl]diazenyl]-, sodium salt (1:3)
5434-21-9	1,2-Benzenedicarboxylic acid, 4-amino-
4121-67-9	7-Benzothiazolesulfonic acid, 2-[2-(2,3-dihydro-1,3-dioxo-1H-inden-2-yl)-6-quinolinyl]-6-methyl-, sodium salt (1:1)
2391-28-8	3-Acridinamine, 9-(4-aminophenyl)-2-methyl-, nitrate (9CI)
998-00-5	Aluminum, tetrakis(2-methylpropyl)-m-oxodi-
526-55-6	1H-Indole-3-ethanol
485-41-6	Benzoic acid, 3,5-diamino-2-[2-[4-(aminosulfonyl)phenyl]diazenyl]-
112-63-0	9,12-Octadecadienoic acid (9Z,12Z)-, methyl ester
98-73-7	Benzoic acid, 4-(1,1-dimethylethyl)-
79-68-5	3-Buten-2-one, 4-(2,2,3-trimethyl-6-methylenecyclohexyl)-
97721-85-2	Amides, from C18-24 fatty acids and triethylenetetramine, reaction products with 1,6-diisocyanatohexane
96152-60-2	Zinc, 3-hydroxy-4-[(2-hydroxy-1-naphthalenyl)azo]-7-nitro-1-naphthalenesulfonate complexes
65689-55-6	Bis[1]benzothieno[3,2-e:2',3'-g]isobenzofuran-5,7-dione
62637-99-4	Cyclohexanebutanoic acid, lead(2+) salt (2:1)
61436-47-3	Titanium, tris(dioctyl phosphato-kO")(2-propanolato)-, (T-4)-
55324-97-5	D-Glucose, 6-amino-6-deoxy-, hydrochloride (1:1)
38133-90-3	2-Naphthalenecarboxamide, N-(2,3-dihydro-2-oxo-1H-benzimidazol-5-yl)-3-hydroxy-4-[2-(2-nitrophenyl)diazenyl]-
34866-46-1	Urea, [5-[2-[(1,1-dimethylethyl)amino]-1-hydroxyethyl]-2-hydroxyphenyl]-, monohydrochloride (9CI)
34662-31-2	Benzonitrile, 5-chloro-2-nitro-
34320-82-6	3H-Pyrazol-3-one, 2,4-dihydro-5-[(4-nitrophenyl)amino]-2-(2,4,6-trichlorophenyl)-
28950-66-5	Benzenesulfonic acid, 2,2'-(1,2-ethenediyl)bis[5-[[4-chloro-6-(4-morpholinyl)-1,3,5-triazin-2-yl]amino]-, disodium salt (9CI)
23786-14-3	Benzeneacetic acid, 4-methoxy-, methyl ester
17696-62-7	Benzoic acid, 4-hydroxy-, phenyl ester
15962-46-6	Benzeneacetic acid, a-(acetylamino)-
9031-72-5	Dehydrogenase, alcohol
7241-13-6	Methane, 1,1'-sulfonylbis[1,1,1-tribromo-
6713-54-8	Pyrimido[5,4-d]pyrimidine-2,4,6,8(3H,7H)-tetrone, 1,5-dihydro-
5426-09-5	4,7-Epoxyisobenzofuran-1,3-dione, 3a,4,7,7a-tetrahydro-
2896-60-8	1,3-Benzenediol, 4-ethyl-

CAS	CAS Index Name
2363-88-4	2,4-Decadienal
2281-11-0	1-Hexadecanaminium, N,N-dimethyl-N-(3-sulfopropyl)-, inner salt
1556-08-7	Cyclooctane, chloro-
874-87-3	Benzene, 1-(chloromethyl)-4-(methylthio)-
537-98-4	2-Propenoic acid, 3-(4-hydroxy-3-methoxyphenyl)-, (2E)-
432-24-6	2-Cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl-

Annex III: Deviating structures from ACD NTS

CAS	EINECS Name	CAS Name	NTS report	Comment
316-93-2	4,6-dichloro-5-nitro-pyridine	Pyrimidine, 4,6-dichloro-5-nitro-	ok	EINECS name error
547-60-4	(1alpha,2alpha,5alpha)-2,6,6-trimethylbicyclo[3.1.1]heptan-3-one	Bicyclo[3.1.1]heptan-3-one, 2,6,6-trimethyl-, (1R,2S,5S)-rel-	ok	NTS with EINECS name mismatch of one stereodescriptor; NTS with CAS name matches
29952-87-2	5-hydroxy-3,4-(hydroxymethyl)-6-methylpyridinium 2-(p-chlorophenoxy)-2-methylpropionate	Propanoic acid, 2-(4-chlorophenoxy)-2-methyl-, compd. with 5-hydroxy-6-methyl-3,4-pyridinedimethanol (1:1)	Dictionary	NTS from CAS: Name corresponds to the salt NTS from name ok
6117-91-5	but-2-en-1-ol	2-Buten-1-ol	Dictionary	NTS from CAS stereodescriptor specified (E); NTS from name ok (undefined double bond)
79-68-5	4-(2,2,3-trimethyl-6-methylenecyclohexyl)-3-buten-2-one	3-Buten-2-one, 4-(2,2,3-trimethyl-6-methylenecyclohexyl)-	Dictionary	NTS from CAS stereodescriptor specified (E); NTS from name ok (undefined double bond)
94086-60-9	disodium (tetrapropenyl)succinate	Butanedioic acid, (tetrapropenyl)-, disodium salt (	Warning Ambiguity is possible	Incompletely defined substance
63373-82-0	(1S-endo)-3,3-dimethylbicyclo[2.2.1]heptane-2-methanol	Bicyclo[2.2.1]heptane-2-methanol, 3,3-dimethyl-, (1S,2S,4R)-	Warning Some stereodescriptors ignored	NTS with CAS name matches
62637-99-4	lead bis(4-cyclohexylbutyrate)	Cyclohexanebutanoic acid, lead(2+) salt (2:1)	Warning Invalid Charge	NTS sets automatically Pb+4
37466-21-0	L-arginine monoacetylsalicylate	L-Arginine, 2-(acetyloxy)benzoate (1:1)	Warning Ambiguity is possible	NTS from CAS name ok
37087-96-0	trans-2-chloro-5-[(3,5-dimethylpiperidin-1-	Benzoic acid, 2-chloro-5-[(3,5-dimethyl-1-piperidinyl)sulfonyl	Warning Some stereodescriptors	NTS ignores stereodescriptors from EINECS & CAS name

CAS	EINECS Name	CAS Name	NTS report	Comment
	yl)sulphonyl]benzoic acid	]-, trans-	ignored	
31024-35-8	3-(dimethoxymethylsilyl)-N-methylpropylamine	1-Propanamine, 3-(dimethoxymethylsilyl)-N-methyl-	Warning structure 1 of 2	NTS from EINECS & CAS name: Structure 2 ok
26968-58-1	(chloromethyl)ethyl benzene	Benzene, (chloromethyl)ethyl-	Warning Ambiguity is possible	Incompletely defined substance
10353-05-6	(6R-trans)-1-[[2-carboxy-8-oxo-7-[(2-thienylacetyl)amino]-5-thia-1-azabicyclo[4.2.0]oct-2-en-3-yl]methyl]pyridinium thiocyanate	Pyridinium, 1-[[(6R,7R)-2-carboxy-8-oxo-7-[(2-thienylacetyl)amino]-5-thia-1-azabicyclo[4.2.0]oct-2-en-3-yl]methyl]-, thiocyanate	Warning Some stereodescriptors ignored	NTS from CAS name ok
3207-09-8	butyl hydroxyethyl fumarate	2-Butenedioic acid (2E)-, 1-butyl 4-(2-hydroxyethyl) ester	Warning Ambiguity is possible	EINECS name is ambigious & does not include stereodescriptor NTS from CAS name error
94248-50-7	dinitro-1,2-bis(2,4,6-trinitrophenoxy)benzene	Benzene, dinitro-1,2-bis(2,4,6-trinitrophenoxy)-	Warning Ambiguity is possible	Incompletely defined substance
93982-15-1	(dodecylbenzyl)dimethyl[2-[(1-oxooctadecyl)oxy]ethyl]ammonium chloride	Benzenemethanaminium, ar-dodecyl-N,N-dimethyl-N-[2-[(1-oxooctadecyl)oxy]ethyl]-, chloride	Warning Ambiguity is possible	Incompletely defined substance
61211-10-7	sodium 1,4-bis(trimethylhexyl) sulphonatosuccinate	Butanedioic acid, sulfo-, 1,4-bis(trimethylhexyl) ester, sodium salt (	Warning Ambiguity is possible	Incompletely defined substance

Annex IV: PRS parent compounds and predicted data

The following data compilation is available on the JRC-IHCP Computational Toxicology webpage as an sdf file (PRS_parent_substances.xls).

Field	Description
CAS Number	All PRS CAS, related to the Parent compound
SMILES Parent	The simplified molecular input line entry specification or SMILES is a specification for describing the structure of chemical molecules using short ASCII strings. Automatically generated by Pipeline Pilot
IUPAC Name	IUPAC names (English) automatically generated by ACD Name.
InChI_Parent	The IUPAC International Chemical Identifier is a textual identifier for chemical substances, designed to provide a standard and human-readable way to encode molecular information and to facilitate the search for such information in databases and on the web. Automatically generated by ACD Name
InChI_key_Parent	The InChI Key is a fixed length (25 characters) condensed digital representation of the InChI that is not human-readable. The InChI Key specification was released in September 2007 in order to facilitate web searches for chemical compounds, since these were problematical with the full-length InChI. Automatically generated by ACD Name

The following data compilation is available on the JRC-IHCP Computational Toxicology webpage as an sdf file (PRS_parents_EPISUITE.sdf).

Field	Description
CAS_Number(s)	All CAS_Numbers relating to the Parent substance
SMILES_Parent	The simplified molecular input line entry specification or SMILES is a specification for describing the structure of chemical molecules using short ASCII strings. Automatically generated by Pipeline Pilot
IUPAC_Name_Parent	IUPAC names (English) automatically generated by ACD Name.
InChI_Parent	The IUPAC International Chemical Identifier is a textual identifier for chemical substances, designed to provide a standard and human-readable way to encode molecular information and to facilitate the search for such information in databases and on the web. Automatically generated by ACD Name
InChI_key_Parent	The InChI Key is a fixed length (25 characters) condensed digital representation of the InChI that is not human-readable. The InChI Key specification was released in September 2007 in order to facilitate web searches for chemical compounds, since these were problematical with the full-length InChI. Automatically generated by ACD Name
logP	Octanol /Water partition coefficient
Log_BCF	Bioconcentration Factor fish(l/kg wet weight)
Log_Sol	Log Water Solubility (moles/l)
Sol	Water Solubility (est = estimated, data exp =experimental data)
BIOWIN	Estimation of persistence
BIOWIN1	Linear Model Prediction (does not biodegradate fast)
BIOWIN2	Non Linear Model Prediction (does not biodegradate fast)
BIOWIN3	Ultimate biodegradation Timeframe: Weeks - Months
BIOWIN4	Primary biodegradation Timeframe: Days-Weeks
BIOWIN5	MITI Linear Model Prediction (does not biodegradate fast)
BIOWIN6	MITI Non Linear Model Prediction (does not biodegradate fast)
BP	Boiling point (est = estimated, data exp =experimental data)
MP	Melting point (est = estimated, data exp =experimental data)
VP	Vapor pressure VP_Pa in Pa, others in mmHg

The following data compilation is available on the JRC-IHCP Computational Toxicology webpage as an sdf file (PRS_parents_chemical_characterisation.sdf).

Field	Description
ACID Pred $pK_{a,}$	Multiprotic ionisation acid constants calculated by ADMET Predictor
ACID Pred pK_B	Multiprotic ionisation base constants calculated by ADMET Predictor
ADME	Adsorption, distribution, metabolism, excretion; as pre-defined by Pipeline Pilot (Lipinski), calculated by Pipeline Pilot: Num_Atoms > 0 AND (N_Count + O_Count) <= 10 AND Molecular_Weight <= 500 AND Num_H_Donors <= 5 AND AlogP <= 5 Tag "No violation" means, ADME properties are within the range of the desriptors
ADMET	As pre-defined by ADMET Predictor
AlogP	Calculates LogP and Molar Refractivity (Ghose 1997), calculated by Pipeline Pilot
CAS Number	All PRS-CAS, related to the Parent compound
FANION	Cumulative Contribution of anionic Species to Fraction Ionized at specified pH (default 7.4) calculated by ADMET Predictor
FCATION	Cumulative Contribution of cationic Species to Fraction Ionized at specified pH (default 7.4) calculated by ADMET Predictor
FormalCharge	Formal Charge calculated by Pipeline Pilot
FUNION	Fraction Un-ionized species of all ionisable groups at specified pH (default 7.4) calculated by ADMET Predictor
FZwitter	Cumulative Contribution of Zwitterionic Species to Fraction Ionized at specified pH (default 7.4) calculated by ADMET Predictor
InChI_key_Parent	The InChI Key is a fixed length (25 characters) condensed digital representation of the InChI that is not human-readable. The InChI Key specification was released in September 2007 in order to facilitate web searches for chemical compounds, since these were problematical with the full-length InChI. Automatically generated by ACD Name
InChI_Parent	The IUPAC International Chemical Identifier is a textual identifier for chemical substances, designed to provide a standard and human-readable way to encode molecular information and to facilitate the search for such information in databases and on the web. Automatically generated by ACD Name
Is_Chiral	calculated by Pipeline Pilot
IUPAC Name	IUPAC names (English) automatically generated by ACD Name.
Molecular Formula	Molecular formula generated by Pipeline Pilot

Field	Description
Molecular Weight	Molecular Weight generated by Pipeline
N_Count	Number of nitrogen atoms, calculated by Pipeline Pilot
Num_Atoms	Number of atoms, calculated by Pipeline Pilot
Num_Bonds	Number of bonds calculated by Pipeline Pilot
Num_H_Acceptors	Number of hydrogen acceptors calculated by Pipeline Pilot
Num_H_Donors	Number of Hydrogen Donors, calculated by Pipeline Pilot
Num_RotatableBonds	Number of rotatable bonds calculated by Pipeline Pilot
Num_StereoAtoms	Atoms marked as Even-Atom-Stereo, Odd-Atom-Stereo, or Unknown-Atom-Stereo calculated by Pipeline Pilot.
Num_StereoBonds	Bonds marked as Cis-Bond-Stereo, as Trans-Bond-Stereo, or as Unknown-Bond-Stereo calculated by Pipeline Pilot.
Num_UnknownStereoAtoms	Atoms marked Unknown-Atom-Stereo calculated by Pipeline Pilot.
Number_of_tautomers	calculated by Pipeline Pilot (Sayle 1999)
Number_UnknownStereoBonds	Bonds marked Unknown-Bond-Stereo calculated by Pipeline Pilot.
O_Count	Number of oxygen atoms, calculated by Pipeline Pilot
S+logD	calculated from $pK_{a/b}$ and S+logP calculated by ADMET Predictor
S+logP	log P; Simulations Plus model calculated by ADMET Predictor
S+Sw	Water solubility calculated by ADMET Predictor
SMILES Parent	The simplified molecular input line entry specification or SMILES is a specification for describing the structure of chemical molecules using short ASCII strings. Automatically generated by Pipeline Pilot

8. References

ACDLabs 12.0, Copyright © 1998-2008 Advanced Chemistry Development Inc.

ADMET predictor 3.0 Copyright © 1998-2008 Simulationsplus, Inc

Bassan A & Worth AP (2008). The Integrated Use of Models for the Properties an Effects of Chemicals by means of a Structured Workflow. QSAR Comb. Sci. 27, No.1, 6-20

Sayle & Delaney (1999). Canonicalization and Enumeration of tautomers. EUROMUG99, 28-29 October 1999, Cambridge, UK, accessible from http://www.daylight.com/meetings/emug99/index.html

SciFinder, Copyright © 2008 American Chemical Society, accessible from https://scifinder.cas.org.

ChemSpider, accessible from http://www.chemspider.com/

Danish QSAR database, accessible from http://ecb.jrc.ec.europa.eu/qsar/

CONCAWE PETROTOX, accessible from http://www.concawe.be/

DSSTOX Distributed Structure Searchable Toxicity Database network, accessible from http://www.epa.gov/ncct/dsstox/index.html

ECHA (2008) Guidance Information Requirements and Chemical Safety Assessment, chapter R7b p.164, accessible from http://guidance.echa.europa.eu/guidance_en.htm

ECHA (2008b) Guidance Information Requirements and Chemical Safety Assessment, chapter R7b p.171, accessible from http://guidance.echa.europa.eu/guidance_en.htm

ECHA (2009) European Chemicals Agency, PRS list, accessible from http://apps.echa.europa.eu/preregistered/PRE_REGISTERED_SUBSTANCE.zip

EPISUITE, accessible from http://www.epa.gov/oppt/exposure/pubs/episuite.htm

European Commission, IHCP, ex-ECB, accessible from http://ecb.jrc.ec.europa.eu/classification-labelling/search-classlab/ (ATP 29)

European Commission, IHCP, ex-ECB, accessible from http://ecb.jrc.ec.europa.eu/esis/

Ghose AK, Viswanadhan VN & Wendoloski JJ (1998). Prediction of Hydrophobic (Lipophilic) Properties of Small Organic Molecules Using Fragment Methods: An Analysis of AlogP and CLogP Methods. J. Phys. Chem. A, 102, 3762-3772

Lipinski CA, Lombardo F, Dominy BW & Feeney PJ (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug Del. Rev., 23, 3-25

FOOTPRINT, Functional Tools for pesticides risk assessment and management, accessible from http://www.eu-footprint.org/

IUPAC International Union of Pure and Applied Chemistry, accessible from http://www.iupac.org/inchi/

Jager T, den Hollander HA, Janssen GB, van der Poel P, Rikken MGJ, Vermeire TG. 2000. Probabilistic risk management for new and existing chemicals: sample calculations, RIVM report No. 679102049. Bilthoven, The Netherlands

Jager T, van der Poel P &, Rikken MGJ (1997). Uncertainty analysis of EUSES: improving risk management by probabilistic risk assessment, RIVM report No. 679102039. Bilthoven, The Netherlands

Jager T (1998). Uncertainty analysis of EUSES: interviews with representatives from Member States and industry, RIVM report No. 679102047. Bilthoven, The Netherlands

JRC (2008). IHCP / Computational Toxicology; EINECS processed file accessible from http://ecb.jrc.ec.europa.eu/qsar/information-sources/

JRC-IHCP Computational Toxicology, accessible from http://ecb.jrc.ec.europa.eu/qsar/

Lipinski CA, Lombardo F, Dominy BW & Feeney PJ (1997). Adv. Drug Del. Rev. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings 23, 3-25

Kavlock RJ, Ankley G,.Blancato J, Breen M., Conolly R, Dix D, Houck K, Hubal E, Judson R, Rabinowitz J, Richard A, Woodrow Setzer R, Shah I, Villeneuve D & Weber E (2008). Computational Toxicology – A state of the Science Mini Review. Toxicological Sciences 103(1), 14-27

Mannhold R, Poda GI, Ostermann C & Tetko I (2008). Calculation of molecular lipophilicity: State-of-the-art and comparison of log P methods on more than 96,000 compounds IV. J. Pharm. Sci. 98(3):861-93

National Cancer Institute, accessible from http://cactus.nci.nih.gov/ncidb2/download.html

National Institute of Standards and Technology, accessible from http://webbook.nist.gov/chemistry/

Pavan M, Netzeva TI & Worth AP (2008). Review of literature based quantitative structure activity relationship models for bioconcentratrion. QSAR Comb. Sci. 27, No.1, 21-31

Pipeline Pilot 7.5, 2008 © Copyright Accelrys Software Inc.

PubChem, accessible from http://pubchem.ncbi.nlm.nih.gov/

European Commission (2006). Directive 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 Europeans Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1484/94 as well as Council Directive76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 200/21/EC

Sayle & Delaney (1999). Canonicalization and Enumeration of tautomers. EuroMUG99 28-29 October, Cambridge UK

Van Leeuwen C & Vermeire T (2007). Risk assessment of chemicals: An introduction. Springer p. 443. ISBN: 0-471-14747-8

European Commission

EUR 24138 EN – Joint Research Centre – Institute for Health and Consumer Protection
Title: Characterisation of the REACH Pre-Registered Substances List by Chemical Structure and Physicochemical Properties
Author(s): Klaus Daginnus
Luxembourg: Office for Official Publications of the European Communities
2009 – 39 pp. – 21.0 x 29.7 cm
EUR – Scientific and Technical Research series – ISSN 1018-5593
ISBN 978-92-14835-4
DOI 10.2788/54409

Abstract

In the European Union, the registrants of chemical substances under the REACH legislation are explicitly encouraged, and even required, to use non-testing methods as a means of identifying the presence or absence of hazardous properties of substances in order to meet the information requirements of REACH while at the same time minimising testing on vertebrate animals. The need to use non-testing methods or other alternative (nonanimal) methods such as in vitro tests, has led to the development and implementation of Integrated Testing Strategies based as far as possible on the integrated use of non-animal data. The use of non-testing methods within such strategies implies the need for computational tools and a structured workflow to facilitate their application.
The list of pre-registered substances (PRS) published by the European Chemicals Agency includes chemicals that industry may register in accordance with the deadlines specified in the REACH legislation. The PRS list does not include information on chemical structures, which are a prerequisite for the development and application of non-testing methods. Therefore, in order to facilitate the implementation of non-testing methods for the regulatory assessment of REACH chemicals, the Computational Toxicology Group within the Joint Research Centre has: a) generated structures for the PRS; b) processed the structures to generate substance identifiers such as IUPAC name, InChI codes and SMILES strings; c) assessed the availability of experimental toxicological data with DSSTOX and FOOTPRINT tags; and d) created a “QSAR-ready” data file to support the application of non-testing methods, such as QSARs.
The application of ACD NTS resulted in a high rate of yield for the generation of structures (85%) for monoconstituent
substances with a high reliability – in total, about 80,000 structures were generated. By comparing these results with inventories of structures available from other publicly available sources of information, additional high quality structures including precise information on stereochemistry were generated. A quality review resulted in the assignment of quality labels to the structures and in the further checking of about 5500 structures.
To support QSAR predictions and to estimate key physicochemical properties of the substances, these structures were processed to obtain an inventory of PRS parent substances, which serves as a standardised input for computational tools containing about 62,000 records. For these parent compounds the key features of the structures were calculated and key physicochemical properties were estimated using EPISUITE, Pipeline Pilot and ADMET Predictor. This chemical characterisation of the parent substances can be used to support a preliminary assessment of hazard and exposure.

How to obtain EU publications

Our priced publications are available from EU Bookshop (http://bookshop.europa.eu), where you can place an order with the sales agent of your choice.

The Publications Office has a worldwide network of sales agents. You can obtain their contact details by sending a fax to (352) 29 29-42758.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

JRC
EUROPEAN COMMISSION

LB-NA-24138-EN-N

ISBN 978-92-79-14835-4

Publications Office
Publications.europa.eu