http://echa.europa.eu

# Guidance for identification and naming of substances under REACH and CLP

Version: 1.3
February 2014

LEGAL NOTICE
This document contains guidance on REACH and CLP explaining the REACH and CLP obligations and how to fulfil them. However, users are reminded that the text of the REACH and CLP Regulations is the only authentic legal reference and that the information in this document does not constitute legal advice. The European Chemicals Agency does not accept any liability with regard to the contents of this document.

***Guidance for identification and naming of substances under REACH and CLP***

**Reference:** ECHA-11-G-10.2-EN
**Publ.date:** February 2014
**Language:** EN

© European Chemicals Agency, 2014
Cover page © European Chemicals Agency

Reproduction is authorised provided the source is fully acknowledged in the form “Source: European Chemicals Agency, http://echa.europa.eu/”, and provided written notification is given to the ECHA Communication Unit (publications@echa.europa.eu).

If you have questions or comments in relation to this document please send them (quote the reference and issue date) using the information request form. The information request form can be accessed via the Contact ECHA page at: http://echa.europa.eu/about/contact_en.asp

**EUROPEAN CHEMICALS AGENCY**

Mailing address: P.O. Box 400, FI-00121 Helsinki, Finland
Visiting address: Annankatu 18, Helsinki, Finland

# PREFACE

This document describes how to name and identify a substance under REACH and CLP. It is part of a series of guidance documents that are aimed to help all stakeholders with their preparation for fulfilling their obligations under the REACH and CLP Regulations. These documents cover detailed guidance for a range of essential REACH and CLP processes as well as for some specific scientific and/or technical methods that industry or authorities need to make use of under REACH and CLP.

The guidance documents were drafted and discussed within the REACH Implementation Projects (RIPs) led by the European Commission services, involving all stakeholders: Member States, industry and non-governmental organisations. These guidance documents can be obtained via the website of the European Chemicals Agency (http://echa.europa.eu/reach_en.asp). Further guidance documents will be published on this website when they are finalised or updated.

.

**Document History**

| Version | Comment | Date |
|---|---|---|
| Version 1 | First edition | June 2007 |
| Version 1.1 | Corrigendum<br><br>The corrigendum includes the following adjustments: editorial corrections, replacing of outdated information, correction of examples and introduction of updated references (as foot notes) with the aim of improving readability.<br><br>These adjustments do not change the requirements for Industry.<br><br>The main changes are identified in Appendix 3. | November 2011 (In English only) |
| Version 1.2 | Corrigendum<br><br>The definition of “phase-in substance” has been aligned with the definition in Regulation (EC) No 1907/2006 introduced by Council Regulation (EC) No 1354/2007 and by Corrigendum, OJ L 36, 5.2.2009, p. 84 (1907/2006).<br><br>Please note that changes in version 1.1 and 1.2 are consolidated into a single translated version 1.2 for languages other than English. | March 2012 |
| Version 1.3 | Corrigendum<br><br>Two missing structural formulae in chapter 7.6 were inserted. | February 2014 |

# TABLE OF CONTENTS


1 GENERAL ..... 1

1.1 OBJECTIVES ..... 1

1.2 SCOPE ..... 2

1.3 STRUCTURE OF THE GUIDANCE DOCUMENT ..... 2

2 DEFINITIONS AND ABBREVIATIONS ..... 4

2.1 ABBREVIATIONS ..... 4

2.2 DEFINITIONS ..... 6

3 FRAMEWORK FOR SUBSTANCE IDENTIFICATION IN REACH AND CLP ..... 9

3.1 SUBSTANCE DEFINITION ..... 9

3.2 EC INVENTORY ..... 9

3.2.1 The role of the EC inventory at entry into force of REACH ..... 10

3.2.2 The list numbers after entry into force of REACH ..... 11

3.3 REQUIREMENTS FOR SUBSTANCE IDENTIFICATION IN REACH AND CLP ..... 12

4 GUIDANCE FOR SUBSTANCE IDENTIFICATION AND NAMING IN REACH AND CLP ..... 14

4.1 INTRODUCTION ..... 14

4.2 SUBSTANCES OF WELL DEFINED COMPOSITION ..... 18

4.2.1 Mono-constituent substances ..... 19

4.2.2 Multi-constituent substances ..... 21

4.2.3 Substances of defined chemical composition and other main identifiers ..... 24

4.3 UVCB SUBSTANCES ..... 25

4.3.1 General guidance on UVCB Substances ..... 26

4.3.2 Specific types of UVCB substances ..... 34

5 CRITERIA FOR CHECKING IF SUBSTANCES ARE THE SAME ..... 42

6 SUBSTANCE IDENTITY WITHIN (LATE) PRE-REGISTRATION AND INQUIRY ..... 47

6.1 (LATE) PRE-REGISTRATION ..... 47

6.2 INQUIRY ..... 47

7 EXAMPLES ..... 49

7.1 DIETHYL PEROXYDICARBONATE ..... 49

7.2 ZOLIMIDINE ..... 50

7.3 MIXTURE OF ISOMERS ..... 50

7.4 FRAGRANCE AH ..... 53

7.5 MINERALS ..... 57

7.6 ESSENTIAL OIL OF LAVANDIN GROSSO ..... 59

7.7 CHRYSANTHEMUM OIL AND ISOMERS ISOLATED THEREOF ..... 63

7.8 PHENOL, ISOPROPYLATED, PHOSPHATE ..... 67

7.9 QUATERNARY AMMONIUM COMPOUNDS ..... 68

7.10 PETROLEUM SUBSTANCES ..... 71

7.10.1 Gasoline blending stream ($C_4$-$C_{12}$) ..... 71

7.10.2 Gas oils (petroleum) ..... 72

**7.11 ENZYMES** ..... **73**
7.11.1 Subtilisin ..... 73
7.11.2 a-Amylase ..... 75

**8 DESCRIPTION OF SUBSTANCES IN IUCLID 5** ..... **76**

**8.1 GENERAL PRINCIPLES** ..... **76**
8.1.1 Inventories ..... 77
8.1.2 Substance data set (IUCLID Sections 1.1, 1.2, 1.3 and 1.4) ..... 79

**8.2 EXAMPLES HOW TO FILL IN IUCLID 5** ..... **81**
8.2.1 Mono-constituent substance ..... 82
8.2.2 Multi-constituent substance ..... 83
8.2.3 Substance defined by its chemical composition plus other identifiers ..... 85
8.2.4 UVCB Substance ..... 86

**8.3 REPORTING OF ANALYTICAL INFORMATION** ..... **87**

**APPENDIX I - GUIDANCE INSTRUMENTS** ..... **89**

**APPENDIX II - TECHNICAL GUIDANCE PER SUBSTANCE IDENTIFICATION PARAMETER** ..... **93**

**APPENDIX III - UPDATE OF THE DOCUMENT** ..... **109**

## Tables

Table 2.1: Abbreviations ..... 4
Table 2.2: Definitions ..... 6
Table 3.1: Substance identification parameters in REACH *Annex VI* section 2 ..... 13
Table 4.1: Grouping of main identifiers for examples that represent various types of well defined similar substances ..... 15
Table 4.2 : Grouping of main identifiers for examples that represent various types of UVCB substances ..... 16

# 1 GENERAL

The REACH Regulation (Regulation (EC) No 1907/2006) sets up a system for the Registration, Evaluation, Authorisation and Restriction of Chemicals and established the European Chemicals Agency (ECHA) to implement the Regulation.[1]

The CLP Regulation (Regulation (EC) No 1272/2008) is the new European Regulation on Classification, Labelling and Packaging of chemical substances and mixtures.[2] The legislation introduces, throughout the EU, a new system for classifying and labelling chemicals, based on the United Nations' Globally Harmonised System (UN GHS).

The REACH Regulation focuses on substances. To ensure that the REACH processes are working properly, correct and unambiguous substance identification is essential. This guidance document on substance identification and naming is intended to support industry, Member States and the European Chemicals Agency.

This guidance document is based on experience with substance identification under the previous chemical legislation (Directive 67/548/EEC) and other currently existing EU chemical legislation e.g. the Biocidal Products Directive (98/8/EEC). However, current practices in relation to substance identity under the REACH Regulation and the Classification, Labelling and Packaging of substances and mixtures Regulation (CLP) form the basis for the refinement of this guidance. In addition, and where appropriate, approaches from other chemical schemes outside the European Union have also been taken into consideration.

Tailored guidance for different types of substances has been included.

This guidance document should be applied to identify and name substances which are regulated under the REACH and CLP Regulation.

## 1.1 OBJECTIVES

The objective of this guidance document is to give guidance for manufacturers and importers on recording and reporting the identity of a substance within the context of REACH and CLP. As an important key element of substance identification the guidance document provides guidance on how to name the substance. It also gives guidance on whether substances may be regarded as the same in the context of REACH and CLP. Identifying the same substances is important for the process of (late) pre-registration of phase-in substances, for inquiries, for data sharing, for Joint Submission of data, for notification to the Classification and Labelling inventory and for harmonisation of Classification and Labelling.

---

[1] 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 ("REACH").

[2] Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006 (Text with EEA relevance) ("CLP").

The identification of substances should preferably be conducted by industry experts. For those parties within industry with little expertise in substance identification, additional guidance on identification parameters is included as an appendix to this guidance document.

In addition, this guidance document lists some links to relevant tools to support the characterisation and checking of the chemical identity of a substance.

## 1.2 SCOPE

According to *Article 1* of REACH, the Regulation concerns the manufacture, import, placing on the market and use of substances on their own and in mixtures and articles. Mixtures and articles as such are not regulated by REACH.

In line with *Article 10* of REACH, a registration requires the substance identity to be recorded using the parameters specified in section 2 of *Annex VI* of REACH (see **Table 3.1**). Similar parameters (as specified in section 2.1 to 2.3.4 to *Annex VI* of REACH) are required for recording the substance identity for the purpose of notification in accordance with *Article 40(1)* of CLP. This guidance document is focused on appropriate identification of substances that fall under the legal definition of a substance in REACH and CLP and provides guidance on the substance identification parameters of Section 2 of *Annex VI* of REACH. The information given on substance identity shall be sufficient to identify each substance. One or more of the substance identification parameters can be omitted if it is not technically possible or if it does not appear scientifically necessary to give the requested information. The reasons for such omissions shall be clearly stated and based on a scientific justification.

The approach to identify a substance depends on the substance type. Therefore, the user of this guidance document is guided to specific chapters for different types of substances.

The EC Inventories used within the framework of Directive 67/548/EEC (EINECS, ELINCS and the NLP-list) are important tools in substance identification. Guidance on the role of these inventories under REACH is given in Chapter 3.2.

Substances within the scope of REACH and CLP (and consequently of this guidance document) are typically the result of chemical reactions as part of the manufacture of the substance and may contain multiple distinct constituent. Substances, as defined in REACH and CLP, also include substances chemically derived or isolated from naturally occurring materials, which may comprise a single element or molecule (e.g. pure metals or certain minerals) or several constituents (e.g. essential oils, metal mattes that are formed when sulfide metal ores are smelted). However, substances which are regulated by other Community legislation are in a number of cases exempted from registration under REACH (see *Article 2* of REACH). Also substances listed in *Annex IV* of REACH and substances fulfilling certain criteria which are specified in *Annex V* of REACH are exempted from registration. It should be noted that although a substance can be exempted from registration, this does not necessarily mean that the substance is exempted from other Titles of the REACH Regulation or from the requirements of the CLP Regulation.

## 1.3 STRUCTURE OF THE GUIDANCE DOCUMENT

Background information such as the objectives and scope of this guidance document are given in Chapter 1 and the abbreviations and definitions used can be found in Chapter 2. Relevant information on the framework for substance identification in REACH, e.g. substance definition and information requirements in the legal text, is given in Chapter 3.

The practical guidance for substance identification and naming is given in Chapter 4.

- Chapter 4.1 describes the differentiation between “well defined” and “poorly defined” substances; and within these two main groups different substance types can be recognised with their own specific guidance for substance identification. A key diagram is presented to guide the user to the appropriate chapter with identification guidance for the specific type of substance.

- In the subsequent chapters, specific guidance is given for each substance type, as a set of rules with explanation and examples.

Chapter 5 provides guidance for checking whether or not substances may be regarded as the same. Guidance on substance identity within the (late) pre-registration and inquiry processes is given in Chapter 6.

Furthermore, in Chapter 7, some detailed examples have been prepared using the practical guidance of Chapter 4 illustrating how industry could work with the guidance in this guidance document.

Finally, Chapter 8 gives guidance concerning the description of substances in IUCLID 5.

Appendix I lists some links to relevant tools to support the characterisation and checking of the chemical identity of a substance.

Appendix II provides more background information on the individual substance identification parameters used in the substance identification process, such as the nomenclature rules, EC numbers and CAS numbers, notations of molecular formula and structural formula and analytical methods.

Appendix III shows the main changes made in each new version of the guidance.

# 2 DEFINITIONS AND ABBREVIATIONS

## 2.1 ABBREVIATIONS

Key abbreviations used in this guidance document are listed and explained in **Table 2.1**.

**Table 2.1**: Abbreviations

| Abbreviation | Meaning |
|---|---|
| AAS | Atomic Absorption Spectroscopy |
| AISE | International Association for Soaps, Detergents and Maintenance Products |
| CAS | Chemical Abstracts Service |
| CLP | Regulation (EC) No 1272/2008 on classification, labelling and packaging of substance and mixtures |
| EC | European Commission |
| EINECS | European Inventory of Existing Commercial Chemical Substances |
| ELINCS | European List of Notified Chemical Substances |
| ENCS | Existing and New Chemical Substances (Japan) |
| ESIS | European Substances Information System |
| EU | European Union |
| GC | Gas chromatography |
| GHS | Globally Harmonised System |
| HPLC | High performance liquid chromatography |
| InChI | IUPAC International Chemical Identifier |
| INCI | International Nomenclature of Cosmetic Ingredients |
| IR | Infrared |
| ISO | International Organization for Standardization |
| IUCLID | International Uniform Chemical Information Database |
| IUBMB | International Union of Biochemistry and Molecular Biology |
| IUPAC | International Union of Pure and Applied Chemistry |
| MS | Mass spectroscopy |
| NLP | No Longer Polymer |
| NMR | Nuclear Magnetic Resonance |
| ppm | Parts per million |
| REACH | Registration, Evaluation, Authorisation and Restriction of Chemicals |
| SIEF | Substance Information Exchange Forum |
| SMILES | Simplified Molecular Input Line Entry Specification |
| TSCA | Toxic Substances Control Act (USA) |
| UVCB | Substances of Unknown or Variable composition, Complex reaction products or Biological materials |
| UV/VIS | Ultra violet /visible |
| w/w | Weight by weight |
| XRD | X-Ray Diffraction |
| XRF | X-Ray Fluorescence |

## 2.2 DEFINITIONS

Key definitions used in this guidance document are listed and described in **Table 2.2**.

These definitions take into account the definitions used in the REACH Regulation and the CLP Regulation. For this reason some terms are defined differently than they were used under Directive 67/548/EEC.

Table 2.2 : Definitions

| Definition | Description |
|---|---|
| Additive | A substance that has been intentionally added to stabilise the substance[3]. |
| Alloy* | A metallic material, homogenous on a macroscopic scale, consisting of two or more elements so combined that they cannot be readily separated by mechanical means.<br>Alloys are regareded as special mixtures. |
| Article* | An object which during production is given a special shape, surface or design which determines its function to a greater degree than does its chemical composition. |
| Chromatographic fingerprint | Representation of the composition of a substance from the characteristic distribution of constituents in an analytical chromatogram. |
| Component | Substance intentionally added to form a mixture. |
| Constituent | Any single species present in a substance that can be characterised by its unique chemical identity. |
| EC Inventory | Although not legally defined in the REACH Regulation, the EC Inventory is a combination of three independent and legally approved European lists of substances from the previous EU chemicals regulatory frameworks: EINECS, ELINCS and the NLP-list (no-longer polymers). The entries in the EC Inventory consist of a chemical name and a number (EC name and EC number), a CAS number, molecular formula (if available) and description (for certain types of substances). |
| EC Number | The EC Number is the numerical identifier for substances in the EC Inventory. |
| Impurity | An unintended constituent present in a substance as manufactured. It may originate from the starting materials or be the result of secondary or incomplete reactions during the manufacture process. While it is present in the final substance it was not intentionally added. |

[3] In other areas an additive can also have other functions, e.g. pH-regulator or colouring agent. However, in the REACH regulation and in this TGD an additive is a stabilising agent.

| Definition | Description |
|---|---|
| Intermediate* | A substance that is manufactured for and consumed in or used for chemical processing in order to be transformed into another substance (hereafter referred to *synthesis*):<br>(a) *non-isolated intermediate* means an intermediate that during synthesis is not intentionally removed (except for sampling) from the equipment in which the synthesis takes place. Such equipment includes the reaction vessel, its ancillary equipment, and any equipment through which the substance(s) pass(es) during a continuous flow or batch process as well as the pipework for transfer from one vessel to another for the purpose of the next reaction step, but it excludes tanks or other vessels in which the substance(s) are stored after the manufacture;<br>(b) *on-site isolated intermediate* means an intermediate not meeting the criteria of a non-isolated intermediate and where the manufacture of the intermediate and the synthesis of (an)other substance(s) from that intermediate take place on the same site, operated by one or more legal entities;<br>(c) *transported isolated intermediate* means an intermediate not meeting the criteria of a non-isolated intermediate and transported between or supplied to other sites; |
| IUCLID | International Uniform Chemical Information Database. IUCLID is a database and management system for the administration of data on chemical substances. |
| List Number | Automatically allocated number assigned by REACH-IT. Applies to all incoming valid submissions (e.g. pre-registrations, PPORD, inquiries, registrations, classification and labelling notifications). A list number has no legal relevance and is only used as a technical identifier for managing submissions within ECHA. |
| Main constituent | A constituent, not being an additive or impurity, in a substance that makes a significant part of that substance and is therefore used in substance naming and detailed substance identification. |
| Manufacturing* | Production and extraction of substances in the natural state. |
| Mixture* | Mixture or solution composed of two or more substances. |
| Monomer* | A substance which is capable of forming covalent bonds with a sequence of additional like or unlike molecules under the conditions of the relevant polymer-forming reaction used for the particular process. |
| Mono-constituent substance | As a general rule, a substance, defined by its composition, in which one main constituent is present to at least 80% (w/w). |
| Multi-constituent substance | As a general rule, a substance, defined by its composition, in which more than one main constituent is present in a concentration ³ 10% (w/w) and < 80% (w/w). |
| Non-phase-in substance | A substance requiring registration which does not benefit from the transitional regime provided for phase-in substances under REACH. |
| Not chemically modified substance* | A substance whose chemical structure remains unchanged, even if it has undergone a chemical process or treatment, or a physical mineralogical transformation, for instance to remove impurities. |

| Definition | Description |
|---|---|
| Notified substance* | A substance for which a notification has been submitted and which could be placed on the market in accordance with Directive 67/548/EEC. |
| Phase-in substance* | A substance which meets at least one of the following criteria:<br>(a) It is listed in the European Inventory of Existing Commercial Chemical Substances (EINECS);<br>(b) it was manufactured in the Community, or in the countries acceding to the European Union on 1 January 1995, on 1 May 2004 or on 1 January 2007, but not placed on the market by the manufacturer or importer, at least once in the 15 years before the entry into force of this Regulation, provided the manufacturer or importer has documentary evidence of this;<br>(c) it was placed on the market in the Community, or in the countries acceding to the European Union on 1 January 1995, on 1 May 2004 or on 1 January 2007, by the manufacturer or importer before the entry into force of this Regulation and it was considered as having been notified in accordance with the first indent of Article 8(1) of Directive 67/548/EEC in the version of Article 8(1) resulting from the amendment effected by Directive 79/831/EEC, but it does not meet the definition of a polymer as set out in this Regulation, provided the manufacturer or importer has documentary evidence of this, including proof that the substance was placed on the market by any manufacturer or importer between 18 September 1981 and 31 October 1993 inclusive; |
| Polymer* | A substance consisting of molecules characterised by the sequence of one or more types of monomer units. Such molecules must be distributed over a range of molecular weights wherein differences in the molecular weight are primarily attributable to differences in the number of monomer units. A polymer compromises the following:<br>(a) a simple weight majority of molecules containing at least three monomer units which are covalent bound to at least one other monomer unit or other reactant;<br>(b) less than a simple weight majority of molecules of the same molecular weight.<br>In the context of this definition a ‘monomer unit’ means the reacted form of a monomer substance in a polymer. |
| Substance* | A chemical element and its compounds in the natural state or obtained by any manufacturing process, including any additive necessary to preserve its stability and any impurity deriving from the process used, but excluding any solvent which may be separated without affecting the stability of the substance or changing its composition. |
| Substance which occurs in nature* | A naturally occurring substance as such, unprocessed or processed only by manual, mechanical gravitational means; by dissolution in water, by flotation, by extraction with water, by steam distillation or by heating solely to remove water, or which is extracted from air by any means. |

* Definitions according *Article 3* of the REACH.

# 3 FRAMEWORK FOR SUBSTANCE IDENTIFICATION IN REACH AND CLP

REACH and CLP include a definition of a substance and REACH lists the substance identification parameters (*Annex VI*, Section 2) that shall be included to identify the substance for the purpose of registration.

This chapter describes the substance definition in REACH and CLP (Chapter 3.1), provides generic guidance on how to use the EC Inventory from the previous chemicals regulatory framework (Chapter 3.2) and provides more background information on substance identification requirements that are specified in REACH (Chapter 3.3).

## 3.1 SUBSTANCE DEFINITION

A substance is defined in REACH (*Article 3(1)*) and in CLP (*Article 2(7)*):

> *Substance means a chemical element and its compounds in the natural state or obtained by any manufacturing process, including any additive necessary to preserve its stability and any impurity deriving from the process used, but excluding any solvent which may be separated without affecting the stability of the substance or changing its composition.*

The substance definition in REACH and CLP is identical to the definition of a substance that was used in the 7th amendment of the Dangerous Substances Directive (Directive 92/32/EEC amending Directive 67/548/EEC). In both cases, the definition goes beyond a pure chemical compound defined by a single molecular structure. The definition of the substance includes different constituents like impurities.

## 3.2 EC INVENTORY

There are three separate inventories established by the previous chemicals regulatory framework. These are the European Inventory of Existing Commercial Chemical Substances (EINECS), the European List of Notified Chemical Substances (ELINCS) and the No-Longer Polymers (NLP) list.

Substances on the European market between 1st January 1971 and 18th September 1981 are listed in the European Inventory of Existing Commercial Chemical Substances (EINECS).[4, 5, 6]

---

[4] EINECS is based on the **E**uropean **CO**re **IN**ventory (ECOIN) to which supplementary substance reporting could be made by industry (according criteria for reporting substances for EINECS). ECOIN was composed by blending different lists of chemicals presumed to be on the European market (e.g. TSCA). EINECS was published on 15th June 1990 and includes more than 100,000 substances. During the use of the inventory, a number of errors were identified (printing errors, e.g. incorrect chemical name, formula or CAS RN). Therefore, a corrigendum was published on 1st March 2002.

[5] ECB (2005) Manual of Decisions for implementation of the sixth and seventh amendments to Directive 67/548/EEC (Directives 79/831/EEC and 92/32/EEC) Non-confidential version. EUR 20519 EN. Updated version of June 2005.

[6] Geiss F, Del Bino G, Blech G, et al. (1992) The EINECS Inventory of existing chemical substances on the EC market. Tox Env Chem Vol. 37, p. 21-33.

This inventory comprises about 100 000 substances identified by a chemical name (and a description for certain types of substances), a CAS number and seven digit number called the EINECS number. EINECS numbers always start with 2 or 3 (2xx-xxx-x; 3xx-xxx-xx). Substances reported to EINECS have gone through a verification step which justifies entering the substance into the inventory.

Substances notified and placed on the market after 18th September 1981 are listed in the European List of New Chemical Substances (ELINCS).[5] This inventory (list) comprises all substances notified until 31 May 2008 in accordance with Directive 67/548/EEC and its amendments. These substances are so called "new substances", as they were not placed on the Community market by 18 September 1981. An ELINCS number was allocated to a substance by the European Commission after review by the Member States Competent Authorities (MSCAs). In contrast to EINECS, ELINCS does not include a CAS number in its entries but the notification number allocated by MSCA, the trade name (if available), the classification and the IUPAC name for classified substances. The ELINCS numbers are also seven digit numbers starting always with 4 (4xx-xxx-x).

Polymers were excluded from reporting to EINECS and were subject to special rules within Directive 67/548/EEC.[7, 8] The term "polymer" was further defined in the 7th amendment of Directive 67/548/EEC (Directive 92/32/EEC). As a consequence of the implementation of this definition, some substances which were considered to be polymers under the reporting rules for EINECS were no longer considered to be polymers under the 7th amendment. As all substances that are not listed in EINECS were notifiable, all "No-Longer Polymers" (NLPs) should, in theory, have been notified. However, the Council of Ministers made it clear that these no-longer polymers should not, retrospectively, become subject to notification. The Commission was requested to draw up a list of No-Longer Polymers (NLP-list). Substances to be included in this list were those which had been on the EU market between 18th September 1981 (the date of entry into force of Directive 79/831/EEC, the 6th Amendment of Directive 67/548/EEC), and 31st October 1993 (the date of entry into force of Directive 92/32/EEC, the 7th Amendment of Directive 67/548/EEC) and which satisfied the requirement that they were considered to be polymers under the reporting rules for EINECS but were no longer considered to be polymers under the 7th amendment. The NLP-list is a non-exhaustive list. The substances in the NLP list are identified with a chemical name, a CAS number and seven digit number called the NLP number. An NLP number always starts with 5 (5xx-xxx-x).

These three lists of substances, EINECS, ELINCS and the NLP-list, in combination are called the EC Inventory. Each substance in this inventory has an EC number allocated by the European Commission (see detailed information on the EC number in Appendix II).

Information on these substances can be obtained through the website of the European Commission Joint Research Centre (http://esis.jrc.ec.europa.eu/). In the future, an inventory of registered substances will be maintained and published by the European Chemicals Agency.

### 3.2.1 The role of the EC inventory at entry into force of REACH

The EC Inventory can be used as a tool for manufacturers and importers to decide whether a substance is a phase-in substance or a non-phase-in substance. Thus, the EC Inventory will

---

[7] ECB (2003) Notification of new chemical substances in accordance with Directive 67/548/EEC on the classification, packaging and labelling of dangerous substances. No Longer Polymer List. EUR 20853 EN.

[8] Rasmussen K, Christ G and Davis JB (1998) Registration of polymers in accordance with Directive 67/548/EEC. Tox Env Chem Vol. 67, pp. 251-261.

help manufacturers and importers to find out *when* the registration of a substance will be required, and if a (late) pre-registration or an inquiry is necessary.

The REACH Regulation sets out different procedures for registration of and data sharing of "existing" ("phase-in") substances (as defined *in Article 3(20)*) and "new" ("non-phase-in") substances.[9]

If a substance was previously notified in accordance with Directive 67/548/EEC and is, thus, listed in ELINCS, the notification submitted shall be regarded as a registration for the purpose of REACH (*Article 24*). These substances are considered to be already registered by the relevant manufacturer or importer who made the notification and require no initial registration from this manufacturer/importer. The manufacturer/importer nevertheless has the obligation to keep the registration up to date. Additional manufacturers/importers of a substance listed in ELINCS (not covered by the previous notification(s)) are liable for registration (as for non-phase-in substance) and sharing of data with the previous registrant shall be established. More guidance on this issue can be found in the Guidance on Registration available on the ECHA guidance website at http://guidance.echa.europa.eu/guidance_en.htm.

### 3.2.2 The list numbers after entry into force of REACH

When setting up the REACH-IT system, ECHA considered it beneficial to allocate automatically a number to substances in all incoming technically complete submissions (pre-registrations, PPORD, inquiries, registrations, classification and labelling notifications, etc.) for which an EC number was not specified (see the criteria of assignment of the list numbers below). This has technically facilitated managing, further processing and identification of the substances in these submissions. These so-called "list numbers" have the same numerical format as used for EINECS, ELINCS and NLP numbers but they start with different digits.

In contrast to EINECS, ELINCS and NLP entries, the list numbers are not based on a legal requirement nor have they been published in the Official Journal of the European Union. Therefore the list numbers do not have the same significance as EC numbers but have only the numerical format in common. They are solely of administrative, and not of regulatory relevance. Most importantly, the vast majority of the list numbers and the substance identification connected to them have never been checked for correctness, validity and whether the conventions outlined in this guidance document have been kept.

For this reason, it was initially not planned to release the list numbers to the public before they had been checked by ECHA. However, since during the pre-registration period approximately 40 000 substances were pre-registered without an EC number, ECHA decided to publish the list numbers with the list of pre-registered substances in order to facilitate the formation of SIEFs.

It has to be highlighted that it is possible that different list numbers are allocated to the same substance when different identifiers (e.g. name) are used for this substance. This is because the list number allocation is a fully automated process without human intervention. As a consequence, it is also possible that a list number will be allocated to an EINECS, ELINCS or NLP listed substance if in a submission to ECHA via REACH-IT a substance name is used which differs from that in the EC inventory.

The list numbers always start with 6, 7 or 9 (6xx-xxx-x; 7xx-xxx-x; 9xx-xxx-x).

[9] Definitions of "phase-in" and non-phase-in" substances are given in the Guidance on registration.

A substance identified in the dossier/submission with a CAS number, which is not linked to an EC number or other list number already assigned by ECHA, is assigned a list number starting with 6.

A substance for which solely a name is indicated in the dossier, which cannot be linked with a name in the EC inventory or with a list name, is assigned a list number starting with 9.

The list numbers starting with 7 are allocated during the inquiry process (*Article 26* of REACH) after verification of the substance identification. These entries have a reliable and verified substance identity.

It is important to note that for some EINECS entries, the description of a substance is relatively broad and could potentially be considered to cover more than one substance identity according to *Article 3(1)* of REACH. In these cases, the potential registrant is invited to describe the substance in question more precisely (e.g. via the IUPAC name or other identifiers). To demonstrate the phase-in status of the substance, the registrant should nevertheless indicate to which EINECS entry the substance belongs. In such cases, the European Chemicals Agency will decide whether or not it is appropriate to allocate a list number to the substance in question.

## 3.3 REQUIREMENTS FOR SUBSTANCE IDENTIFICATION IN REACH AND CLP

Under the REACH Regulation when a registration is required, it shall include information on the identification of the substance as specified in Section 2 of *Annex VI*. This information shall be adequate and sufficient to enable each substance to be identified. If it is not technically possible, or if it does not appear scientifically necessary, to give information on one or more of the substance identification parameters, the reasons shall be clearly stated as indicated in *Note 1* of *Annex VI*.

Similarly under the CLP Regulation, when a notification is required to be made (*Article 40* of CLP) it shall include information on the identification of the substance as specified in Section 2.1 to 2.3.4 of *Annex VI* of the REACH. This information shall be adequate to enable each substance to be identified. If it is not technically possible, or if it does not appear scientifically necessary, to give information on one or more of the substance identification parameters, the reasons shall be clearly stated as indicated in *Note 1* of *Annex VI*.

An overview of the substance identification parameters within REACH *Annex VI* is given in **Table 3.1**.

**Table 3.1**: Substance identification parameters in REACH *Annex VI* section 2

| | |
|---|---|
| **2.** | **IDENTIFICATION OF THE SUBSTANCE**<br>For each substance the information given shall be sufficient to enable each substance to be identified. If it is not technically possible or if it does not appear scientifically necessary to give information on one or more items below, the reason shall be clearly stated. |
| **2.1** | **Name or other identifier of each substance** |
| *2.1.1* | *Name(s) in the IUPAC nomenclature or other international chemical name(s)* |
| *2.1.2* | *Other names (usual name, trade name, abbreviation)* |
| *2.1.3* | *EINECS or ELINCS number (if available and appropriate)* |
| *2.1.4* | *CAS name and CAS number (if available)* |
| *2.1.5* | *Other identity code (if available)* |
| **2.2** | **Information related to molecular and structural formula of each substance** |
| *2.2.1* | *Molecular and structural formula (including SMILES notation, if available)* |
| *2.2.2* | *Information on optical activity and typical ratio of (stereo) isomer (if applicable and appropriate)* |
| *2.2.3* | *Molecular weight or molecular weight range* |
| **2.3.** | **Composition of each substance** |
| *2.3.1* | *Degree of purity (%)* |
| *2.3.2* | *Nature of impurities, including isomers and by-products* |
| *2.3.3* | *Percentage of (significant) main impurities* |
| *2.3.4* | *Nature and order of magnitude (.....ppm, .....%) of any additives (e.g. stabilising agents or inhibitors)* |
| *2.3.5* | *Spectral data (ultra-violet, infra-red, nuclear magnetic resonance or mass spectrum)* |
| *2.3.6* | *High performance liquid chromatogram, gas chromatogram* |
| *2.3.7* | *Description of the analytical methods or the appropriate bibliographical references for the identification of the substance and, where appropriate, for the identification of impurities and additives. This information shall be sufficient to allow the methods to be reproduced.* |

# 4 GUIDANCE FOR SUBSTANCE IDENTIFICATION AND NAMING IN REACH AND CLP

## 4.1 INTRODUCTION

Rules for identification and naming are different for various types of substances. For practical reasons, this guidance document is structured in such a way that, for each type of substance, the user is directly guided to the chapter where the appropriate guidance is given. To this end, some explanation about different substance types is given below and finally a key is given to find the appropriate chapter.

Substance identification should be based on at least the substance identification parameters listed in REACH *Annex VI*, section 2 (see **Table 3.1**). Therefore, any substance needs to be identified by a combination of the appropriate identification parameters:

- The IUPAC- and/or other name and other identifiers, e.g. CAS-number, EC-number (*Annex VI*, section 2.1);
- The molecular and structural information (*Annex VI*, section 2.2);
- The chemical composition (*Annex VI*, section 2.3);

A substance is completely identified by its chemical composition i.e. the chemical identity and the content of each constituent in the substance. Although such straightforward identification may be possible for most substances, for certain substances it is not feasible or not adequate within the scope of REACH and CLP. In those cases, other or additional substance identification information is required.

Thus, substances can be divided into two main groups:

1. “Well defined substances”: Substances with a defined qualitative and quantitative composition that can be sufficiently identified based on the identification parameters of REACH *Annex VI* section 2.

2. “UVCB substances”: Substances of Unknown or Variable composition, Complex reaction products or Biological materials. These substances cannot be sufficiently identified by the above parameters.

Variability of composition for well defined substances is specified by upper and lower limit of the concentration range(s) of the main constituent(s). For UVCB substances the variability is relatively large and/or poorly predictable.

It is recognised that there will be borderline cases between well defined substances (reaction products with many constituents, each within a broad range) and UVCB substances (reaction products with variable and poorly predictable composition). It is the responsibility of the registrant to identify a substance in the most appropriate way.

Rules for identification and naming differ for “well defined substances” with one main constituent and for “well defined substances” with more than one main constituent. And for the various substance types under the umbrella of “UVCB”, different identification and naming rules are described.

In **Tables 4.1** and **4.2**, the main identifiers are listed for several examples of various types of substances. These examples are grouped in such a way that similarities and differences for the substance identification are easily recognised.

**Tables 4.1** and **4.2** do not represent a comprehensive list of all possible substance types. This grouping of substances with identification and naming rules should not be considered as an official categorization system for substances, but as a practical help to apply the specific rules suitably and to find the appropriate guidance in this guidance document.

**Table 4.1**: Grouping of main identifiers for examples that represent various types of well defined similar substances

| Common features | Examples or representatives | Main identifiers |
|---|---|---|
| Well defined substances by chemical composition<br>*[Chapter 4.2.]* | Mono-constituent substances, e.g.<br>- benzene (95%)<br>- nickel (99%)<br>*[Chapter 4.2.1]* | Chemical composition: one main constituent ≥ 80%:<br>- Chemical identity of the main constituent (chemical name, CAS-number, EC-number, etcetera)<br>- Typical concentration and upper and lower limit |
| | Multi-constituent substances, e.g. defined reaction products such as<br>Reaction mass of 2-, 3-, and 4-chlorotoluene (30% each)<br>*[Chapter 4.2.2]* | Chemical composition: a mixture (reaction mass) of main constituents each between ≥10 - <80%:<br>- Chemical identity of each main constituent<br>- Typical concentrations and upper and lower limit for each constituent and for the reaction mass itself |
| | Substances defined by more than the chemical composition, e.g.<br>Graphite and diamond<br>*[Chapter 4.2.3]* | Chemical composition as mono- or multi-constituent substance<br>AND<br>Other physical or characterisation parameters: e.g. crystallomorphology, (geological) mineral composition, etc. |

**Table 4.2** : Grouping of main identifiers for examples that represent various types of UVCB substances

| **Common features** | | **Examples or representatives** | **Main identifiers** | | |
|---|---|---|---|---|---|
| | | | ***Source*** | ***Process*** | ***Other Identifiers*** |
| UVCB substances<br>(Substances of Unknown or Variable composition, Complex reaction products or Biological materials)<br>*[Chapter 4.3]* | Biological materials (B) | Extracts of biological materials e.g. natural fragrances, natural oils, natural dyes and pigments | · Plant or animal species and family<br>· Part of plant/animal | · Extraction<br>· Fractioning, concentrating, isolation, purification, etc.<br>· Derivation* | · Known or generic composition<br>· Chromatographic and other fingerprints<br>· Reference to standards<br>· Colour index |
| | | Complex biological macromolecules e.g. enzymes, proteins, DNA or RNA-fragments, hormones, antibiotics | | | · Standard enzyme index<br>· Genetic code<br>· Stereo configuration<br>· Physical properties<br>· Function/activity<br>· Structure<br>· Amino acid sequence |
| | | Fermentation products<br>antibiotics, biopolymers, enzyme mixtures, vinasses (products of sugar fermentation) etc. | · Culture medium<br>· Micro-organism applied | · Fermentation<br>· Isolation of products<br>· Purification steps | · Type of products: e.g. antibiotics, biopolymers, proteins etc<br>· Known composition |
| | Chemical and mineral substances with poorly defined, complex or variable composition (UVC) | Reaction mixtures with poorly predictable and/or variable composition | · Starting materials | · Chemical reaction type, e.g. esterification, alkylation, hydrogenation | · Known composition<br>· Chromatographic and other fingerprints<br>· Reference to standards |
| | | · Fractions or distillates, e.g. petroleum substances<br>· Clay e.g. bentonite<br>· Tars | · Crude oils<br>· Coal/peat<br>· Mineral gases<br>· Minerals | · Fractionation, distillation<br>· Conversion of fractions<br>· Physical processing<br>· Residues | · Cut off ranges<br>· Range of chain length<br>· Ratio aromatic/ aliphatic<br>· Known composition<br>· Standard index |
| | | Concentrates or melts, e.g. metallic minerals, or residues of various melting or metallurgic processes, e.g. slags | · Ores | · Smelting<br>· Heat treatment<br>· Various metallurgic processes | · Known or generic composition<br>· Concentration of metals |

* Underlined processes indicate synthesis of new molecule

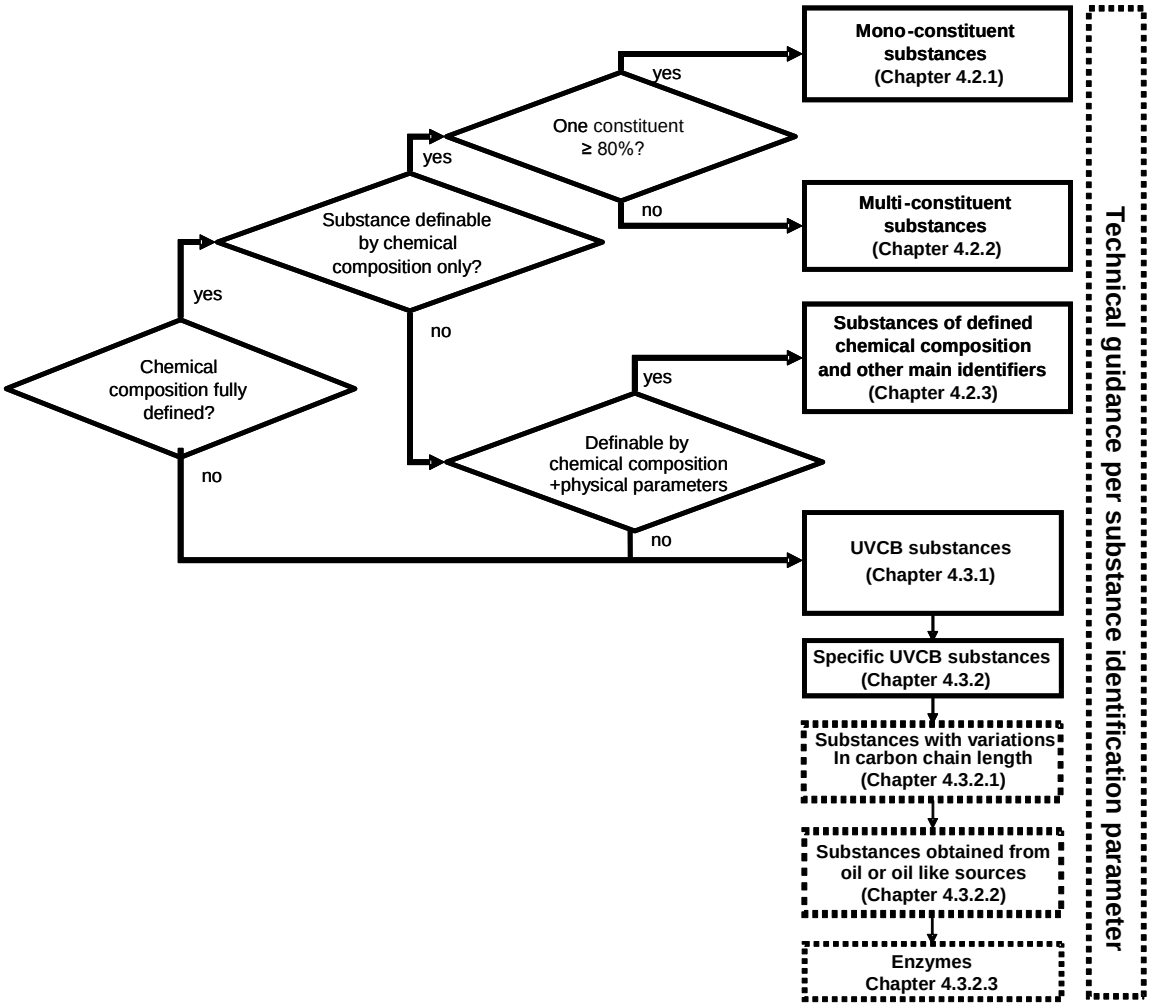
This chapter is divided into sub-chapters that contain specific guidance for the substance identification of various types of substances. A key to the appropriate chapters is given in **Figure 4.1**.

The key at **Figure 4.1** is based on criteria that are "rules of thumb". The registrant is responsible for selecting the most appropriate chapter and recording the substance identity in line with the rules and criteria for that type of substance.

The basic rule is that substances are defined as much as possible by the chemical composition and the identification of the constituents. Only if this is not technically feasible other identifiers should be used, as specified for the various types of UVCB substances.

If the registrant deviates from the substance identification rules and criteria of this guidance document, justification should be given. The substance identification should be transparent, accountable and ensure consistency.

**Figure 4.1**:Key to guidance document chapters and appendices for appropriate guidance for various types of substances




A description of the analytical methods and/or the appropriate bibliographic references for the identification of the substance and where appropriate, for the identification of the impurities and additives needs to be given (REACH Annex VI, sections 2.3.5, 2.3.6 and 2.3.7). This information should be sufficient to allow the methods to be reproduced. Typical results when applying the analytical techniques should also be provided.

## 4.2 SUBSTANCES OF WELL DEFINED COMPOSITION

Substances of well defined chemical composition are named according to the main constituent(s). For some types of substances, the chemical composition alone is not enough for characterisation. In these cases, some additional physical parameters in relation to the chemical structures have to be added to the substance identification.

As a general rule, the aim should be to cover the composition up to 100%, and each constituent requires a complete chemical specification, including structural information. For substances that are defined by their chemical composition, a distinction is made between:

- Main constituent: A constituent, not being an additive or impurity, in a substance that makes up a significant part of that substance and is therefore used in substance naming and detailed substance identification.

- Impurity: An unintended constituent present in a substance, as produced. It may originate from the starting materials or be the result of secondary or incomplete reactions during the production process. While impurities are present in the final substance, they were not intentionally added.

- Additive: A substance that has been intentionally added to stabilise the substance.

All constituents (except additives) which are not the main constituent(s) in the mono-constituent substance or a multi-constituent substance are considered to be impurities. Although in some sectors it is general practice to use the term “traces”, only the term “impurities” is used in this guidance document.

The different constituents have different identification requirements:

- Main constituents contribute to the naming of the substance and each main constituent shall be completely specified by all relevant identifiers;

- Impurities do not contribute to the naming of the substance and need only to be specified by name, CAS-number and EC-number and/or molecular formula.

- Additives contribute to the substance composition (but not to the naming) and should always be fully identified.

Some conventions are used to distinguish between mono-constituent and multi-constituent substances:

- A mono-constituent substance is a substance in which one constituent is present at a concentration of at least 80% (w/w) and which contains up to 20% (w/w) of impurities.

A mono-constituent substance is named according to the one main constituent;

- A multi-constituent substance is a substance consisting of several main constituents present at concentrations generally ≥ 10% and < 80% (w/w).

A multi-constituent substance is named as a reaction mass of two or more main constituents.

The above mentioned rules are intended as guidance. Deviation is acceptable if a robust justification can be given.

Normally, impurities present in a concentration ³ 1% should be specified. However, impurities that are relevant for the classification and/or for PBT assessment[10] shall always be specified, irrespective of the concentration. As a general rule, the compositional information should be completed up to 100%.

Additives in the sense of the REACH Regulation and CLP Regulation and in this guidance document are agents necessary to preserve the substance's stability. Thus, additives are an essential constituent of the substance and are taken into account when making the mass balance. However, outside the definition of REACH and this guidance document the wording 'additive' is also used for intentionally added substances with other functions, e.g. pH-regulators or colouring agents. These intentionally added substances are not part of the substance as such, and therefore not taken into account when making the mass balance.

Mixtures, as defined in REACH and CLP, are intentional mixtures of substances and are consequently not to be considered as multi-constituent substances.

Specific guidance on mono-constituent substances can be found in Chapter 4.2.1, and specific guidance on multi-constituent substances in Chapter 4.2.2. For substances that require additional information (e.g. certain minerals), guidance can be found in Chapter 4.2.3.

## 4.2.1 Mono-constituent substances

A mono-constituent substance is a substance, defined by its quantitative composition, in which one main constituent is present to at least 80% (w/w).

### 4.2.1.1 Naming convention

A mono-constituent substance is named after the main constituent. In principle, the name should be given in English language according to the IUPAC nomenclature rules (see Appendix I). Other internationally accepted designations can be given in addition.

### 4.2.1.2 Identifiers

A mono-constituent substance is identified by the chemical name and other identifiers (including the molecular and structural formula) of the main constituent and the chemical identity of the impurities and/or additives, and their typical concentration(s) and concentration range(s), which is proven by the spectroscopic and analytical information.

| Main constituent | Content (%) | Impurity | Content (%) | Substance Identity |
|---|---|---|---|---|
| m-xylene | 91 | o-xylene | 5 | m-xylene |
| o-xylene | 87 | m-xylene | 10 | o-xylene |

Normally, the main constituent is present ≥ 80% and should be specified completely by all above mentioned parameters. Impurities present in a concentration ≥ 1% should be specified by at least one of the following identifiers: chemical name (IUPAC and/or CAS name), CAS-number and EC-number and/or molecular formula. Impurities that are relevant for the classification and/or PBT assessment[11] shall always be specified by the same identifiers, independently from their concentration.

[10] More information on PBT assessment and relevant criteria can be found in the Guidance on information requirements and chemical safety assessment, chapter R11: PBT assessment.

[11] More information on PBT assessment and relevant criteria can be found in the Guidance on information requirements and chemical safety assessment, chapter R11: PBT assessment.

For correct application of the 80% rule, intentionally added substances like pH-regulators or colouring agents shall not be included in the mass balance.

The “80%-rule” has been applied for the notification of new substances (Directive 67/548/EEC). It can be seen as a “rule of thumb”. However, deviation from this 80% rule has to be justified. Possible examples for a justified deviation are:

- If the main constituent is < 80% but the substance can be shown to have similar physico-chemical properties and the same hazard profile as other mono-constituent substances with the same identity that fulfil the 80% rule.

- The range of concentrations for the main constituent and the impurities overlap the 80% criterion and the main constituent is only occasionally ≤ 80%.‘

**Examples**

| Subst. | Main constituent | Upper content (%) | Typical content (%) | Lower content (%) | Impurity | Upper content (%) | Typical content (%) | Lower content (%) | Substance identity |
|---|---|---|---|---|---|---|---|---|---|
| 1 | o-xylene | 90 | 85 | 65 | m-xylene | 35 | 15 | 10 | o-xylene |
| 2 | o-xylene<br>m-xylene | 90<br>35 | 85<br>15 | 65<br>10 | p-xylene | 5 | 4 | 1 | o-xylene |

Due to the concentration ranges of the main constituent and the impurity, substances 1 and 2 may be considered as a multi-constituent of the two main constituents, o-xylene and m-xylene, or as mono-constituent substances. The decision in such a case is to consider both as mono-constituent substance and this is triggered by the fact that o-xylene is typically present > 80%.

Guidance, how to describe mono-constituent substances in IUCLID 5, is given in Chapter 8.2.1. Additional information can also be found in the Data Submission Manual Part 18 – “How to report the substance identity in IUCLID 5 for registration under REACH”.

#### 4.2.1.3 Analytical Information

Sufficient spectral data is needed to confirm the structure of a mono-constituent substance. Several spectroscopic methods can be suitable, in particular Ultraviolet and Visible Absorption Spectroscopy (UV/Vis), Infrared Spectroscopy (IR) Nuclear Magnetic Resonance Spectroscopy (NMR) and Mass spectroscopy (MS). For inorganic substances, the use of X-Ray Diffraction (XRD) or X-Ray Fluorescence (XRF) or Atomic Absorption Spectroscopy (AAS) may be more suitable.

Chromatographic methods, such as Gas Chromatography (GC) or High-Performance Liquid Chromatography (HPLC) are needed to confirm the composition of the substance. If appropriate, also other valid constituent separation techniques may be used.

Spectroscopic and analytical methods are subject to continuous change. Therefore, it is the responsibility of the registrant to present appropriate spectral and analytical data.

### 4.2.2 Multi-constituent substances

A multi-constituent substance is a substance, defined by its quantitative composition, in which more than one main constituent is present in a concentration ≥ 10% (w/w) and < 80% (w/w). A multi-constituent substance is the result of a manufacturing process[12].

REACH requires the registration of a substance as produced. If a multi-constituent substance is manufactured, the multi-constituent substance needs to be registered[13] [14]. It is a case by case decision to establish to what extent the different steps in producing the substance are covered by the definition 'manufacturing'. All substances covered previously by EINECS (e.g. multi-constituent substances were covered if all individual constituents were listed on EINECS) would qualify as phase-in substances. There is no need to test the substance as such, if the hazard profile of the substance can be sufficiently described by the information of the individual constituents.

#### 4.2.2.1 Naming convention

A multi-constituent substance is named as a reaction mass of the main constituents of the substance as such i.e. not the starting materials needed to produce the substance .The generic format is: "Reaction mass of [names of the main constituents]". It is recommended that the names of the constituents are presented in alphabetical order and they are separated by the conjunction "and". Only main constituents typically ≥ 10% contribute to the name. In principle, the names should be given in English language according to the IUPAC nomenclature rules. Other internationally accepted designations can be given in addition.

#### 4.2.2.2 Identifiers

A multi-constituent substance is identified by the chemical name and identifiers of the substance as such, and the quantitative and qualitative chemical composition (chemical identity, including the molecular and structural formula) of the constituents, and is proven by analytical information.

| **Example** | | | | |
|---|---|---|---|---|
| **Main constituents** | **Content (%)** | **Impurity** | **Content (%)** | **Substance identity** |
| m-xylene<br>o-xylene | 50<br>45 | p-xylene | 5 | Reaction mass of m-xylene and o-xylene |

For multi-constituent substances, the chemical composition is known and more than one main constituent is relevant for the identification of the substance. Furthermore, the chemical composition of the substance is predictable, as typical values and ranges. The main constituents shall be specified completely by all relevant parameters. The sum of typical concentrations for main constituents (≥ 10%) and impurities (< 10%) shall be 100%.

For correct application of the 10% and 80% rule, intentionally added substances, e.g. pH-regulators or colouring agents, shall not be included in the mass balance.

[12] The difference between mixture and multi-constituent substance is that a mixture is obtained by blending of two or more substances without chemical reaction. A multi-constituent substance is the result of a chemical reaction.

[13] A number of substances are exempted for registration in REACH (e.g. the substances listed in Annex IV).

[14] This approach does not apply to a number of specific substances like minerals (see chapter 7.5 for more details).

Impurities present in a concentration ³ 1% should be specified by at least one of the following identifiers: chemical name, CAS-number and EC-number and/or molecular formula. Impurities that are relevant for the classification and/or PBT assessment shall always be specified by the same identifiers, independently from their concentration.

| **Example** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Main constituent** | **Upper content (%)** | **Typical content (%)** | **Lower content (%)** | **Impurity** | **Upper content (%)** | **Typical content (%)** | **Lower content (%)** | **Substance identity** |
| aniline<br>naphthalene | 90<br>35 | 75<br>20 | 65<br>10 | phenanthrene | 5 | 4 | 1 | Reaction mass of aniline and naphthalene |

According to the rules in this guidance document, this substance is a multi-constituent substance. Although the range of one constituent is > 80%, this happens only occasionally and the typical composition is < 80%.

Occasionally it is convenient to consider a substance as a multi-constituent substance even when one constituent is present at ≥ 80%. For example, a substance contains two constituents, one at 85% and another at 10%, the balance being impurities. Both constituents contribute towards and are essential for the desired technical effect of the substance. In this case, despite one constituent being present at > 80%, the substance can be described as a two-constituent substance.

Guidance, how to describe multi-constituent substances in IUCLID 5, is given in Chapter 8.2.2. Additional information can be also found in the Data Submission Manual Part 18 – “How to report the substance identity in IUCLID 5 for registration under REACH”.

#### 4.2.2.3 Analytical Information

In cases where spectral data provides information on the composition of the multi-constituent substance this information should be given. Several spectroscopic methods can be suitable, in particular Ultraviolet and Visible Absorption Spectroscopy (UV/Vis), Infrared Spectroscopy (IR) Nuclear Magnetic Resonance Spectroscopy (NMR) and Mass Spectroscopy (MS). For inorganic substances, the use of X-Ray Diffraction (XRD), X-Ray Fluorescence (XRF) or Atomic Absorption Spectroscopy (AAS) may be more suitable.

The use of chromatographic methods, such as Gas Chromatography (GC) and/or High-Performance Liquid Chromatography (HPLC) is needed to confirm the composition of the substance. If appropriate, also other valid constituent separation techniques may be used.

Spectroscopic and analytical methods are subject to continuous change. Therefore, it is the responsibility of the registrant to present appropriate spectral and analytical data.

#### 4.2.2.4 Registration of individual constituents of a multi-constituent substance

In general, recording the identity of substances for the purpose of (pre)registration should follow the multi-constituent substances approach (i.e. registration of the multi-constituent substance). As a deviation from that approach, individual constituents can be registered, if justifiable. The possibility to deviate from the standard case to identify (and potentially register) substances by their individual constituents is given, when

- there is no reduction in information requirements;

- there is sufficient existing data to justify the approach of registering the individual constituents i.e. the approach should normally not instigate additional (vertebrate animal) testing compared to the standard approach;

- registering the individual constituents leads to a more efficient situation (i.e. avoiding numerous registrations of substances which are composed of the same constituents);

- the information on the composition of the individual reaction masses is given.

The flexibility offered should not be abused to avoid data requirements. In the case of e.g. 1200 ton per year (tpa) of a multi-constituent substance "(C + D)", with a composition of 50% C and 50 % D, this approach would lead to two registrations with the following information:

Substance C

- Tonnage 600

- Data requirements to be fulfilled for >1000 ton (*Annex X*)

Substance D

- Tonnage 600

- Data requirements to be fulfilled for >1000 ton (*Annex X*)

This approach has to be combined with the REACH requirement to sum up volumes of the same substance per legal entity. The proposal is to establish the data requirements as follows:

- add up all volumes of the individual constituents (according to the quantities in the substance)

- refer to the highest volume of a substance that contains that constituent

The information requirements should be established based on the highest result. For the reporting of tonnages, the result of the summation of the tonnage for each individual constituent should be taken. Simplified examples are given hereafter to illustrate the practical implementation of this approach:

Example 1

Multi-constituent substance "C+D+E" is a result of a process within one legal entity, from which different substances are the result:

Substance 1: 50% C and 25 % D and 25 % E, 1100 tpa

Substance 2: 50% C and 50 % D 500 tpa

Also in this case the reaction product is the starting point: the two substances should be registered as multi-constituent substances. If the approach of registration of individual constituents is followed[15], the following would apply:

The reporting of substance D would in this case mean:

Tonnage: (25% * 1100) + (50% * 500) = 525 tpa

[15] The example is only intended to illustrate the establishment of the information requirements and the reporting of volumes. It does not address whether the approach is justifiable in this case.

Determination of the information requirements is based on the most severe requirement. In this case: >1000 tpa, as the total tonnage of the multi-constituent substance “C+D+E” is above 1000 tpa.

Note: in this example, the substances C and E should be registered accordingly.

Example 2

Multi-constituent substance “G+H+I” is a result of a process within one legal entity, from which different substances are the result:

Substance 3: 65% G and 15 % H and 20 % I, 90 tpa

Substance 4: 60% G and 40 % H, 90 tpa

Reporting of substance G:

Tonnage: (65% * 90) + (60% * 90) = 112.5 tpa

Determination of the information requirements is based on the most severe requirement. In this case: >100 tpa, as the total tonnage of the constituent G is above 100 tpa.

Note: in this example, the substances H and I should be registered accordingly

Besides the establishment of the information requirement mentioned, another consideration is the number of new studies (on vertebrate animals) that need to be executed. Before deciding on a strategy, potential registrants have to consider if there are sufficient existing studies (on vertebrate animals) and if the proposed flexibility will lead to less or more new testing (on vertebrate animals). The strategy that avoids new testing (on vertebrate animals) should be taken.

In case of doubt the standard route for recording the substance identity for the purpose of registration should always be the identification of the substance as it is manufactured.

### 4.2.3 Substances of defined chemical composition and other main identifiers

Some substances (e.g. inorganic minerals) which can be identified by their chemical composition need to be further specified by additional identifiers to get their own substance identification. These substances can be either mono-constituent substances or multi-constituent substances, but need, in addition to the substance identification parameters described in the previous chapters, other main identifiers to record the substance identity unequivocally.

| **Examples** |
|---|
| Some non-metallic minerals (from natural sources or man-made) with unique structures also need the morphology and mineral composition to identify the substance unequivocally. An example is kaolin (CAS 1332-58-7) composed of kaolinite, potassium aluminium silicate, feldspar and quartz. |

The current developments in nano-technology and insights in related hazard effects may cause the need for additional information on size of the substances in the future. The current state of development is not mature enough to include guidance on the identification of substances in the nanoform in this guidance document.

#### 4.2.3.1 Naming convention

In principal, the same naming convention, as for mono-constituent substances (see Chapter 4.2.1) or multi-constituent substances (see Chapter 4.2.2), needs to be followed.

For inorganic minerals the mineralogical names can be used for the constituents. For example, apatite is a multi-constituent substance comprised of a group of phosphate minerals, usually referred to as hydroxylapatite, fluorapatite, and chlorapatite, named for high concentrations of $OH^-$, $F^-$, or $Cl^-$ ions, respectively, in the crystal. The formula of the mixture of the three most common species is $Ca_5(PO_4)_3(OH, F, Cl)$. Another example is aragonite, one of the special crystalline structures of calcium carbonate.

#### 4.2.3.2 Identifiers

These substances are identified and named according to the rules for mono-constituent substances (see Chapter 4.2.1) or multi-constituent substances (see Chapter 4.2.2). The other specific main identification parameters to be added depend on the substance. Examples of other main identifiers can be elemental composition with spectral data, the crystalline structure as revealed by X-ray diffraction (XRD), Infra Red absorption peaks, swelling index, cation exchange capacity or other physical and chemical properties.

For minerals, it is important to combine the results of the elemental composition with the spectral data to identify the mineralogical composition and crystalline structure. This is then confirmed by characteristic physical-chemical properties like crystalline structure (as revealed by X-ray diffraction), shape, hardness, swelling capacity, density and/or surface area.

Examples of specific additional main identifiers can be given for specific minerals, as minerals have characteristic physical-chemical properties which enable the completion of their identification, e.g.: very low hardness for talc, swelling capacity of bentonite, shapes of diatomite, very high density of barite and surface area (nitrogen adsorption).

Guidance, how to describe substances of defined chemical composition and other main identifiers in IUCLID 5, is given in Chapter 8.2.3.

#### 4.2.3.3 Analytical Information

The same analytical information as for mono-constituent substances (see Chapter 4.2.1) or multi-constituent substances (see Chapter 4.2.2) should be given. For those substances for which spectral data, GC or HPLC chromatograms are not sufficient for identification, information resulting from other analytical techniques shall be given, e.g. X-ray diffraction for minerals, elementary analysis etc. The criterion is that sufficient information should be provided to confirm the structure of the substance.

## 4.3 UVCB SUBSTANCES

Substances of **U**nknown or **V**ariable composition, **C**omplex reaction products or **B**iological materials,[16, 17, 18] also called UVCB substances, cannot be sufficiently identified by their chemical composition, because:

---

[16] Rasmussen K, Pettauer D, Vollmer G et al. (1999) Compilation of EINECS: Descriptions and definitions used for UVCB substances. Tox Env Chem Vol. 69, pp. 403-416.

[17] US EPA (2005-B) Toxic Substances Control Act Inventory Registration for Combinations of two or more substances: complex reaction products.

[18] US EPA (2005-D) Toxic Substances Control Act Inventory Registration for Chemical Substances of Unknown or Variable Composition, Complex Reaction Products and Biological Materials: UVCB Substances.

- The number of constituents is relatively large and/or
- The composition is, to a significant part, unknown and/or
- The variability of composition is relatively large or poorly predictable.

As a consequence, UVCB substances require other types of information for their identification, in addition to what is known about their chemical composition.

It can be seen from **Table 4.2** that the main identifiers for the various types of UVCB substances are related to the source of the substance and the process used; or they belong to a group of "other main identifiers" (e.g. "chromatographic or other fingerprints"). The number and kind of identifiers given in **Table 4.2** represent an illustration of variability of types and shall not be regarded as a comprehensive overview. Where the chemical composition of e.g. a complex reaction product or a substance of biological origin is known, substance identification should be identified either as a mono- or multi-constituent substance, as appropriate. The consequence of defining a substance as UVCB is that any significant change of source or process would be likely to lead to a different substance that should be registered again. If a reaction mixture is identified as a "multi-constituent substance", the substance may be derived from a different source and/or by different processes as long as the composition of the final substance stays within the specified range. Hence, a new registration would not be required.

Generic guidance on UVCB substances can be found in Chapter 4.3.1 and specific guidance on substances with variation in the carbon-chain lengths, substances obtained from oil or oil like sources and enzymes, as specific types of UVCB substances, in Chapter 4.3.2.

## 4.3.1 General guidance on UVCB Substances

This chapter of the guidance document provides generic guidance on how to use certain main identifiers, besides the substance identification parameters of REACH *Annex VI* (section 2), to identify UVCB substances.

### 4.3.1.1 Information on chemical composition

UVCB substances either cannot be uniquely specified with the IUPAC name of the constituents, as not all the constituents can be identified; or they may be generically specified but with a lack of specificity due to variability of the exact composition. Due to the lack of differentiation between constituents and impurities, the terms "main constituents" and "impurities" should not be regarded as relevant for UVCB substances.

However, the chemical composition and the identity of the constituents should still be given as far as known. The description of the composition can often be given in a more generic way, for example "linear fatty acids C8-C16" or "alcohol ethoxylates with alcohols C10-C14 and 4-10 ethoxylate units". Additionally, information on chemical composition can be given on the basis of well-known reference samples or standards; and in many cases indexes and existing codes can be used in addition. Other generic information on the composition can consist of so called "fingerprints", that is, e.g. chromatographic or spectral images that show a characteristic peak distribution pattern.

For a UVCB substance, all known constituents and all constituents present at concentrations ≥ 10% should be specified by at least an English-language IUPAC name and preferably a CAS number; the typical concentrations and concentrations ranges of the known constituents should be given as well. Constituents that are relevant for the classification and/or PBT assessment[19]

[19] More information on PBT assessment and relevant criteria can be found in the Guidance on information

of the substance shall always be identified by the same identifiers, independently from their concentration.

Unknown constituents should be identified as far as possible by a generic description of their chemical nature. Additives should be completely specified in a similar way to that described for well defined substances.

#### 4.3.1.2 Main identification parameters – name, source and process

As the chemical composition alone is not sufficient for substance identification, the substance shall in general be identified by its name, its origin or source and the most relevant steps taken during processing. Other substance properties can also be important identifiers, either as relevant generic identifiers (e.g. boiling point) or as crucial identifiers for specific groups of substances (e.g. catalytic activity for enzymes).

### 1. Name convention

In general, the name of a UVCB substance is a combination of source and process with the general format: first the source and then the process(es).

- A substance derived from biological sources is identified by the name of the species.
- A substance derived from non-biological sources is identified by the starting materials.
- Processes are identified by the type of chemical reaction if synthesis of new molecules is involved, or as a type of refinement step e.g. extraction, fractioning, concentration, or as residue.

| **Examples** | |
|---|---|
| **EC number** | **EC Name** |
| 296-358-2 | Lavender, Lavandula hybrida, ext., acetylated |
| 307-507-9 | Lavender, Lavandula latifolia, ext., sulfurized, palladium salt |

In case of reaction products different formats have been used in the EC Inventory, e.g.

- EINECS: Main starting material, reaction product(s) of other starting material(s)
- ELINCS: Reaction product(s) of starting material(s)

| **Examples** | |
|---|---|
| **EC number** | **EC Name** |
| 232-341-8 | Nitrous acid, reaction products with 4-methyl-1,3-benzenediamine hydrochloride |
| 263-151-3 | Fatty acids, coco, reaction products with diethylenetriamine |
| 400-160-5 | Reaction products of tall-oil fatty acids, diethanolamine and boric acid |
| 428-190-4 | Reaction product of: 2,4-diamino-6-[2-(2-methyl-1H-imidazol-1-yl)ethyl]-1,3,5-triazine and cyanuric acid |

In this guidance document, the generic format of the name of reaction products is “Reaction product of [names of the starting materials]”. In principle, the names should be given in English language according to the IUPAC nomenclature rules. Other internationally accepted designations can be given in addition. It is recommended to substitute the word “reaction” in the

requirements and chemical safety assessment, chapter R11: PBT assessment.

name with the specific type of reaction described in a generic way e.g. esterification or salt formation etc. (see guidance in the four specific UVCB sub-classes, below).

## 2. Source

The source can be divided into two groups:

### 2.1. Sources of biological nature

Substances of biological origin have to be defined by the genus, species and the family e.g. *Pinus cembra, Pinaceae* means *Pinus* (genus), *cembra* (species), *Pinaceae* (family), and strain or genetic type, if relevant. If appropriate, the tissue or the part of the organism used for extraction of the substance, e.g. bone marrow, pancreas; or stem, seeds or roots, should be given as well.

| **Examples** | |
|---|---|
| **EC number** | **EC name** |
| 283-294-5 | Saccharomyces cerevisiae, ext. |
| | **EC description** |
| | Extractives and their physically modified derivatives such as tinctures, concretes, absolutes, essential oils, oleoresins, terpenes, terpene-free fractions, distillates, residues, etc., obtained from Saccharomyces cerevisiae, Saccharomycelaceae. |
| 296-350-9 | Arnica mexicana, ext. |
| | **EC description** |
| | Extractives and their physically modified derivatives such as tinctures, concretes, absolutes, essential oils, oleoresins, terpenes, terpene-free fractions, distillates, residues, etc., obtained from Arnica mexicana, Compositae. |

### 2.2. Chemical or mineral sources

In the case of reaction products of chemical reactions, the starting materials have to be described with their IUPAC name in English language. Mineral sources have to be described in generic terms e.g. phosphate ores, bauxite, china clay, mineral gas, coal, peat.

## 3. Process

Processes are identified by the type of chemical reaction if synthesis of new molecules is involved; or as a type of refinement steps, e.g. extraction, fractioning, concentration; or as a residue of a refinement.

For some substances, e.g. chemical derivates, the process shall be described as a combination of refinement and synthesis.

- Synthesis

A certain chemical or biochemical reaction occurs between the starting materials resulting in the substance. For example, the Grignard reaction, sulfonation, enzymatic splitting by protease or lipase etc. Many derivation reactions belong also to this type.

For newly synthesised substances, for which the chemical composition cannot be given, the starting materials are the main identifier together with a specification of the reaction, i.e. the type of chemical reaction. The type of chemical reaction is indicative for the molecules expected to be present in the substance. There are several types of final chemical reaction: hydrolysis, esterification, alkylation, chlorination etc. As this gives only generic information about the possible substances produced, in many cases a chromatographic fingerprint will also be necessary for full substance characterisation and identification.

| **Examples** | |
|---|---|
| **EC number** | **EC Name** |
| 294-801-4 | Linseed oil, epoxidised, reaction products with tetraethylenepentamine |
| 401-530-9 | Reaction product of (2-hydroxy-4-(3-propenoxy)benzophenone and triethoxysilane) with (hydrolysis product of silica and methyltrimethoxysilane) |

- Refinement

Refinement can be applied in many ways to substances of natural or mineral origin, where the chemical identity of the constituents is not changed, but the concentration of the constituents are changed, e.g. cold processing of plant tissue followed by extraction with an alcohol.

Refinement can be further defined in processes like extraction. The substance identification depends on the type of process:

- For substances derived by physical methods, e.g. refinement or fractionating, the cut-off range and parameter of the fraction shall be specified (e.g.: molecular size, chain length, boiling point, volatility range etc.);

- For substances derived by concentrating, e.g. products from metallurgical processes, centrifuged precipitates, filter residues etc., the concentration step shall be specified together with the generic composition of the resulting substance in comparison to the starting material;

| **Examples** | |
|---|---|
| **EC number** | **EC Name** |
| 408-250-6 | Organotungsten compound concentrate (reaction products of tungsten hexachloride with 2-methylpropan-2-ol, nonylphenol and pentane-2,4-dione) |

- For residues of a specific reaction, e.g. slags, tars and heavy ends, the process is to be described together with the generic composition of the resulting substance;

| **Examples** | |
|---|---|
| **EC number** | **EC Name** |
| 283-659-9 | Tin, melting residues |
| | **EC description** |
| | Substance resulting from the use and production of tin and its alloys obtained from primary and secondary sources and including recycled plant intermediates. Composed primarily of tin compounds and may contain other residual nonferrous metals and their compounds. |
| 293-693-6 | Soybean meal, protein extn. Residue |
| | **EC description** |
| | By-product, containing primarily carbohydrates, produced by an ethanolic extraction of defatted soybean. |

- For extracts, the extraction method, the solvent used for the extraction and other relevant conditions, e.g. temperature/temperature range) shall be given.

- For combined processing, each process step shall be specified (in a generic way) in addition to the source information. This combined processing is of particular relevance in the case of chemical derivations.

Examples:

- o A plant is first extracted, the extract is distilled and the distilled fraction of the plant extract is used for chemical derivation. The resulting substance may be further purified. The purified product might eventually be well defined by its chemical composition and there is no need to identify the substance as a UVCB. If the product is still to be considered as UVCB, the combined processing can be described as a "purified chemical derivate of a distilled fraction of a plant extract."

  If the further processing of an extract includes only physical derivation, the composition will change but without intended synthesis of new molecules. Nevertheless, the change of composition results in a different substance, e.g. a distillate or precipitate of a plant extract.

- o For the production of petroleum products, chemical derivation and fractioning are often used in combination. For example, oil distillation followed by cracking generates a fraction of the starting material and also new molecules. Thus, in that case, both types of processes should be identified or the distillate should be specified as the starting material of the cracking. In particular, this applies to petroleum derivates that often result from a combination of processes. However, a separate specific system can be used for identification of petroleum substances (see Chapter 4.3.2.2).

As a chemical derivate of an extract will not contain the same constituents as the parent extract, it shall be regarded as a different substance. This rule may have as a consequence that the identification by name and description deviates from the earlier EINECS name and description. At the time of the setting up of the EINECS inventory, extracts from different processes, different solvents and even physical or chemical derivates were often covered under one single entry. However, under REACH these substances should be registered as separate substances.

## 4. Other substance identification parameters

Besides the chemical name, the source and the specification of the process, a UVCB substance should include any other relevant information, as required by REACH *Annex VI*, section 2.

Especially for specific types of UVCB substances other identification parameters can be relevant. Additional other identifiers may include:

- Generic description of chemical composition;

- Chromatographic fingerprint or other types of fingerprint;
- Reference material (e.g. ISO);
- Physical-chemical parameters (e.g. boiling point);
- Colour Index number;
- AISE number.

Specific guidance on the rules and criteria, how to use the name, source and process information for the identification of UVCB substances, is included below for various types of sources and processes. In the following paragraphs four sub-types of UVCB substances are described as a combination of biological or chemical/mineral sources and processes (synthesis or refinement).

Guidance, how to describe UVCB substances in IUCLID 5, is given in Chapter 8.2.4.

**UVCB sub-type 1, where the source is biological and the process is a synthesis**

Substances of biological nature can be modified in (bio)chemical processing to generate constituents that were not present in the starting material, e.g. chemical derivates of plant extracts or products of enzymatic treatment of the extracts. For example, proteins can be hydrolysed by protease to generate oligopeptides, or cellulose from wood can be carboxylated to yield Carboxy Methyl Cellulose (CMC).

Products of fermentation may also belong to this UVCB sub-type. For example, vinasse is a product of sugar fermentation that, compared to the sugar, contains many different constituents. When fermentation products are further purified, the substances may eventually become fully identifiable by their chemical composition and should no longer be identified as a UVCB substance.

Enzymes are a special group of substances that can be derived by extraction and further refinement from a source of biological origin. Although the source and the process could be specified in detail, this does not generate the specific information on the enzyme. For these substances, a specific system for classification, naming and identification shall be used (see Chapter 4.3.2.3).

For substance identification, the final process step shall be given and/or any other process step that is relevant for the identity of the substance.

A description of the chemical process shall be a generic description of the type of process (esterification, alkaline hydrolysis, alkylation, chlorination, substitution etc.), together with relevant process circumstances.

A description of the biochemical process can be a generic description of the catalysed reaction, together with the name of the enzyme catalysing the reaction.

For substances produced by fermentation or (tissue) cultures of species, the fermenting species, type and general conditions of fermentation (batch or continuous, aerobic, anaerobic, anoxic, temperature, pH, etc) should be given, together with any further process steps applied to isolate the fermentation products, e.g. centrifugation, precipitation, extraction, etc. If these substances are further refined, this may yield a fraction, a concentrate or a residue. These further processed substances are identified with additional specification of the further process steps.

**UVCB sub-type 2, where the source is chemical or mineral and the process is a synthesis**

UVCB substances obtained from chemical or mineral sources, derived via a process in which new molecules are synthesized, are "reaction products". Examples of chemical reaction products are esterification, alkylation or chlorination products. Biochemical reactions by application of isolated enzymes are special types of chemical reactions. However, if a complex biochemical pathway of synthesis is applied using complete micro-organisms, it is better to consider the resulting substance as a fermentation product and identify it by the fermentation process and fermenting species rather than by the starting materials (see UVCB sub-type 4).

Not every reaction product should automatically be specified as a UVCB. If a reaction product can be sufficiently defined by the chemical composition (including some variability), identification as a multi-constituent substance (see Chapter 4.2.2) should be preferred. Only when the composition of the reaction product is insufficiently known or poorly predictable the substance should be identified as a UVCB substance ("reaction product"). The identification of a reaction product is based on the starting materials for the reaction and on the (bio)chemical reaction process in which the substance is generated.

| **Examples** | | |
|---|---|---|
| **EC number** | **EINECS Name** | **CAS-number** |
| 294-006-2 | Nonanedioic acid, reaction products with 2-amino-2-methyl-1-propanol | 91672-02-5 |
| 294-148-5 | Formaldehyde, reaction products with diethylene glycol and phenol | 91673-32-4 |

A main identifier for reaction products is the description of the manufacturing process. For substance identification, the final or most relevant process step shall be given. The chemical process description shall be a generic description of the type of process (e.g. esterification, alkaline hydrolysis, alkylation, chlorination, substitution etc.), together with relevant process circumstances. A biochemical process shall be described by the type of reaction, together with the name of the enzyme catalysing the reaction.

**UVCB sub-type 3, where the source is biological and the process is refinement**

UVCB substances of biological origin, resulting from a refinement process in which no new molecules are intentionally generated can be e.g. extracts, fractions of an extract, concentrates of an extract, purified extract or process residues of substances of biological origin.

As soon as an extract is further processed, the substance is no longer identical with the extract but is another substance that belongs to another UVCB sub-type, e.g. a fraction or a residue of an extract. These substances shall be specified with additional (further) processing parameters. If the extract is modified in chemical or biochemical reactions, generating new molecules (derivates), the identification of the substance is covered using the guidance of UVCB sub-type 2 or Chapter 4.2 for a well defined substance.

This differentiation of further processed extracts may have the consequence that the new name and description will differ from those in the EINECS inventory. At the time of setting up the inventory, such a differentiation has not been made and all types of extracts with different solvents and further process steps might have been covered under a single entry.

The first main identifier for this sub-type of UVCB substances is the family, genus and species of the organism from which the substance originates. If appropriate, the tissue or the part of the organism used for extraction of the substance should be given, e.g. bone marrow, pancreas; or stem, seeds or roots. For substances of microbiological origin, the strain and genetic type of the species shall be defined.

If the UVCB substance is derived from a different species, it will be regarded as a different substance, even if the chemical composition might be similar.

| **Examples** | |
|---|---|
| **EC number** | **EINECS name** |
| 290-977-1 | Oxidised logwood (Haematoxylon campechianum) extract |
| | **EC description** |
| | This substance is identified in the colour index by colour index constitution No C.I. 75290 oxidised. |
| 282-014-9 | Pancreatic extracts, deproteinated |

The second main identifier is the processing of the substance, e.g. the extraction process, the fractioning, purification or concentration process or the process that influences the composition of the residue. Thus, refinements of extracts made by different processes, e.g. using different solvents or different purification steps, will result in different substances.

The more steps are applied for refinement, the more feasible it will become to define the substance by its chemical composition. In that case, different source species or different process modifications do not lead automatically to a different substance.

A main identification parameter for substances of biological origin is the description of the relevant processes. For extracts, the extraction process shall be described to the level of detail relevant for the identity of the substance. At least the solvent used shall be specified.

When further process steps are used for manufacturing the substance, such as fractioning or concentration, the combination of relevant process steps shall be described, e.g. the combination of extraction and fractioning including the cut-off ranges.

**UVCB sub-type 4, where the source is chemical or mineral and the process is a refinement**

Substances of non-biological origin, i.e. that are or originate from minerals, ores, coal, natural gas and crude oil, or other raw materials for the chemical industry, and resulting from processing without intentional chemical reactions can be (purified) fractions, concentrates or residues of these processes.

Coal and crude oil are used in distillation or gasification processes to produce a wide variety of substances, e.g. petroleum substances and fuel gases etc., and also residues such as tars and slags. Very often, a distilled or otherwise fractionated product is immediately further processed, including chemical reactions. In such cases, substance identification shall follow the guidance given for UVCB sub-type 2, as the process is more relevant than the source.

For petroleum substances a special identification system is used (see Chapter 4.3.2.2). Substances covered by that system include fractions and chemical reaction products.

Other substances in UVCB sub-type 4 may include ores, ore concentrates and slags containing varying amounts of metals that may be extracted by metallurgical processing.

Minerals such as bentonite or calcium carbonate can be processed by e.g. acid dissolution and/or chemical precipitation or in ion-exchange columns. When the chemical composition is fully defined, minerals should be identified according to the guidance in the appropriate part of Chapter 4.2. If minerals are processed only by mechanical methods, e.g. by grinding, sieving, centrifugation, flotation etc., they are still considered to be the same as the minerals as mined. Minerals that are produced through a manufacturing process can – for the purpose of

identification[20] - be regarded to be the same as their naturally occurring equivalent provided the composition is similar and the toxicity profile identical.

A main identification parameter for substances of non-biological origin is the description of the relevant process step(s).

For fractions, the fractioning process shall be described with the parameters and cut-off range for the isolated fraction, together with a description of previous process steps when relevant.

For the concentration step, the type of process, e.g. evaporation, precipitation etc. shall be given and the ratio between the starting concentration and the end concentration of the main constituents shall be given, in addition to information about the previous process step(s).

A main identification parameter for residues of non-biological origin is the description of the process from which the residue originates. The process can be any physical reaction that generates residues, e.g. purification, fractioning, concentration process.

#### 4.3.1.3 Analytical information

In cases where spectral data provides information on the composition of the UVCB substance, this information should be given. Several spectroscopic methods are used for generating spectra (UV/Vis, infra-red, nuclear magnetic resonance or mass spectrum). Methods and insights into how to use these methods are subject to continuous change. Therefore, it is the responsibility of the registrant to present appropriate spectral data.

A chromatogram that can be used as a fingerprint shall be provided to characterise the composition of the substance. If applicable, also other valid constituent separation techniques might be used.

### 4.3.2 Specific types of UVCB substances

This section gives guidance on specific groups of UVCB substances: substances with variation in the carbon-chain length (4.3.2.1); substances obtained from oil or oil like sources (4.3.2.2); and enzymes (4.3.2.3).

#### 4.3.2.1 Substances with variation in the carbon-chain lengths

This group of UVCB substances deals with long-chain alkyl substances with variation in the carbon-chain length, e.g. paraffins and olefins. These substances are either derived from natural fats or oils or produced synthetically. The natural fats originate either from plants or animals. Long carbon-chain substances derived from plants have normally only even number chain lengths, whereas long carbon-chain substances obtained from animal sources also include (some) odd number chain lengths. Synthetically produced long carbon-chain substances can comprise the whole range of carbon chains, even and odd numbered.

**Identifiers and naming convention**

The group comprises substances whose individual constituents have a common structural feature: One or more long-chain alkyl group(s) often with an attached functional group. The constituents differ from each other with respect to one or more of the following alkyl-chain group characteristics:

[20] The same approach for identification for natural occurring and chemically produced minerals does not necessarily mean that the legal requirements (e.g. exemptions from registration) are the same.

- Length of carbon chain (carbon number)
- Saturation
- Structure (linear or branched)
- Position of the functional group

The chemical identity of the constituents can be described sufficiently and systematically named by using the following three descriptors:

- The **alkyl descriptor** which describes the number of carbon atoms in the carbon-chain length(s) of the alkyl group(s).
- The **functionality descriptor** which identifies the functional group of the substance, e.g. amine, ammonium, carboxylic acid.
- The **salt descriptor**, the cation / anion of any salt, e.g. sodium ($Na^+$), carbonate ($CO_3^{2-}$), chloride ($Cl^-$).

**Alkyl descriptor**

- In general, the alkyl descriptor $C_{x-y}$ refers to saturated, linear alkyl-chains comprising all chain lengths from x to y, e.g. $C_{8-12}$ corresponds to $C_8$, $C_9$, $C_{10}$, $C_{11}$ and $C_{12}$.
- It has to be indicated, if the alkyl descriptor refers only to even or odd numbered alkyl chains, e.g. $C_{8-12\ (\text{even numbered})}$
- It has to be indicated if the alkyl descriptor refers (also) to branched alkyl chains, e.g. $C_{8-12\ (\text{branched})}$ or $C_{8-12\ (\text{linear and branched})}$
- It has to be indicated if the alkyl descriptor refers (also) to unsaturated alkyl chains, e.g. $C_{12-22\ (\text{C18 unsaturated})}$
- A narrow alkyl chain lengths distribution does not cover a broader one and vice versa, e.g. $C_{10-14}$ does not correspond to $C_{8-18}$
- The alkyl descriptor can also refer to the source of the alkyl chains, e.g. coco or tallow. However, the carbon-chain length distribution must correspond to that of the source.

The above described system should be used to describe substances with variation in the carbon chain lengths. It is not suitable for well defined substances, which can be identified by a definite chemical structure.

The information on the alkyl descriptor, the functionality descriptor and the salt descriptor is the basis for the naming of this type of UVCB substance. In addition, information on the source and the process may be useful to identify the substance more precisely.

| Examples | | |
|---|---|---|
| **Descriptors** | | **Name** |
| **Alkyl descriptor**<br>**Functionality descriptor**<br>**Salt descriptor** | alkyl chain lengths $C_{10-18}$<br>fatty acids (carboxylic acid)<br>cadmium salts | fatty acids (C10-18) cadmium salts |
| **Alkyl descriptor**<br>**Functionality descriptor**<br>**Salt descriptor** | di-$C_{10-18}$-alkyl-dimethyl<br>ammonium<br>chloride | di-C10-18-alkyl-dimethylammonium chloride |
| **Alkyl descriptor**<br>**Functionality descriptor**<br>**Salt descriptor** | trimethyl tallow-alkyl<br>ammonium<br>chloride | trimethyl-tallowalkyl-ammonium chloride |

#### 4.3.2.2 Substances obtained from oil or oil like sources

Substances obtained from oil (petroleum substances) or oil like sources (e.g. coal) are substances of very complex and variable or partly undefined composition. In this chapter petroleum substances are used to demonstrate, how to identify this specific type of a UVCB substance. However, the same approach could be applied to other substances obtained from oil like sources as coal.

The starting materials used in the petroleum refining industry may be crude oil, or any specific refinery stream obtained by one or more processes. The composition of the final products depends on the crude oil used for the manufacture (as the composition of the crude oil varies depending on the place of origin) and the subsequent refinery processes. Therefore, there is natural, process-independent variation in composition of petroleum substances.[16]

**1. Naming convention**

For the identification of petroleum substances, it is recommended to give the name according to an established nomenclature system.[21] This name consists usually of the refinery process, the stream's source and general composition or characteristics. If the substance contains > 5 w/w-% of 4- to 6-membered condensed ring aromatic hydrocarbons, this information shall be included in the description. For petroleum substances with an EINECS number, the name given in the EC Inventory shall be used.

**2. Identifiers**

Terms and definitions for identification of petroleum substances generally include the stream's source, refinery process, general composition, carbon number, boiling range or other appropriate physical characteristics, and predominant hydrocarbon type.[21]

The identification parameters of REACH *Annex VI*, section 2 should be given. It is recognised that petroleum substances are manufactured to performance specifications rather than to compositional specifications. Therefore, characteristics like the name, carbon-chain length range, boiling point, viscosity, cut-off values and other physical properties are generally more helpful than compositional information in order to identify the petroleum substance as clearly as possible.

---

[21] US EPA (1978) TSCA PL 94-469 Candidate list of chemicals substances Addendum I. Generic terms covering petroleum refinery process streams. US EPA, Office of Toxic Substances, Washington DC 20460.

Although chemical composition is not the primary identifier for UVCB substances, the known main constituents (≥ 10%) shall be given and the composition shall be described in generic terms e.g. molecular weight range, aliphatics or aromatics, degree of hydrogenation and other essential information. Moreover, any other constituent at lower concentration which has impact on the hazard classification shall be identified with name and typical concentration.

#### 4.3.2.3 Enzymes

Enzymes are most often produced by fermentation of micro organisms, but occasionally from plant or animal origin. The liquid enzyme concentrate, resulting from the fermentation or extraction and subsequent purification steps contains, besides water, the active enzyme protein and other constituents comprising residues from the fermentation, i.e. proteins, peptides, amino acids, carbohydrates, lipids and inorganic salts.

The enzyme protein together with the other constituents deriving from the fermentation or extraction process, but excluding any water, which may be separated without affecting the stability of the enzyme protein or changing its composition, should be regarded as the substance for identification purposes.

The enzyme substance typically contains 10-80 % (w/w) of the enzyme protein. The other constituents vary in percentage and depend on the production organism used, the fermentation medium, and operational parameters of the fermentation process as well as the downstream purification applied, but the composition will typically be within the ranges indicated in the following table.

| | |
|---|---|
| Active enzyme protein | 10 - 80% |
| Other proteins + peptides and amino acids | 5 - 55% |
| Carbohydrates | 3 - 40% |
| Lipids | 0 - 5% |
| Inorganic salts | 1 - 45% |
| Total | 100% |

The enzyme substance should be regarded as a 'UVCB-substance' due to its variability and partly unknown composition. The enzyme protein should be regarded as a constituent of the UVCB substance. Highly purified enzymes may be identified as substances of well defined composition (mono-constituent or multi-constituent) and should be identified accordingly.

In EINECS, the main identifier for enzymes is the catalytic activity. Enzymes are listed as generic entries without further specification or with specific entries indicating the source organism or the substrate.

| **Examples** | | |
|---|---|---|
| **EC number** | **EINECS name** | **CAS number** |
| 278-547-1 | Proteinase, Bacillus neutral | 76774-43-1 |
| 278-588-5 | Proteinase, Aspergillus neutral | 77000-13-6 |
| 254-453-6 | Elastase (pig pancreas) | 39445-21-1 |
| 262-402-4 | Mannanase | 60748-69-8 |

A study on enzymes commissioned by the European Commission suggested identifying enzymes according to the international system for enzyme nomenclature, IUBMB (International Union of Biochemistry and Molecular Biology).[22] This approach is taken over in this guidance

[22] UBA (2000) Umweltbundesamt Austria. Collection of Information on Enzymes. Final report. Co-operation between Federal Environment Agency Austria and Inter-University Research Center for Technology, Work and Culture

document and will enable a more systematic, detailed and comprehensive identification of enzymes compared to EINECS.

## 1. Naming convention

Enzymes are named according to the IUBMB nomenclature conventions.

The IUBMB classification system provides a unique four digit number for each enzyme type and catalytic function (e.g. 3.2.1.1 for a-amylase)[23]. Each number may comprise enzymes of variable amino acid sequence and origin but the enzyme functionality is identical. The name and number from the IUBMB nomenclature should be used for substance identification. The IUBMB nomenclature classifies the enzymes into six main groups:

1. Oxidoreductases
2. Transferases
3. Hydrolases
4. Lyases
5. Isomerases
6. Ligases

The following example is given to illustrate an entry according to the IUBMB nomenclature:

**EC 3.4.22.33**

**Accepted name:** fruit bromelain

**Reaction:** Hydrolysis of proteins with broad specificity for peptide bonds. Bz-Phe-Val-Arg┼NHMec is a good synthetic substrate, but there is no action on Z-Arg-Arg-NHMec (*c.f.* stem bromelain)

**Other name(s):** juice bromelain; ananase; bromelase; bromelin; extranase; juice bromelain; pinase; pineapple enzyme; traumanase; fruit bromelain FA2

**Comments:** From the pineapple plant, *Ananas comosus*. Scarcely inhibited by chicken cystatin. Another cysteine endopeptidase, with similar action on small molecule substrates, pinguinain (formerly EC 3.4.99.18), is obtained from the related plant, *Bromelia pinguin*, but pinguinain differs from fruit bromelain in being inhibited by chicken cystatin [4].[24] In peptidase family C1 (papain family). Formerly EC 3.4.22.5 and included in EC 3.4.22.4, CAS registry number: 9001-00-7

**Links to other databases:** BRENDA, EXPASY, MEROPS,

**General References:**

---

(IFF/IFZ). Contract No B4-3040/2000/278245/MAR/E2.

23 The terms “EC number” (≡ Enzyme Commission number) and “IUBMB number” are often used as synonyms. In order to avoid misunderstandings, it is recommended to use the term “IUBMB number” for the four numbers code from the IUBMB.

24 Rowan, A.D., Buttle, D.J. and Barrett, A.J. The cysteine proteinases of the pineapple plant. Biochem. J. 266 (1990) 869-875. [Medline UI: 90226288]

Sasaki, M., Kato, T. and Iida, S. Antigenic determinant common to four kinds of thiol proteases of plant origin. *J. Biochem. (Tokyo)* 74 (1973) 635-637. [Medline UI: 74041600]

Yamada, F., Takahashi, N. and Murachi, T. Purification and characterization of a proteinase from pineapple fruit, fruit bromelain FA2. *J. Biochem. (Tokyo)* 79 (1976) 1223-1234. [Medline UI: 76260156]

Ota, S., Muta, E., Katanita, Y. and Okamoto, Y. Reinvestigation of fractionation and some properties of the proteolytically active components of stem and fruit bromelains. *J. Biochem. (Tokyo)* 98 (1985) 219-228. [Medline UI: 86008148]

| **Examples for enzyme classification according IUBMB system** (http://www.chem.qmul.ac.uk/iubmb/enzyme/index.html) | |
|---|---|
| **Proteases are numbered by the following criteria:** | |
| 3. | **Hydrolases** |
| 3.4 | **Acting on peptide bonds (peptidases), with subclasses:** |
| 3.4.1 | α-Amino-Acyl-Peptide Hydrolases (now in EC 3.4.11) |
| 3.4.2 | Peptidyl-Amino-Acid Hydrolases (now in EC 3.4.17) |
| 3.4.3 | Dipeptide Hydrolases (now in EC 3.4.13) |
| 3.4.4 | Peptidyl Peptide Hydrolases (now reclassified within EC 3.4) |
| 3.4.11 | Aminopeptidases |
| 3.4.12 | Peptidylamino-Acid Hydrolases or Acylamino-Acid Hydrolases (now reclassified within 3.4) |
| 3.4.13 | Dipeptidases |
| 3.4.14 | Dipeptidyl-peptidases and tripeptidyl-peptidases |
| 3.4.15 | Peptidyl-dipeptidases |
| 3.4.16 | Serine-type carboxypeptidases |
| 3.4.17 | Metallocarboxypeptidases |
| 3.4.18 | Cysteine-type carboxypeptidases |
| 3.4.19 | Omega peptidases |
| 3.4.21 | Serine endopeptidases |
| | **And further, specific enzymes are identified:** |
| 3.4.21.1 | chymotrypsin |
| 3.4.21.2 | chymotrypsin C |
| 3.4.21.3 | metridin |
| 3.4.21.4 | trypsin |
| 3.4.21.5 | thrombin |
| 3.4.21.6 | coagulation factor Xa |
| 3.4.21.7 | plasmin |
| 3.4.21.8 | now covered by EC 3.4.21.34 and EC 3.4.21.35 |
| 3.4.21.9 | enteropeptidase |
| 3.4.21.10 | acrosin |
| 3.4.21.11 | now covered by EC 3.4.21.36 and EC 3.4.21.37 |
| 3.4.21.12 | 12 a-Lytic endopeptidase |
| … | |
| 3.4.21.105 | |
| 3.4.99 | Endopeptidases of unknown catalytic mechanism |

| Examples from EINECS with IUBMB number added | | | |
|---|---|---|---|
| **EC number** | **EINECS name** | **CAS number** | **IUBMB number** |
| 278-547-1 | Proteinase, Bacillus neutral | 76774-43-1 | 3.4.24.28 |
| 232-752-2 | Subtilisin | 9014-01-1 | 3.4.21.62 |
| 232-734-4 | Cellulase | 9012-54-8 | 3.2.1.4 |

## 2. Identifiers

Enzyme substances are identified by the containing enzyme protein (IUBMB nomenclature) and the other constituents from the fermentation. Beside the enzyme protein, each specific constituent is usually not present in concentrations above 1%. If the identities of these specific constituents are not known, they can be indicated in a grouping approach (i.e. proteins, peptides, amino acids, carbohydrates, lipids and inorganic salts). However, constituents have to be indicated if their identities are known and they have to be identified if their concentration exceeds 10 % or if they are relevant for classification and labelling and-or PBT assessment[25].

**Enzyme proteins**

Enzyme proteins in the concentrate should be identified by

- IUBMB number
- Names given by IUBMB (systemic name, enzyme names, synonyms)
- Comments given by IUBMB
- Reaction and reaction type
- EC number and name, if appropriate
- CAS number and name, if available

The reaction induced by the enzyme should be specified. This reaction is defined by IUBMB.

| Example |
|---|
| .alpha.-amylase: Polysaccharide containing .alpha.-(1-4)-linked glucose units + H2O = maltooligosaccharides; endohydrolysis of 1,4-.alpha.-d-glucosidic linkages in polysaccharides containing three or more 1,4-.alpha.-linked d-glucose units. |

According to the enzyme class, a type of reaction shall be allocated. This can be oxidation, reduction, elimination, addition or a reaction name.

[25] More information on PBT assessment and relevant criteria can be found in the Guidance on information requirements and chemical safety assessment, chapter R11: PBT assessment.

| **Example** |
| --- |
| .alpha.-amylase: O-glycosyl bond hydrolysis (endohydrolysis). |

**Constituents other than the enzyme protein**

All constituents ≥ 10 % (w/w) or relevant for classification and labelling and-or PBT assessment[26] should be identified. The identity of constituents less than 10% can be indicated as a chemical group. Their typical concentration(s) or concentration ranges have to be given, i.e.:

- (Glyco)Proteins
- Peptides and Amino acids
- Carbohydrates
- Lipids
- Inorganic Material (e.g. sodium chloride or other inorganic salts)

If it is not feasible to identify the other constituents of an enzyme concentrate sufficiently, the name of the production organism (Genus and the strain or genetic type if relevant) should be given as for other UVCB substances of biological origin.

If available, additional parameters can be given, e.g. functional parameters (i.e. pH or temperature optima and ranges), kinetic parameters (i.e. specific activity or turnover number), ligands, substrates and products and co-factors.

[26] More information on PBT assessment and relevant concentration limits can be found in RIP 3.2 TGD Chemical Safety assessment section on PBT assessment

# 5 CRITERIA FOR CHECKING IF SUBSTANCES ARE THE SAME

When checking whether or not the substances from different manufacturers/importers can be regarded as the same, some rules should be respected. These rules which were applied for establishing EINECS should be regarded as a common base for identifying and naming a substance and thus finding a potential co-registrant of this particular substance.[5, 6, 16, 27, 28] Substances which are not regarded as the same may, however, be regarded as structurally related by application of expert judgement. Data sharing might, nevertheless, be possible for these substances if scientifically justified. However, this is not a subject of this guidance document rather it is addressed in Guidance on Data Sharing.

- The "≥ 80%" rule for mono-constituent substances as well as the "< 80%/≥ 10%" rule for multi-constituent substances should be applied.

No differentiation is made between technical, pure or analytical grades of the substances. This means that the "same" substance may have a different purity/impurity profile depending on its grade. However, well defined substances should contain the same main constituent(s) and the only impurities allowed are those derived from the production process (for details see Chapter 4.2) and additives which are necessary to stabilize the substance.

- Hydrated and anhydrous forms of compounds shall be regarded as the same substance for the purpose of a registration

| **Examples** | | | |
|---|---|---|---|
| **Name and formula** | **CAS number** | **EC number** | **Rule** |
| Copper sulphate ($Cu \cdot H_2O_4S$) | 7758-98-7 | 231-847-6 | |
| Sulphuric acid copper(2+) salt (1:1), pentahydrate<br>($Cu.H_2O_4\ S \cdot 5\ H_2O$) | 7758-99-8 | | This substance is covered by a registration of it's anhydrous form (EC number: 231-847-6) |

Hydrated and anhydrous forms have different chemical names and different CAS numbers. Detailed information on how to make use of the specific provision for the registration of hydrated forms of a substance in *Annex V(6)* of the REACH Regulation, is given in the Data Submission Manual Part 18 – "How to report the substance identity in IUCLID 5 for registration under REACH".

[27] Vollmer et al. (1998) Compilation of EINECS: Descriptions and definitions used for substances, impurities and mixtures. Tox Env Chem Vol. 65, p. 113-122.

[28] Manual of Decisions, Criteria for reporting substances for EINECS, ECB web-site; Geiss et al. 1992, Vollmer et al. 1998, Rasmussen et al. 1999.

- Acids or bases and their salts shall be regarded as different substances.

| **Examples** | | |
|---|---|---|
| **EC number** | **Name** | **Rule** |
| 201-186-8 | Peracetic acid<br>$C_2H_4O_3$ | This substance shall not be regarded as the same as, for example, its sodium salt (EINECS 220-624-9) |
| 220-624-9 | Sodium glycollate<br>$C_2H_4O_3$ . Na | This substance shall not be regarded as the same as its corresponding acid (EINECS 201-186-8) |
| 202-426-4 | 2-Chloroaniline<br>$C_6H_6ClN$ | This substance shall not be regarded as the same as, for example, 2-chloroaniline hydrobromide (1:1) ($C_6H_6ClN$ . HBr) |

- Individual salts (e.g. sodium or potassium) shall be regarded as different substances.

| **Examples** | | |
|---|---|---|
| **EC number** | **Name** | **Rule** |
| 208-534-8 | Sodium benzoate<br>$C_7H_5O_2$ . Na | This substance shall not be regarded as the same as, for example, the potassium salt (EINECS 209-481-3) |
| 209-481-3 | Potassium benzoate<br>$C_7H_5O_2$ . K | This substance shall not be regarded as the same as, for example, the sodium salt (EINECS 208-534-8) |

- Branched or linear alkyl chains shall be regarded as different substances.

| **Examples** | | |
|---|---|---|
| **EC number** | **Name** | **Rule** |
| 295-083-5 | Phosphoric acid, dipentyl ester, branched and linear | This substance shall not be regarded as the same as the individual substances phosphoric acid, dipentyl ester, branched or phosphoric acid, dipentyl ester, linear |

- Branched groups shall be mentioned as such in the name. Substances containing alkyl groups without any further information cover only the unbranched linear chains unless otherwise specified.

| **Examples** | | |
|---|---|---|
| **EC number** | **Name** | **Rule** |
| 306-791-1 | Fatty acids, C12-16 | Only substances with linear and unbranched alkyl groups are regarded as the same substance |
| 279-420-3 | Alcohols, C12-14 | |
| 288-454-8 | Amines, C12-18-alkylmethyl | |

- Substances with alkyl groups using additional terms like iso, neo, branched etc, shall not be regarded the same as the substances without that specification.

| Examples | | |
|---|---|---|
| **EC number** | **Name** | **Rule** |
| 266-944-2 | Glycerides, $C_{12-18}$<br>This substance is identified by SDA Substance Name: C12-C18 trialkyl glyceride and SDA Reporting Number: 16-001-00 | This substance shall not be regarded the same as $C_{12-18-iso}$ Substance with saturated alkyl chains which is branched at any position |

- Without explicit specification, alkyl chains in acids or alcohols etc. shall be considered to represent only the saturated chains. Unsaturated chains shall be specified as such and are regarded as different substances.

| Examples | | |
|---|---|---|
| **EC number** | **Name** | **Rule** |
| 200-313-4 | Stearic acid, pure $C_{18}H_{36}O_2$ | This substance shall not be regarded the same as Oleic acid, pure, $C_{18}H_{34}O_2$ (EINECS 204-007-1) |

- Substances with chiral centres

A substance with one chiral centre can exist in left and right-handed forms (enantiomers). In the absence of any indication to the contrary, it is assumed that a substance is an equal (racemic) mixture of the two forms.

| Examples | | |
|---|---|---|
| **EC number** | **Name** | **Rule** |
| 201-154-3 | 2-chloropropan-1-ol | The individual enantiomers (R)-2-chloropropan-1-ol and (S)-2-chloropropan-1-ol are not regarded equal to this entry |

Where a substance has been enriched with a single enantiomeric form, the rules for multi-constituent substances apply. Similarly, racemates are considered as multi-constituent substances.

Substances with multiple chiral centres can exist in 2n forms (where n is the number of chiral centres). These different forms can have different physico-chemical, toxicological and/or eco-toxicological properties to each other. They should be regarded as separate substances.

- Inorganic catalysts

Inorganic catalysts are regarded as mixtures. For identification purposes, component metals or metallic compounds should be considered as individual substances (without specification of use).

| Examples | | |
|---|---|---|
| | **Name** | **Rule** |
| | Cobalt oxide-aluminium oxide catalyst | Should be identified seperately as:<br>- Cobalt II oxide<br>- Cobalt III oxide<br>- Aluminium oxide<br>- Aluminium cobalt oxide |

- Enzyme concentrates with the same IUBMB number can be regarded as the same substance, despite using different production organism, provided that the hazardous properties do not differ significantly and warrant the same classification.

### Multi-constituent substances

Directive 67/548/EEC regulated the placing of substances on the market. The production manner of the substance was not relevant. Therefore, a marketed multi-constituent substance was covered by EINECS, if all the single constituents were listed on EINECS; e.g. the isomeric mixture difluorobenzenes was covered by the EINECS entries 1,2-Difluorobenzene (206-680-7), 1,3-Difluorobenzene (206-746-5) and 1,4-Difluorobenzene (208-742-9) although the isomeric mixture itself was not listed on EINECS.

REACH instead requires the registration of the manufactured substance. It is a case by case decision to establish to what extent the different steps while producing the substance are covered by the definition 'manufacturing' (e.g. different purification or distillation steps). If a multi-constituent substance is produced it has to be registered (unless it is covered by a registration of the individual constituents, see chapter 4.2.2.4); e.g. the isomeric mixture diflurobenzene is produced, thus "diflurobenzene", as an isomeric mixture, has to be registered. However, for multi-constituent substances, there is no need to test the substance as such, if the hazard profile of the substance can be sufficiently described by the information of the individual constituents. If the individual isomers 1,2-Difluorobenzene, 1,3-Difluorobenzene and 1,4-Difluorobenzene are produced and mixed afterwards, the individual isomers have to be registered and the isomeric mixture would be regarded as a mixture.

A multi-constituent substance of main constituents A, B and C shall not be regarded as same as a multi-constituent substance of main constituents A and B or as a reaction mass of A, B, C and D.

- A multi-constituent substance is not regarded equal to a substance with only a subset of the single constituents.

| **Examples** | | |
|---|---|---|
| **EC number** | **Name** | **Rule** |
| 207-205-6 | 2,5-Difluorotoluene | These two substances are not regarded as the same as the isomeric mixture difluorotoluenes because these two substances are only a subset of all possible isomers. |
| 207-211-9 | 2,4-Difluorotoluene | |

- The registration of a multi-constituent substance does not cover the individual constituents.

| **Examples** | | |
|---|---|---|
| **EC number** | **Name** | **Rule** |
| 208-747-6 | 1,2-Dibromoethylene | This substance describes a mixture of cis- and trans-isomers. The individual substances (1Z)-1,2-Dibromoethene and (1E)-1,2-Dibromoethene are not covered by the registration of the isomeric mixture. |

### UVCB substances

- A UVCB substance with a narrow distribution of constituents is not regarded as equal to a UVCB substance with a broader composition and vice versa.

| **Examples** | | |
|---|---|---|
| **EC number** | **Name** | **Rule** |
| 288-450-6 | Amines, $C_{12\text{-}18\text{-alkyl}}$, acetates | The substances "amines, $C_{12\text{-}14\text{-alkyl}}$, acetates" or "amines, $C_{12\text{-}20\text{-alkyl}}$, acetates" or "amines, dodecyl ($C_{12\text{-alkyl}}$), acetates" |

| | | |
|---|---|---|
| | | or substances with only even-numbered alkyl-chains are not regarded equal to this substance |

- A substance which is characterised by a species/genus is not regarded as the same as a substance isolated from another species/genus.

| **Examples** | | |
|---|---|---|
| **EC number** | **Name** | **Rule** |
| 296-286-1 | Glycerides, sunflower-oil di- | This substance is not regarded as the same as Glycerides, soya di- (EINECS: 271-386-8), nor the same as Glycerides, tallow di- (EINECS: 271-388-9) |
| 232-401-3 | Linseed oil, epoxidized | This substance is not regarded as the same as linseed oil, oxidized (EINECS: 272-038-8), nor the same as linseed oil, maleated (EINECS: 268-897-3), nor as castor oil, epoxidized (not listed in EINECS). |

- A purified extract or a concentrate is regarded as a different substance than the extract.

| **Examples** | | |
|---|---|---|
| **EC number** | **Name** | **Rule** |
| 232-299-0 | Rape oil<br><br>Extractives and their physically modified derivates. It consists primarily of the glycerides of the fatty acids erucic, linoleic and oleic. (Brassica napus, Cruciferae) | The substance “(Z)-Docos-13-enoic acid (erucic acid)” is a constituent of the substance “rape oil”. Erucic acid is not regarded as the same as rape oil as it is isolated as a pure substance from the rape oil; Erucic acid has its own EINECS entry (204-011-3).<br><br>An isolated mixture of palmitic acid, oleic acid, linoleic acid, linolenic acid, erucic acid and eicosenoic acid is not regarded as the same as rape oil as these constituents do not represent the whole oil. |

# 6 SUBSTANCE IDENTITY WITHIN (LATE) PRE-REGISTRATION AND INQUIRY

Guidance on how to identify and name substances is provided in Chapter 4 of this guidance document. This guidance should be followed to determine whether substances could be considered to be the same for the purpose of REACH and CLP. This is further elaborated below for (late) pre-registration of phase-in substances and inquiry of non phase-in substances.

According to *Article 4*, any manufacturer or importer may, whilst retaining full responsibility for complying with his obligation under the REACH Regulation, appoint a third party representative for all proceedings under Title III involving discussions with other manufacturers or importers.

## 6.1 (LATE) PRE-REGISTRATION

The aim of the (late) pre-registration process is to bring potential registrants of the same substance together to avoid duplication of studies, in particular testing on vertebrate animals. (Late) pre-registration only applies to phase-in substances.

More information on (late) pre-registration can be found in Guidance on Data Sharing at http://guidance.echa.europa.eu/guidance_en.htm.

## 6.2 INQUIRY

For non-phase in substances, or phase-in-substances that have not been pre-registered, there is a duty for the potential registrant to inquire from the Agency prior to registration, whether a registration has already been submitted for the same substance (*Article 26* of REACH). This inquiry shall contain:

- the identity of the potential registrant as specified in section 1 of *Annex VI* of the REACH Regulation with the exception of the use sites;
- the identity of the substance, as specified in section 2 of *Annex VI* of the REACH Regulation;
- which information requirements would require new studies involving vertebrate animals to be carried out by the potential registrant;
- which information requirements would require other new studies to be carried out by the potential registrant.

The potential registrant should provide the identity and the name of the substance according the rules laid down in Chapter 4 of this guidance document.

The Agency shall establish whether the same substance has previously been registered. This shall also be done, by applying the rules laid down in Chapter 4 of this guidance document. The result is communicated back to the potential registrant and any previous or other potential registrants are informed.

More information on inquiry process can be found in the Guidance on data sharing at http://guidance.echa.europa.eu/guidance_en.htm and on the dedicated ECHA web page http://echa.europa.eu/reachit/inquiry_en.asp.

# 7 EXAMPLES

The examples given in the following pages are meant only to illustrate how the user could work with the guidance in this guidance document. They do not present any precedent regarding duties concerning REACH.

The following examples are included:

- 'Diethyl peroxydicarbonate' is an example for a mono-constituent substance including a solvent which is also acting as a stabilizing agent (see Chapter 7.1);
- 'Zolimidine' is an example for a substance which could be identified as mono-constituent or as multi-constituent substance (see Chapter 7.2);
- A 'mixture of isomers' formed during the manufacturing reaction is included as an example for a multi-constituent substance (see Chapter 7.3). This substance was previously covered by the EINECS entries of the individual isomers;
- 'Fragrance AH' is an example for a substance produced in different qualities, which can be described by a reaction mass of five constituents with concentration ranges (Chapter 7.4). It is also an example for a justified deviation from the 80% rule and the 10% rule;
- Non-metallic 'minerals', including montmorillonite as an example of a well defined substance, that requires additional physical characterization, are included in Chapter 7.5;
- An 'essential oil of lavendula' is an example for a UVCB substance obtained from plants (Chapter 7.6);
- 'Chrysanthemum oil and isomers isolated thereof' is an example for a UVCB substance of biological origin, which is further processed (Chapter 7.7);
- 'Phenol, isopropylated, phosphate' is an example for a variable UVCB substance, which cannot be fully defined (Chapter 7.8);
- 'Quaternary ammonium compounds' are examples for substances with variation in the carbon-chain length (Chapter 7.9);
- Two examples for 'petroleum substances', a gasoline blending stream and gas oils, are included in Chapter 7.10;
- Two examples, how to identify enzymes, laccase and amylase, are given in Chapter 7.11.

## 7.1 DIETHYL PEROXYDICARBONATE

The substance 'diethyl peroxydicarbonate' (EC 238-707-3, CAS 14666-78-5, $C_6H_{10}O_6$) is produced as an 18% solution in isododecane (EC 250-816-8, CAS 31807-55-3). Isododecane is also acting as a stabilizing agent against explosive properties. The highest possible concentration which guarantees safe handling of the substance is a 27% solution.

How should the above described substance be identified and named for registration?

According to the substance definition in REACH, solvents which may be separated without affecting the stability of the substance or changing its composition should be excluded. As in the above case, isododecane is also acting as stabilizing agent and cannot be totally separated due to explosive properties of the substance, isododecane has to be regarded as an additive and not as a solvent only. However the substance should still be regarded as a mono-constituent substance. Therefore, the substance should be registered as the solution with the lowest concentration of isododecane which guarantees safe handling:

Diethyl peroxydicarbonate (upper concentration limit: 27%). Isodecane should be reported under “Additives” and the stabilising function should be specified.

## 7.2 ZOLIMIDINE

The manufactured methanolic solution contains ‘zolimidine’ (EC 214-947-4; CAS 1222-57-7, $C_{14}H_{12}N_2O_2S$) and ‘imidazole’(EC 206-019-2; CAS 288-32-4, $C_3H_4N_2$). After removing the solvent “methanol” and optimizing the manufacturing process the substance has purity range of 74 – 86% zolimidine and 4-12%.imidazole.

How should the above described substance be identified and named for registration?

According to the substance definition in REACH, solvents which may be separated without affecting the stability of the substance or changing its composition should be excluded. As in the above case, methanol can be separated without any difficulties; the solvent-free substance has to be registered.

In general, a substance is regarded as a mono-constituent substance, if one main constituent is present ≥ 80%. A substance is regarded as multi-constituent substance, if more than one main constituent is ≥ 10% and < 80%. The above example is a borderline case, as the threshold values are crossed. Therefore the substance could be regarded as a mono-constituent substance “zolimidine” or as a multi-constituent substance, a reaction mass of “zolimidine” and “imidazole”.

In such a borderline case, the typical concentration of the main constituents of the substance can be used to decide how best to describe this substance as follows:

(1) If the typical concentration for zolimidine is 77% and for imidazole it is 11%, then it is recommended to regard the substance as a reaction mass of zolimidine and imidazole;

(2) If the typical concentration is for zolimidine is 85% and for imidazole it is 5%, then it is recommended to regard the substance as mono-constituent substance “zolimidine”.

## 7.3 MIXTURE OF ISOMERS

The substance in question is a mixture (reaction mass) of two isomers formed during the manufacturing reaction. The individual isomers were reported for EINECS. Directive 67/548/EEC regulated the placing of substances on the market. As the production manner of the substance was not significant, the mixture was covered by the EINECS entries of the two individual isomers. REACH requires the registration of manufactured substances. It is a case by case decision to establish to what extent the different steps conducted while producing the substance are covered by the definition of ‘manufacturing’. If the isomer mixture is registered as a multi-constituent substance (following the guidance of Chapter 4.2.2), there is no need to test the substance as such, if the hazard profile of the substance can be sufficiently described by the

information of the individual constituents. However, reference should be made to the EINECS entries of the individual isomers to demonstrate the phase-in status.

## 1. Name and other identifiers

| | |
|---|---|
| **IUPAC name or other international chemical name (of the substance)** | Reaction mass of 2,2'-[[(4-methyl-1H-benzotriazol-1-yl)methyl]imino]bisethanol and 2,2'-[[(5-methyl-1H-benzotriazol-1-yl)methyl]imino]bisethanol |
| **Other names (of the substance)** | 2,2'-[[(methyl-1H-benzotriazol-1-yl)methyl]imino]bisethanol<br>Reaction mass of Ethanol, 2,2'-[[(methyl-1H-benzotriazol-1-yl)methyl]imino]bis- and water<br>Ethanol, 2,2'-[[(methyl-1H-benzotriazol-1-yl)methyl]imino]bis- (9CI) isomeric compound |
| **EC number (of the substance)**<br>**EC name**<br>**EC description** | There exists no EC number for the mixture, as the mixture was not reported for EINECS. However, the substance was covered by the EINECS entries for the constituents (279-502-9, 279-501-3). Therefore, the mixture should be regarded as phase-in substance. |
| **CAS number (of the substance)**<br>**CAS name** | not available<br>not available |
| **EC number (constituent A)**<br>**EC name**<br>**EC description** | 279-502-9<br>2,2'-[[(4-methyl-1H-benzotriazol-1-yl)methyl]imino]bisethanol<br>/ |
| **EC number (constituent B)**<br>**EC name**<br>**EC description** | 279-501-3<br>2,2'-[[(5-methyl-1H-benzotriazol-1-yl)methyl]imino]bisethanol<br>/ |
| **CAS number (constituent A)**<br>**CAS name** | 80584-89-0<br>Ethanol, 2,2'-[[(4-methyl-1H-benzotriazol-1-yl)methyl]imino]bis- |
| **CAS number (constituent B)**<br>**CAS name** | 80584-88-9<br>Ethanol, 2,2'-[[(5-methyl-1H-benzotriazol-1-yl)methyl]imino]bis- |
| **Other identity code**<br>**Reference** | ENCS number 5-5917 |

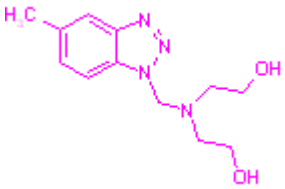
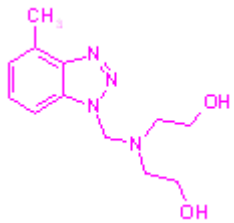
## 2. Composition information – main constituents

| **Main constituents** | | | | | | |
|---|---|---|---|---|---|---|
| | **IUPAC name** | **CAS number** | **EC number** | **Mol. formula**<br>**Hill method** | **Typical conc. (%w/w)** | **Conc. range (%w/w)** |
| **A** | Ethanol, 2,2'-[[(4-methyl-1H-benzotriazol-1-yl)methyl]imino]bis- | 80584-89-0 | 279-502-9 | C12H18N4O2 | 60 | 50-70 |
| **B** | Ethanol, 2,2'-[[(5-methyl-1H-benzotriazol-1-yl)methyl]imino]bis- | 80584-88-9 | 279-501-3 | C12H18N4O2 | 40 | 30-50 |

| Main constituents | |
|---|---|
| | **Other names:** |
| **A** | 2,2'-[[(4-methyl-1H-benzotriazol-1-yl)methyl]imino]bisethanol |
| **B** | 2,2'-[[(5-methyl-1H-benzotriazol-1-yl)methyl]imino]bisethanol |

| Main constituents | | |
|---|---|---|
| | **EC name** | **EC description** |
| **A** | 2,2'-[[(4-methyl-1H-benzotriazol-1-yl)methyl]imino]bisethanol | / |
| **B** | 2,2'-[[(5-methyl-1H-benzotriazol-1-yl)methyl]imino]bisethanol | / |

| Main constituents | | |
|---|---|---|
| | **CAS name** | **CAS numbers** |
| **A** | Ethanol, 2,2'-[[(4-methyl-1H-benzotriazol-1-yl)methyl]imino]bis- | 80584-89-0 |
| **B** | Ethanol, 2,2'-[[(5-methyl-1H-benzotriazol-1-yl)methyl]imino]bis- | 80584-88-9 |

| Main constituents | | | |
|---|---|---|---|
| | **Molecular Formula**<br>**CAS method** | **Structural formula** | **SMILES code** |
| **A** | / |  | OCCN(CCO)Cn2nnc1cc(C)ccc12 |
| **B** | / |  | OCCN(CCO)Cn2nnc1c(C)cccc12 |

| Main constituents | | |
|---|---|---|
| | **Molecular weight [g mol$^{-1}$]** | **Molecular weight range** |
| **A** | 250 | / |
| **B** | 250 | / |

## 7.4 FRAGRANCE AH

Fragrance AH consists of gamma (iso-alpha) methyl ionone and its isomers. It is produced in three different qualities (quality A, B and C), which differ in the ratio of the isomers.

The following table gives an overview about the composition of the different qualities.

Composition of the different qualities of Fragrance AH

| Concentration range [%] | Quality A | Quality B | Quality C | Overall ranges |
|---|---|---|---|---|
| gamma (iso-alpha) methyl ionone | 80 - 85 | 65 - 75 | 50 - 60 | 50 - 85 |
| delta (iso-beta) methyl ionone | 6 - 10 | 3 - 7 | 3 – 7 | 3 - 10 |
| alpha n-methyl ionone | 3 - 11 | 10 - 20 | 20 - 30 | 3 - 30 |
| gamma n-methyl ionone | 0.5 - 1.5 | 2 - 4 | 2 - 4 | 0.5 - 4 |
| beta n-methyl ionone | 0.5 - 1.5 | 4 - 6 | 5 - 15 | 0.5 -15 |
| pseudo methyl ionones | 0.5 - 1.5 | 1 - 3 | 1 - 3 | 0.5 - 3 |

There are several options for substance identification:

- Quality A contains at least 80% of the gamma (iso-alpha) methyl ionone isomer and could therefore be regarded as a mono-constituent substance based on the gamma (iso-alpha) methyl ione isomer with the other isomers as impurities.

- Qualities B & C contain less than 80% of the gamma (iso-alpha) methyl ionone isomer and ≥ 10% of other isomers. Therefore they could be regarded as multi-constituent substances:
    - Quality B: as a reaction mass of gamma(iso-alpha) methyl ionone (65–75%) and alpha-n methyl ionone (10-20%) with the other isomers as impurities.
    - Quality C: as a reaction mass of gamma(iso-alpha) methyl ionone (50-60%) and alpha-n methyl ionone (20-30%) with the other isomers as impurities.

The composition is variable and sometimes an isomer is present as ≥ 10% (therefore normally called main constituent) and sometimes < 10% (therefore normally called impurity).

It would be possible to register the different qualities separately. This would imply three registrations. However, read-across of data may be justified.

Alternatively one can consider:

- One registration as a mono-constituent substance with two sub-qualities. In this case the sub-qualities deviate from the 80% rule (see Chapter 4.2.1);

- One registration as a defined reaction mass of 5 isomers (multi-constituent substance). In this case some isomers (main constituents) deviate from the 10% rule which distinguishes main constituents from impurities (see Chapter 4.2.2).

- One registration as a defined reaction mass where the variability of the composition is covered by the full range for each isomer.

It may be important to consider that

- The three qualities have the same or very similar physico-chemical properties.

- The three qualities have similar use and exposure scenarios.

- All qualities have the same hazard classification and labelling and the contents of the safety data sheets and safety reports are identical

- Available test data (and future testing) cover the variability of the three qualities.

In this example the identification of the substance as a defined reaction mass of 5 isomers (multi-constituent substance) is described. A justification is needed because of the deviation from the 80% rule (see Chapter 4.2.1) and the 10% rule (see Chapter 4.2.2). As each quality is produced as such, the composition of each of the three qualities should be specified in the registration dossier. However, under formal conditions at least two registrations could be necessary: (1) Gamma (iso-alpha) methyl ionone and (2) Reaction mass of gamma (iso-alpha) methyl ionone and alpha-n-methyl ionone.

**Substance identification**

Fragrance AH is produced in three different qualities (A, B and C) with the same qualitative but different quantitative composition. All three qualities are described in one registration dossier for a multi-constituent substance. Although this implies that the 80% and the 10% rule are not applied strictly, the registration as one multi-constituent substance is justified, as (1) available test data cover the variability of the three qualities, (2) the three qualities have very similar physico-chemical properties, (3) all qualities have the same hazard classification and labelling (thus, the safety data sheets are identical), and (4) the three qualities have similar use and exposure scenarios (thus, similar chemical safety reports).

**1. Name and other identifiers**

| | |
|---|---|
| IUPAC name or other international chemical name | Reaction mass of<br>3-methyl-4-(2,6,6-trimethyl-2-cyclohexen-1-yl)but-3-en-2-one;<br>3-methyl-4-(2,6,6-trimethyl-1-cyclohexen-1-yl)but-3-en-2-one;<br>[R-(E)]-1-(2,6,6-trimethyl-2-cyclohexen-1-yl)pent-1-en-3-one;<br>1-(6,6-methyl-2-methylenecyclohex-1-yl)pent-1-en-3-one;<br>1-(2,6,6-trimethyl-1-cyclohexen-1-yl)pent-1-ten-3-one |
| Other names | Methyl Ionone Gamma Quality A<br>Methyl Ionone Gamma Quality B<br>Methyl Ionone Gamma Quality C |
| EC number<br>EC name<br>EC description | not available<br>/<br>/ |

| | |
|---|---|
| CAS number | not available |
| CAS name | / |

## 2. Composition information – main constituents

In theory, additional entiomeres are possible. However, the following isomers were analysed:

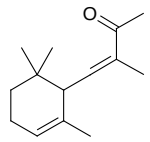
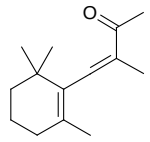
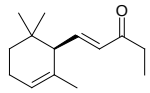
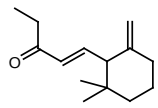
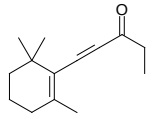
| Main constituents | | | | | | |
|---|---|---|---|---|---|---|
| | **IUPAC name** | **CAS number** | **EC number** | **Mol. formula Hill method** | **Min. conc. (%w/w)** | **Max. conc. (%w/w)** |
| **A** | 3-methyl-4-(2,6,6-trimethyl-2-cyclohexen-1-yl)but-3-en-2-one | 127-51-5 | 204-846-3 | C14H22O | 50 | 85 |
| **B** | 3-methyl-4-(2,6,6-trimethyl-1-cyclohexen-1-yl)but-3-en-2-one | 79-89-0 | 201-231-1 | C14H22O | 3 | 10 |
| **C** | [R-(E)]-1-(2,6,6-trimethyl-2-cyclohexen-1-yl)pent-1-en-3-one | 127-42-4 | 204-842-1 | C14H22O | 3 | 30 |
| **D** | 1-(6,6-methyl-2-methylenecyclohex-1-yl)pent-1-en-3-one | not available | not available | C14H22O | 0.5 | 4 |
| **E** | 1-(2,6,6-trimethyl-1-cyclohexen-1-yl)pent-1-en-3-one | 127-43-5 | 204-843-7 | C14H22O | 0.5 | 15 |

| Main constituents | |
|---|---|
| | **Other names:** |
| **A** | alpha-iso-methyl ionone; gamma methyl ionone |
| **B** | beta-iso-methyl ionone; delta methyl ionone |
| **C** | alpha-n-methyl ionone |
| **D** | gamma-n-methyl ionone |
| **E** | beta-n-methyl ionone |

| Main constituents | | |
|---|---|---|
| | **EC name** | **EC description** |
| **A** | 3-methyl-4-(2,6,6-trimethyl-2-cyclohexen-1-yl)-3-buten-2-one | / |
| **B** | 3-methyl-4-(2,6,6-trimethyl-1-cyclohexen-1-yl)-3-buten-2-one | / |
| **C** | [*R-(E)*]-1-(2,6,6-trimethyl-2-cyclohexen-1-yl)pent-1-en-3-one | / |
| **D** | 1-(2,6,6-trimethyl-2-cyclohexen-1-yl)pent-1-en-3-one | / |
| **E** | 1-(2,6,6-trimethyl-1-cyclohexen-1-yl)pent-1-en-3-one | / |

| Main constituents | | |
|---|---|---|
| | **CAS name** | **CAS number** |
| **A** | 3-Buten-2-one, 3-methyl-4-(2,6,6-trimethyl-2-cyclohexen-1-yl)- | 127-51-5 |
| **B** | 3-Buten-2-one, 3-methyl-4-(2,6,6-trimethyl-1-cyclohexen-1-yl)- | 79-89-0 |
| **C** | 1-Penten-3-one, 1-[(1R)-2,6,6-trimethyl-2-cyclohexen-1-yl)]-, (1E)- | 127-42-4 |
| **D** | not available | not available |
| **E** | 1-Penten-3-one, 1-(2,6,6-trimethyl-1-cyclohexen-1-yl)- | 127-43-5 |

| Main constituents | | |
|---|---|---|
| | **Other identity code** | **Reference** |
| **A** | 2714<br>07.036 | FEMA<br>EU Flavour Register |
| **B** | 07.041 | EU Flavour Register |
| **C** | 2711<br>07.009 | FEMA<br>EU Flavour Register |
| **D** | not available | not available |
| **E** | 2712<br>07.010 | FEMA<br>EU Flavour Register |

| Main constituents | | | |
|---|---|---|---|
| | **Molecular Formula**<br>**CAS method** | **Structural formula** | **SMILES code** |
| **A** | $C_{14}H_{22}O$ |  | O=C(C(=CC(C(=CCC1)C)C1(C)C)C)C |
| **B** | $C_{14}H_{22}O$ |  | O=C(C(=CC(=C(CCC1)C)C1(C)C)C)C |
| **C** | $C_{14}H_{22}O$ |  | O=C(C=CC(C(=CCC1)C)C1(C)C)CC |
| **D** | $C_{14}H_{22}O$ |  | C=C1CCCC(C)(C)C1/C=C/C(=O)CC |
| **E** | $C_{14}H_{22}O$ |  | O=C(C=CC(=C(CCC1)C)C1(C)C)CC |

| Main constituents | | |
|---|---|---|
| | **Molecular weight / gmol$^{-1}$** | **Molecular weight range** |
| **A** | 206.33 | / |
| **B** | 206.33 | / |
| **C** | 206.33 | / |
| **D** | 206.33 | / |
| **E** | 206.33 | / |

### 3. Composition information – impurities and additives

| Impurities | | | | | | |
|---|---|---|---|---|---|---|
| | **IUPAC name** | **CAS number** | **EC number** | **Mol. formula** | **Typical conc. (%w/w)** | **Conc. range (%w/w)** |
| **F** | | | | | | |
| number of non-specified impurities: | | | | 11 (pseudo methyl ionones) | | |
| total concentration of non-specified impurities: | | | | 0.5 – 3%w/w | | |
| **Additives** | | | | | | |
| | **IUPAC name** | **CAS number** | **EC number** | **Mol. formula** | **Typical conc. (%w/w)** | **Conc. range (%w/w)** |
| **G** | Butylated Hydroxytoluene (BHT) | 128-37-0 | 204-881-4 | C15H24O | 0.1 | 0.05 – 0.15 |

### 4. Information on the different qualities

Below are the ranges of the five main constituents in the three different qualities:

| Concentration range [%] | Quality A | Quality B | Quality C |
|---|---|---|---|
| gamma (iso-alpha) methyl ionone | 80 - 85 | 65 - 75 | 50 - 60 |
| delta (iso-beta) methyl ionone | 6 - 10 | 3 - 7 | 3 – 7 |
| alpha n-methyl ionone | 3 - 11 | 10 - 20 | 20 - 30 |
| gamma n-methyl ionone | 0.5 - 1.5 | 2 - 4 | 2 - 4 |
| beta n-methyl ionone | 0.5 - 1.5 | 4 - 6 | 5 - 15 |
| pseudo methyl ionones | 0.5 - 1.5 | 1 - 3 | 1 - 3 |

## 7.5 MINERALS

A mineral is defined as a combination of inorganic constituents as found in the earth's crust, with a characteristic set of chemical compositions, crystalline forms (from highly crystalline to amorphous), and physico-chemical properties.

Minerals are exempted from registration, if they meet the definition of a substance which occurs in nature (*Article 3(39)* of REACH) and if they are not chemically modified (*Article 3(40)* of REACH). This applies to minerals whose chemical structure remains unchanged, even if it has

undergone a chemical process or treatment, or a physical mineralogical transformation, for instance to remove impurities.

While some minerals can be described uniquely by their chemical composition (see Chapter 4.2.1 and 4.2.2 for mono-constituent and multi-constituent substances), for others the chemical composition alone is not sufficient to uniquely identify these substances (see Chapter 4.2.3).

Contrary to other mono- or multi-constituent substances, the identification of many minerals must be based on chemical composition and internal structure (e.g. as revealed by X-ray diffraction), because these together represent the essence of the mineral and determine its physico-chemical properties.

As for other multi-constituent substances, the CAS number for the mineral shall be used as part of the identification (i.e. the combination of inorganic constituents). The CAS numbers of the inorganic constituents (as defined by systematic mineralogy) are used to describe the different constituents. If an individual inorganic constituent would be produced (a mono-constituent substance) the CAS number of this substance should be used for the identification of the substance. For instance:

- The mineral Kaolin (EINECS: 310-194-1, CAS: 1332-58-7) is basically composed of primary and secondary Kaolinites (EINECS: 215-286-4, CAS: 1318-74-7) which is a hydrated aluminosilicate clay.

In the case that a refinement process would be applied to Kaolin to produce a single constituent of Kaolone, e.g. Kaolinites than the CAS- / EINECS-number for the substance would be EINECS: 215-286-4, CAS: 1318-74-7.

- The mineral Bentonite (EINECS: 215-108-5, CAS: 1302-78-9) which is described in EINECS as “A colloidal clay. Consists primarily of montmorillonite” contains in a high proportion the inorganic constituent Montmorillonite (EINECS: 215-288-5, CAS: 1318-93-0) but not only.

In the case that the pure Montmorillonite (EINECS: 215-288-5, CAS: 1318-93-0) would be produced than the CAS number to be used to identify the substance is the one of Montmorillonite.

It has to be emphasized that Bentonite (EINECS: 215-108-5, CAS: 1302-78-9) and Montmorillonite (EINECS: 215-288-5, CAS: 1318-93-0) are not regarded as the same substance.

In conclusion, a mineral is generally named according to its inorganic constituent(s) in combination. They can be regarded as mono-constituent or multi-constituent substances (general guidance in Chapter 4.2.1 and 4.2.2). Some minerals cannot be described uniquely by their chemical composition, but require additional physical characterisation or processing parameters to identify them sufficiently (see Chapter 4.2.3). Some examples are given in the following table.

Mineral examples

| Name | CAS | EINECS | Additional description[29] |
|---|---|---|---|
| Cristobalite | 14464-46-1 | 238-455-4 | $O_2Si$ (crystal structure: cubic symmetry) |
| Quartz | 14808-60-7 | 238-878-4 | $O_2Si$ (crystal structure: rhombohedra symmetry) |
| Kieselguhr | 61790-53-2 | - | Also known as Diatomite, Kieselgur and Celite<br>Description:<br>A soft siliceous solid composed of skeletons of small prehistoric aquatic plants. Contains primarily silica. |
| Dolomite | 16389-88-1 | 240-440-2 | $CH_2O_3.1/2Ca.1/2Mg$ |
| Feldspar-group minerals | 68476-25-5 | 270-666-7 | An inorganic substance that is the reaction product of high temperature calcination in which aluminium oxide, barium oxide, calcium oxide, magnesium oxide, silicon oxide, and strontium oxide in varying amounts are homogeneously and ionically interdiffused to form a crystalline matrix. |
| Talc | 14807-96-6 | 238-877-9 | $Mg_3H_2(SiO_3)_4$ |
| Vermiculite | 1318-00-9 | - | $(Mg_{0.33}[Mg_{2-3}(Al_{0-1}Fe_{0-1})_{0-1}](Si_{2.33-3.33}\,Al_{0.67-1.67})(OH)_2O_{10}.4H_2O)$ |

### Analytical information required for minerals

| | |
|---|---|
| **Elemental composition** | The chemical composition gives an overall overview of the composition of the mineral regardless of the numbers of constituents and its proportions in the mineral. By convention the chemical composition is expressed for oxides. |
| **Spectral data (XRD or equivalent)** | XRD or other techniques can identify minerals based on their crystallographic structure.<br>The characteristic XRD or IR peaks identifying the mineral should be given together with a short description of the analytical method or bibliographical reference. |
| **Typical physical-chemical properties** | Minerals have characteristic physical-chemical properties which enable the completion of their identification, e.g.<br>- Very low hardness<br>- Swelling capacity<br>- Shapes of diatomite (optical microscope)<br>- Very high density<br>- Surface area (nitrogen adsorption) |

## 7.6 ESSENTIAL OIL OF LAVANDIN GROSSO

Essential oils are substances which are obtained from plants. Therefore, essential oils can also be characterised as botanically-derived substances.

In general, botanically-derived substances are complex natural substances obtained by processing a plant or its parts by a treatment such as extraction, distillation, pressing, fractionation, purification, concentration or fermentation. The composition of these substances

[29] *Definition as given in Commission Directive 2001/30/EC (OJ L 146, 31.05.2001, p.1)*

varies depending on the genus, species, growing conditions and harvest period of the sources, and the process techniques applied.

Essential oils could be defined by their main constituents as it is practice for multi-constituent substances. However, essential oils can consist of up to several hundreds of constituents, which can vary considerably depending on many factors (e.g. genus, species, growing conditions, harvest period, processes used). Therefore, a description of the main constituents is often not sufficient to describe these UVCB substances. The essential oils should be described by the plant source and the treatment process as described in Chapter 4.3.1 (using UVCB sub-type 3).

In many cases industrial standards are available for essential oils (for many essential oils also ISO-Standards). Information on standards can be given in addition. However, the substance identification should be based on the substance as manufactured.

The example below describes the “essential oil of Lavandin grosso”, for which an ISO-Standard is available (ISO 8902-1999).

### 1. Names and other identifiers

#### Source

| | |
|---|---|
| **Species** | *Lavendula hybrida grosso* (Lamiaceae) |

#### Process

| **Description of (bio)chemical reaction processes used for the manufacture of the substance:** |
|---|
| Water steam distillation of the flowering tops of *Lavendula hybrida grosso* (Lamiaceae) and subsequent separation of the water from the Essential Oil;<br>The subsequent separation is a spontaneous, physical process, which normally takes place in a separator (a so-called "florentine flask") enabling an easy isolation of the separated oil. The temperature at this stage of the distillation process is about 40 °C. |

#### Name

| | |
|---|---|
| **IUPAC name or other international chemical name** | Essential oil of *Lavendula hybrida grosso* (Lamiaceae) |
| **EC number**<br>**EC name**<br>**EC description** | 297-385-2<br>Lavender, *Lavandula hybrida grosso*, ext.<br>Extractives and their physically modified derivatives such as tinctures, concretes, absolutes, essential oils, oleoresins, terpenes, terpene-free fractions, distillates, residues, etc., obtained from Lavandula hybrida grosso, Labiatae[30]. |
| **CAS number**<br>**CAS name** | 93455-97-1<br>Lavender, *Lavandula hybrida grosso*, ext. |

### 2. Composition information – known constituents

| **Known constituents** | | | | | |
|---|---|---|---|---|---|
| | **Chemical name**<br>**EC**<br>**CAS** | **Number**<br>**EC**<br>**CAS** | **Mol. Formula Hill method** | **Typical conc. % (w/w)** | **Conc. range % (w/w)** |

30 “Labiatae” and “Lamiaceae” are synonyms

| | | | | | |
|---|---|---|---|---|---|
| | **IUPAC**<br>**other** | | | | |
| **A** | **EC**<br>linalyl acetate<br>**CAS**<br>1,6-Octadien-3-ol, 3,7-dimethyl-, acetate<br>**IUPAC**<br>3,7-Dimethyl octa-1,6-dien-3-yl acetate | **EC**<br>204-116-4<br>**CAS**<br>115-95-7 | $C_{12}H_{20}O_2$ | 33 | 28 – 38 |
| **B** | **EC**<br>linalool<br>**CAS**<br>1,6-octadien-3-ol, 3,7-dimethyl-<br>**IUPAC**<br>3,7-Dimethyl octa-1,6-diene-3-ol | **EC**<br>201-134-4<br>**CAS**<br>78-70-6 | $C_{10}H_{18}O$ | 29,5 | 24 – 35 |
| **C** | **EC**<br>Bornan-2-one<br>**CAS**<br>Bicyclo[2.2.1] heptan-2-one, 1,7,7-trimethyl-<br>**IUPAC**<br>1,7,7-Trimethylbicyclo[2.2.1]-2-heptanone<br>**Other**<br>camphor | **EC**<br>200-945-0<br>**CAS**<br>76-22-2 | $C_{10}H_{16}O$ | 7 | 6 – 8 |
| **D** | **EC**<br>Cineole<br>**CAS**<br>2-oxabicyclo [2.2.2]octane, 1,3,3-trimethyl-<br>**IUPAC**<br>1,3,3-Trimethyl-2-oxabicyclo[2.2.2]octane<br>**Other**<br>1,8-cineole | **EC**<br>207-431-5<br>**CAS**<br>470-82-6 | $C_{10}H_{18}O$ | 5,5 | 4 – 7 |
| **E** | **EC**<br>P-menth-1-en-4-ol<br>**CAS**<br>3-Cyclohexen-1-ol, 4-methyl-1-(1-methylethyl)-<br>**IUPAC**<br>1-(1-Methylethyl)-4-methyl-3-cyclohexen-1-ol<br>**Other**<br>terpinene-4-ol | **EC**<br>209-235-5<br>**CAS**<br>562-74-3 | $C_{10}H_{18}O$ | 3,25 | 1,5 – 5 |
| **F** | **EC**<br>2-Isopropenyl-5-methylhex-4-enyl acetate<br>**CAS** | **EC**<br>247-327-7<br>**CAS** | $C_{12}H_{20}O_2$ | 2,25 | 1,5 – 3 |

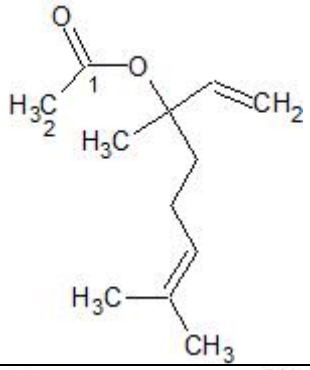
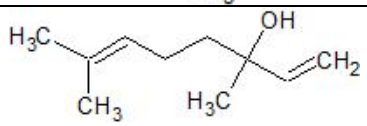
| | | | | | |
|---|---|---|---|---|---|
| | 4-Hexen-1-ol, 5-methyl-2-(1-methylethenyl)-, acetate<br>**IUPAC**<br>2-(1-Methylethenyl)-5-methylhex-4-en-1-ol<br>**Other**<br>(±)-Lavandulol acetate | 25905-14-0 | | | |
| **G** | **EC**<br>DL-borneol<br>**CAS**<br>Bicyclo[2.2.1]heptan-2-ol, 1,7,7-trimethyl-, (1R,2S,4R)-rel-<br>**IUPAC**<br>(1R,2S,4R)-rel-1,7,7-trimethyl bicyclo[2.2.1]heptan-2-ol<br>**Other**<br>borneol | **EC**<br>208-080-0<br>**CAS**<br>507-70-0 | $C_{10}H_{18}O$ | 2,25 | 1,5 – 3 |
| **H** | **EC**<br>Caryophyllene<br>**CAS**<br>Bicyclo[7.2.0]undec-4-ene, 4,11,11-trimethyl-8-methylene-, (1R,4E,9S)-<br>**IUPAC**<br>(1R,4E,9S)-4,11,11-trimethyl-8-methylene bicyclo[7.2.0]undec-4-ene<br>**Other**<br>trans-beta-caryophyllene | **EC**<br>201-746-1<br>**CAS**<br>87-44-5 | $C_{15}H_{24}$ | 1,75 | 1 – 2,5 |
| **I** | **EC**<br>(E)-7,11-dimethyl-3-methylenedodeca-1,6,10-triene<br>**CAS**<br>1,6,10-Dodecatriene, 7,11-dimethyl-3-methylene-, (6E)-<br>**IUPAC**<br>(E)-7,11-Dimethyl-3-methylene-1,6,10-dodecatriene<br>**Other**<br>trans-beta-farnesene | **EC**<br>242-582-0<br>**CAS**<br>18794-84-8 | $C_{15}H_{24}$ | 1,1 | 0,2 – 2 |
| **J** | **EC**<br>(R)-p-mentha-1,8-diene<br>**CAS**<br>cyclohexen, 1-methyl-4-(1-methylethenyl)-, (4R)-<br>**IUPAC**<br>(4R)-1-Methyl-4-(1-methylethenyl)cyclohexene<br>**Other**<br>limonene | **EC**<br>227-813-5<br>**CAS**<br>5989-27-5 | $C_{10}H_{16}$ | 1 | 0,5 – 1,5 |
| **K** | **EC**<br>3,7-dimethylocta-1,3,6-triene | **EC**<br>237-641-2 | $C_{10}H_{16}$ | 1 | 0,5 – 1,5 |

| | | | | | |
|---|---|---|---|---|---|
| | **CAS**<br>1,3,6-Octatriene, 3,7-dimethyl-<br>**IUPAC**<br>3,7-Dimethylocta-1,3,6-triene<br>**Other**<br>cis-beta-ocimene | **CAS**<br>13877-91-3 | | | |

### Known constituents ≥ 10%

| **Known constituents** | | |
|---|---|---|
| | **EC name** | **EC description** |
| **A** | linalyl acetate $C_{12}H_{20}O_2$ | |
| **B** | linalool $C_{10}H_{18}O$ | |

| **Known constituents** | | |
|---|---|---|
| | **CAS name** | **Related CAS numbers** |
| **A** | linalyl acetate $C_{12}H_{20}O_2$ | 115-95-7 |
| **B** | linalool $C_{10}H_{18}O$ | 78-70-6 |

| **Known constituents** | | | |
|---|---|---|---|
| | **Molecular Formula**<br>**CAS method** | **Structural formula** | **SMILES code** |
| **A** | $C_{12}H_{20}O_2$ |  | |
| **B** | $C_{10}H_{18}O$ |  | |

| **Known constituents** | | |
|---|---|---|
| | **Molecular weight** | **Molecular weight range** |
| A | 196.2888 | / |
| B | 154.2516 | / |

## 7.7 CHRYSANTHEMUM OIL AND ISOMERS ISOLATED THEREOF

A company is producing a chrysanthemum oil which is extracted after crushing of blossoms and leaves from *Chrysanthemum cinerariafolim*, Compositae with a solvent containing a mixture of

water/ethanol (1:10). After extraction the solvent is removed and the "pure" extract is refined in further steps resulting in the final chrysanthemum oil.

In addition, two isomers are isolated from the extract as a reaction mass of:

Jasmolin I

(Cyclopropanecarboxylic acid, 2,2-dimethyl-3-(2-methyl-1-propenyl)-, (1S)-2-methyl-4-oxo-3-(2Z)-2-pentenyl-2-cyclopenten-1-yl ester, (1R,3R)-; CAS number 4466-14-2), and

Jasmolin II

(Cyclopropanecarboxylic acid, 3-[(1E)-3-methoxy-2-methyl-3-oxo-1-propenyl]-2,2-dimethyl-, (1S)-2-methyl-4-oxo-3-(2Z)-2-pentenyl-2-cyclopenten-1-ylester, (1R,3R)-; CAS number 1172-63-0

Furthermore, the company decided to also synthesize the isomeric reaction mass of Jasmolin I and II.

The company is asking the following questions:

1. How identify the chrysanthemum oil for registration purposes?

2. Is the reaction mass of the isolated isomers Jasmolin I and II covered by the registration of the oil?

3. Can the synthesized mixture of the two isomers be regarded as the same as the mixture of the isomers isolated to from the chrysanthemum oil?

**1. How to identify the chrysanthemum oil for registration purposes?**

Chrysanthemum oil is regarded as a UVCB substance which cannot be sufficiently identified by its chemical composition (for detailed guidance see Chapter 4.3). Other identification parameters, like source and process, are essential. Chrysanthemum oil is of biological nature and should be identified via the species and the part of the organism from which it is obtained, and the refinement process (extraction with solvent). However, the chemical composition and the identity of the constituents should be given as far as known.

The following information is regarded as necessary to identify the substance sufficiently:

| | | | | |
|---|---|---|---|---|
| **Name of the substance** | *Chrysanthemum cinerariafolium*, Compositae; oil obtained from crushed blossoms and leaves by extraction with water:ethanol (1:10) | | | |
| **Source** | | | | |
| Genus, specie, sub-specie | Chrysanthemum, cinerariafolium, Compositae | | | |
| Part of plant used for oil | Blossoms and leaves | | | |
| **Process** | | | | |
| Method of manufacture | Crushing followed by extraction | | | |
| Solvent used for extraction | Water:ethanol (1:10) | | | |
| **Composition information – known constituents in % (w/w)** | | | | |
| **Name of constituent** | **EC-no** | **CAS-no** | **Min %** | **Max %** |

| | | | | |
|---|---|---|---|---|
| **Pyrethrin I**:<br>2-methyl-4-oxo-3-(penta-2,4-dienyl) cyclopent-2-enyl [1R-[1α[S*(Z)],3β]]-chrysanthemate | 204-455-8 | 121-21-1 | 30 | 38 |
| **Pyrethrin II**:<br>2-methyl-4-oxo-3-(penta-2,4-dienyl) cyclopent-2-enyl [1R-[1α[S*(Z)],3β]]-3-(3-methoxy-2-methyl-3-oxoprop-1-enyl)-2,2-dimethylcyclopropanecarboxylate | 204-462-6 | 121-29-9 | 27 | 35 |
| **Cinerin I**:<br>3-(but-2-enyl)-2-methyl-4-oxocyclopent-2-enyl 2,2-dimethyl-3-(2-methylprop-1-enyl)cyclopropanecarboxylate | 246-948-0 | 25402-06-6 | 5 | 10 |
| **Cinerin II**:<br>3-(but-2-enyl)-2-methyl-4-oxocyclopent-2-enyl 2,2-dimethyl-3-(3-methoxy-2-methyl-3-oxoprop-1-enyl)cyclopropane carboxylate | 204-454-2 | 121-20-0 | 8 | 15 |
| **Jasmolin I**:<br>2-methyl-4-oxo-3-(pent-2-enyl)cyclopent -2-enyl [1R-[1α [S*(Z)],3β]]-2,2-di methyl-3-(2-methylprop-1-enyl)cyclo propanecarboxylate | none | 4466-14-2 | 4 | 10 |
| **Jasmolin II**:<br>2-methyl-4-oxo-3-(pent-2-enyl)cyclo pent-2-en-1-yl [1R-[1α [S*(Z)],3β (E)]]-<br>2,2-dimethyl-3-(3-methoxy-2-methyl-3-oxoprop-1-enyl)cyclopropanecarboxylate | none | 1172-63-0 | 4 | 10 |
| Furthermore the substance contains up to 40 constituents below 1%. | | | | |

One can also consider identifying the substance as a well defined multi-constituent substance with six main constituents (Reaction mass of Pyrethrin I, Pyrethrin II, Cinerin I, Cinerin II, Jasmolin I and Jasmolin II).

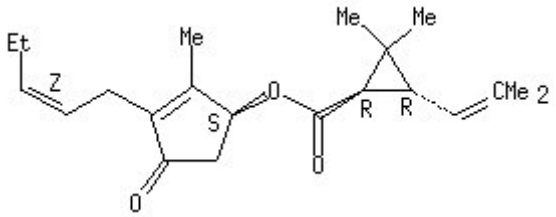
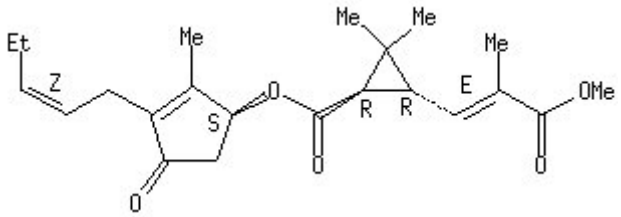
The substance would be regarded as a “substance occurring in the nature” if the manufacturing process would be only “crushing” and would be exempted from the obligation to register unless meeting the criteria for classification as dangerous according to Directive 67/548/EEC.

**2. Is the reaction mass of the isolated isomers Jasmolin I and II covered by the registration of the oil?**

The reaction mass of the isolated isomers Jasmolin I and II is not covered by the registration of the “*Chrysanthemum cinerariafolium*, Compositae oil” as single constituent(s) are not covered by the whole UVCB-substance and vice versa. The reaction mass of Jasmolin I and II is regarded as a different substance.

The reaction mass of Jasmolin I and Jasmolin II can be considered as a multi-constituent substance (detailed guidance see Chapter 4.2.3) with two main constituents.

The following information is regarded as necessary to identify the substance sufficiently:

| | | | | |
|---|---|---|---|---|
| **IUPAC name of the substance** | **Reaction mass of**<br>(2-methyl-4-oxo-3-(pent-2-enyl)cyclopent -2-enyl [1R-[1α [S*(Z)],3β]]-2,2-di methyl-3-(2-methylprop-1-enyl)cyclo propanecarboxylate)<br>**and**<br>(2-methyl-4-oxo-3-(pent-2-enyl)cyclopent-2-en-1-yl [1R-[1α [S*(Z)],3β (E)]]-2,2-dimethyl-3-(3-methoxy-2-methyl-3-oxoprop-1-enyl)cyclopropanecarboxylate) | | | |
| **Other name** | Reaction mass of Jasmolin I and Jasmolin II | | | |
| **Purity of the substance** | 95 – 98% (w/w) | | | |
| **Composition information – main constituents in % (w/w)** | | | | |
| **Name of constituent** | **EC-no** | **CAS-no** | **Min %** | **Max %** |
| **Jasmolin I**:<br>2-methyl-4-oxo-3-(pent-2-enyl)cyclopent -2-enyl [1R-[1α [S*(Z)],3β]]-2,2-di methyl-3-(2-methylprop-1-enyl)cyclo propanecarboxylate | none | 4466-14-2 | 40 | 60 |
| **Molecular formula**<br><br>**Structural formula**<br>**Molecular weight** | |  <br>$C_{22}H_{30}O_5$<br>M = 374 g/mol | | |
| **Jasmolin II**:<br>2-methyl-4-oxo-3-(pent-2-enyl)cyclo pent-2-en-1-yl [1R-[1α [S*(Z)],3β (E)]]-<br>2,2-dimethyl-3-(3-methoxy-2-methyl-3-oxoprop-1-enyl)cyclopropanecarboxylate | none | 1172-63-0 | 35 | 65 |
| **Molecular formula**<br><br>**Structural formula**<br>**Molecular weight** | |  <br>C21H30O3<br>M = 330 g/mol | | |

**3. Can the synthesized mixture (reaction mass) of the two isomers be regarded as the same as the mixture of the isomers isolated from the chrysanthemum oil?**

For chemically well defined substances, which are sufficiently described by their constituents, it is not relevant whether the substance is isolated from an extract or synthesized by a chemical process. Therefore, the synthesised reaction mass of Jasmolin I and Jasmolin II can be regarded as the same as the isomer mixture isolated from the Chrysanthemum, even if derived from different manufacturing processes, provided that the purity of the mixture and the concentration range of the main constituents are the same.

### 4. Conclusion

Two substances are identified:

1. *Chrysanthemum cinerariafolium*, Compositae; oil obtained from crushed blossoms and leaves by extraction with water:ethanol (1:10)
2. Reaction mass of the isomers Jasmolin I and Jasmolin II, independent from the manufacture process of the substance.

If the above substances would be used only in plant protection and biocidal products they would be regarded as registered under REACH (*Article 15*).

## 7.8 PHENOL, ISOPROPYLATED, PHOSPHATE

Phenol, isopropylated, phosphate (3:1) is a UVCB where the variability of the isopropylated entity cannot be fully defined.

### 1. Name and other identifiers

| | |
|---|---|
| **IUPAC name or other international chemical name** | Phenol, isopropylated, phosphate (3:1) |
| **Other names** | Phenol, isopropylated, phosphate<br>Phenol, isopropylated, phosphate (3:1) (based on a 1:1 mol ratio propylene to phenol) |
| **EC number**<br>**EC name**<br>**EC description** | 273-066-3<br>Phenol, isopropylated, phosphate (3:1)<br>/ |
| **CAS number**<br>**CAS name** | 68937-41-7<br>Phenol, isopropylated, phosphate (3:1) |

### 2. Composition information – main constituents

| **Main constituents** | | | | | |
|---|---|---|---|---|---|
| **IUPAC name** | **CAS number** | **EC number** | **Mol. formula**<br>**Hill method** | **Typical conc. (%w/w)** | **Conc. range (%w/w)** |
| Phenol, isopropylated, phosphate (3:1) | 68937-41-7 | 273-066-3 | Unspecified | | |

| **Main constituents** | |
|---|---|
| **EC name** | **EC description** |
| Phenol, isopropylated, phosphate (3:1) | / |
| **CAS name** | **CAS number** |
| Phenol, isopropylated, phosphate (3:1) | 68937-41-7 |

## 7.9 QUATERNARY AMMONIUM COMPOUNDS

A company is synthesizing the following substances:

**Substance A**

Quaternary ammonium compounds, di-$C_{10-18}$-alkyldimethyl, chlorides

EC number 294-392-2

CAS number 91721-91-4

Carbon-chain-lengths-distribution:

| | |
|---|---|
| $C_{10}$ | 10% |
| $C_{11}$ | 5.5% |
| $C_{12}$ | 12% |
| $C_{13}$ | 7.5% |
| $C_{14}$ | 18% |
| $C_{15}$ | 8% |
| $C_{16}$ | 24% |
| $C_{17}$ | 7% |
| $C_{18}$ | 8% |

**Substance B**

Quaternary ammonium compounds, dicoco alkyldimethyl, chlorides

EC number 263-087-6

CAS number 61789-77-3

The exact composition of this substance is not known by the company.

**Substance C**

Didodecyldimethylammonuim bromide

**Substance D**

Didodecyldimethylammonium chloride

**Substance E**

Substance E is manufactured as a reaction mass of Didodecyldimethylammonuim bromide and Didodecyldimethylammonium chloride (Reaction mass of substance C and D)

Substance F

Quaternary ammonium compounds, di-$C_{14-18}$-alkyldimethylammonium, chlorides

EC number 268-072-8

CAS number 68002-59-5

Carbon-chain-lengths-distribution:

| | |
|---|---|
| $C_{14}$ | 20% |
| $C_{15}$ | 10% |
| $C_{16}$ | 40% |
| $C_{17}$ | 10% |
| $C_{18}$ | 20% |

**Substance G**

Quaternary ammonium compounds, di-$C_{4-22}$-alkyldimethyl, chlorides

Carbon-chain-lengths-distribution (a single prime indicates one double bond, a double prime indicates one triple bond):

| | |
|---|---|
| C4 | 0.5% |
| C6 | 3.0% |
| C8 | 6.0% |
| C10 | 10.0% |
| C12 | 12.0% |
| C14 | 24.0% |
| C16 | 20.0% |
| C18 | 16.0% |
| C18’ | 2.0% |
| C18’’ | 0.5% |
| C20 | 4.0% |
| C22 | 2.0% |

So far, the company is using only substance B (Quaternary ammonium compounds, dicoco alkyldimethyl chlorides, EC number 263-087-6, CAS number 61789-77-3) for naming because it fits best to all substances (substance A to G). The company would like to know, whether it is possible to cover all substances (A to G) under one registration of substance B.

**1. General remarks**

Hydrocarbons (parffins, olefins) derived from fats and oils or synthetic substitutes are identified by their carbon chain distribution or by their origin (alkyl descriptor), by a functional group (functionality descriptor), e.g. ammonium, and the anion/cation (salt descriptor), e.g. chloride. The chain length distribution, e.g. $C_{8-18}$, refers to

saturated

linear (unbranched)

all carbon numbers inclusive ($C_8$, $C_9$, $C_{10}$, $C_{11}$,…., $C_{18}$) whereas a narrow distribution does not cover a broader one and vice versa

Otherwise it should be indicated in this way:

unsaturated ($C_{16\ unsaturated}$)

branched ($C_{10\ branched}$)

even-numbered ($C_{12\text{-}18\ \text{even-numbered}}$)

Carbon chains described by the source have to comprise the distribution which occurs in the source, e.g. tallow alkyl amines:

The tallow alkyl amines are 99% primary linear chain alkyl amines with the following carbon chain-length distribution (Ullmann, 1985) [a single prime indicates one double bond, a double prime indicates one triple bond]:

| | |
|---|---|
| C12 | 1% |
| C14 | 3% |
| C14’ | 1% |
| C15 | 0.5% |
| C16 | 29% |
| C16’ | 3% |
| C17 | 1% |
| C18 | 23% |
| C18’ | 37% |
| C18’’ | 1.5% |

**2.How to identify the substances for registration purposes?**

Each substance is compared to substance B (which was used for naming so far) in order to decide whether the two substances can be regarded as the same.

**Comparison of substance A and B**

The following chain lengths distribution can be found for “coco” of substance B (Ullmann, 1985) [a single prime indicates one double bond, a double prime indicates one triple bond]:

| | |
|---|---|
| C6 | 0.5% |
| C8 | 8% |
| C10 | 7% |
| C12 | 50% |
| C14 | 18% |
| C16 | 8% |
| C18 | 1.5% |
| C18’ | 6% |
| C18’’ | 1% |

Thus, the chain lengths distribution of substance A deviates from the carbon chain lengths distribution of the “coco” substance B. As the qualitative and quantitative composition of the two substances deviates significantly, they cannot be regarded as the same.

**Comparison of substance B and C**

Substance B “Quaternary ammonium compounds, dicoco alkyldimethyl, chlorides” describes a mixture of constituents with different carbon chain lengths ($C_6$ to $C_{18}$ even-numbered, linear, saturated and unsaturated), whereas substance C describes only one constituent with one defined and saturated chain length ($C_{12}$) with a different anion (bromide). Therefore, substance C cannot be regarded as the same as substance B.

**Comparison of substance B and D**

Substance B “Quaternary ammonium compounds, dicoco alkyldimethyl, chlorides” describes a mixture of constituents with different carbon chain lengths ($C_6$ to $C_{18}$ even-numbered, linear, saturated and unsaturated), whereas substance D describes one constituent with a defined and

saturated chain length ($C_{12}$) and the same anion (chloride). Substance B and D have different names and cannot be regarded as the same substance, as a single constituent is not covered by a mixture containing a certain constituent and vice versa.

**Comparison of substance B and E**

Substance E is a mixture of the substances C and D. Both have a saturated chain length of $C_{12}$ but different anions (bromide and chloride). Substance B "Quaternary ammonium compounds, dicoco alkyldimethyl, chlorides" describes a mixture of constituents with different carbon chain lengths ($C_6$ to $C_{18}$ even-numbered, linear, saturated and unsaturated) and chloride as anion. However, substance E is described only by the $C_{12}$ carbon chain length with bromide as additional anion. Therefore the substances B and E cannot be regarded as the same. As a consequence a separate registration for substance E is necessary.

**Comparison of substance B and F**

Substance F "Quaternary ammonium compounds, di-$C_{14-18}$-alkyldimethylammonium, chlorides" is a mixture of constituents with different carbon chain lengths ($C_{14}$ to $C_{18}$ even- and odd-numbered, linear and saturated). Substance F differs in the composition and in the range of the carbon chain distribution from substance B. Substance F has a narrow carbon chain length distribution, and in addition the $C_{15}$- and $C_{17}$-carbon chains. Therefore, the substances B and F and cannot be regarded as the same.

**Comparison of substance B and G**

The substances B and G seem to be very similar, as the carbon chain distribution is almost in the same range. However, substances G includes in addition the carbon chain lengths $C_4$, $C_{20}$ and $C_{22}$. The carbon chain lengths distribution of substance G comprises a wider range than that of substance B. Therefore, substance B and G cannot be regarded as the same.

**3. Conclusion**

Hydrocarbons (paraffins, olefins) can only be regarded as the same substance when all three descriptors (alkyl, functionality and salt) are the same.

In the given example above the descriptors are always different from each other. Therefore, the substances cannot be covered by one registration of substance B.

## 7.10 PETROLEUM SUBSTANCES

Using the guidance for specific UVCB substances in Chapter 4.3.3.2, two examples are included.

### 7.10.1 Gasoline blending stream ($C_4$-$C_{12}$)

**1. Name and other identifiers**

**Name**

| | |
|---|---|
| **IUPAC name or other international chemical name** | Naphtha (petroleum), catalytic reformed |

**Source**

| | |
|---|---|
| **Identification or description of stream source** | Crude oil |

### Process

| | |
|---|---|
| **Refinery process description** | **Catalytic reforming process** |
| **Carbon range** | C4-C12 |
| **Boiling point range or cut off** | 30˚C to 220˚C |
| **Other physical properties, e.g. viscosity** | below 7 $mm^2$ /s at 40˚C (Viscosity) |
| **EC number**<br>**CAS number**<br>**EC name/CAS name**<br>**EC description/CAS description** | 273-271-8<br>68955-35-1<br>Naphtha (petroleum), catalytic reformed<br>A complex combination of hydrocarbons produced by the distillation of products from a catalytic reforming process. It consists of hydrocarbons having carbon numbers predominantly in the range of C4 through C12 and boiling in the range of approximately 30˚C to 220˚C (90°F to 430°F). It contains a relatively large proportion of aromatic and branched chain hydrocarbons. This stream may contain 10 vol-% or more benzene. |

### 2. Composition information

| **Known constituents** | | | |
|---|---|---|---|
| **IUPAC name** | **CAS number** | **EC number** | **Conc. range (%w/w)** |
| Benzene | 71-43-2 | 200-753-7 | 1-10 |
| Toluene | 108-88-3 | 203-625-9 | 20-25 |
| Xylene | 1330-20-7 | 215-535-7 | 15-20 |

## 7.10.2 Gas oils (petroleum)

### 1. Name and other identifiers

| | |
|---|---|
| **IUPAC name or other international chemical name** | Gas oils (petroleum), heavy atmospheric |

### Source

| | |
|---|---|
| **Identification or description of stream source** | Crude oil |

### Process

| Refinery process description | Atmospheric distillation |
|---|---|
| **Carbon range** | C7 - C35 |
| **Boiling point range or cut off** | 121˚C to 510˚C |
| **Other physical properties, e.g. viscosity** | 20 $mm^2/s$ at 40˚C (Viscosity) |
| **EC number**<br>**CAS number**<br>**EC name/CAS name**<br>**EC description/CAS description** | 272-184-2<br>68783-08-4<br>Gas oils (petroleum), heavy atmospheric<br>A complex combination of hydrocarbons obtained by the distillation of crude oil. It consists of hydrocarbons having carbon numbers predominantly in the range of C7 to C35 and boiling in the range of approximately 121˚C to 510˚C (250°F to 950°F). |

**2. Chemical composition**

No information available.

## 7.11 ENZYMES

Using the guidance for specific UVCB substances in Chapter 4.3.2.3, two examples for enzyme concentrates are included: subtilisin (identified by IUBMB nomenclature + other constituents) and a-amylase (identified by IUBMB nomenclature + production organism)

### 7.11.1 Subtilisin

| | |
|---|---|
| **Enzyme protein** | Subtilisin |
| **IUBMB number** | 3.4.21.62 |
| **Names given by IUBMB**<br>(Systemic name, enzyme name, synonyms) | Subtilisin;<br>alcalase; alcalase 0.6L; alcalase 2.5L; ALK-enzyme; bacillopeptidase A; bacillopeptidase B; Bacillus subtilis alkaline proteinase bioprase; bioprase AL 15; bioprase APL 30; colistinase; (see also comments); subtilisin J; subtilisin S41; subtilisin Sendai; subtilisin GX; subtilisin E; etc. |
| **Comments given by IUBMB** | Subtilisin is a serine endopeptidase, type example of peptidase family S8. It contains no cysteine residues (although these are found in homologous enzymes). Species variants include subtilisin BPN' (also subtilisin B, subtilopeptidase B, subtilopeptidase C, Nagarse, Nagarse proteinase, subtilisin Novo, bacterial proteinase Novo) and subtilisin Carlsberg (subtilisin A, subtilopeptidase A, alcalase Novo). Formerly EC 3.4.4.16 and included in EC 3.4.21.14. Similar enzymes are produced by various *Bacillus subtilis* strains and other *Bacillus* species [1,3] |

| | |
|---|---|
| **Reaction** | Hydrolysis of proteins with broad specificity for peptide bonds, and a preference for a large uncharged residue in P1. Hydrolyses peptide amides |
| **Reaction type** | Hydrolases;<br>Acting on peptide bonds (peptidases);<br>Serine endopeptidases |
| **EC number** | 232-752-2 |
| **EC name** | Subtilisin |
| **CAS number** | 9014-01-1 |
| **CAS name** | Subtilisin |
| **Concentration of enzyme protein** | 26% |
| **Other constituents** | |
| Other proteins, peptides and amino acids | 39% |
| Carbohydrates | 11% |
| Lipids | 1% |
| Inorganic salts | 23% |
| **Additional parameters** | |
| **Substrates and products** | proteins or oligopeptides, water<br>peptides |

### 7.11.2 a-Amylase

| | |
|---|---|
| **Enzyme protein** | a-Amylase |
| **IUBMB number** | 3.2.1.1 |
| **Names given by IUBMB**<br>(Systemic name, enzyme name, synonyms) | 1,4-a-D-glucan glucanohydrolase;<br>glycogenase;<br>a-amylase;<br>alpha-amylase;<br>endoamylase;<br>Taka-amylase A |
| **Comments given by IUBMB** | Acts on starch, glycogen and related polysaccharides and oligosaccharides in a random manner; reducing groups are liberated in the α-configuration. The term 'α' relates to the initial anomeric configuration of the free sugar group released and not to the configuration of the linkage hydrolysed. |
| **Reaction** | Endohydrolysis of 1,4-a-D-glucosidic linkages in polysaccharides containing three or more 1,4-a-linked D-glucose units |
| **Reaction type** | hydrolases;<br>glycosidases;<br>glycosidases, i.e. enzymes hydrolysing O- and S-glycosyl compounds |
| **EC number** | 232-565-6 |
| **EC name** | Amylase, α- |
| **CAS number** | 9000-90-2 |
| **Related CAS numbers** | 9001-95-0, 9036-05-9, 9077-78-5, 135319-50-5, 106009-10-3, 70356-39-7, 144133-13-1<br>(all deleted) |
| **CAS name** | Amylase, α- |
| **Concentration of enzyme protein** | 37% |
| **Other constituents** | |
| Other proteins, peptides and amino acids | 30% |
| Carbohydrates | 19% |
| Inorganic salts | 14% |
| **Additional parameters** | |
| **Substrates and products** | starch; glycogen; water;<br>polysaccharide; oligosaccharide; |

# 8 DESCRIPTION OF SUBSTANCES IN IUCLID 5

This section illustrates how the different types of substances - mono-constituent, multi constituent, substances defined by their chemical composition plus other identifiers, and UVCB substances - can be described in IUCLID 5. Detailed information on how to describe different types of substances in IUCLID 5 is given in the Data Submission Manual Part 18 – "How to report the substance identity in IUCLID 5 for registration under REACH".

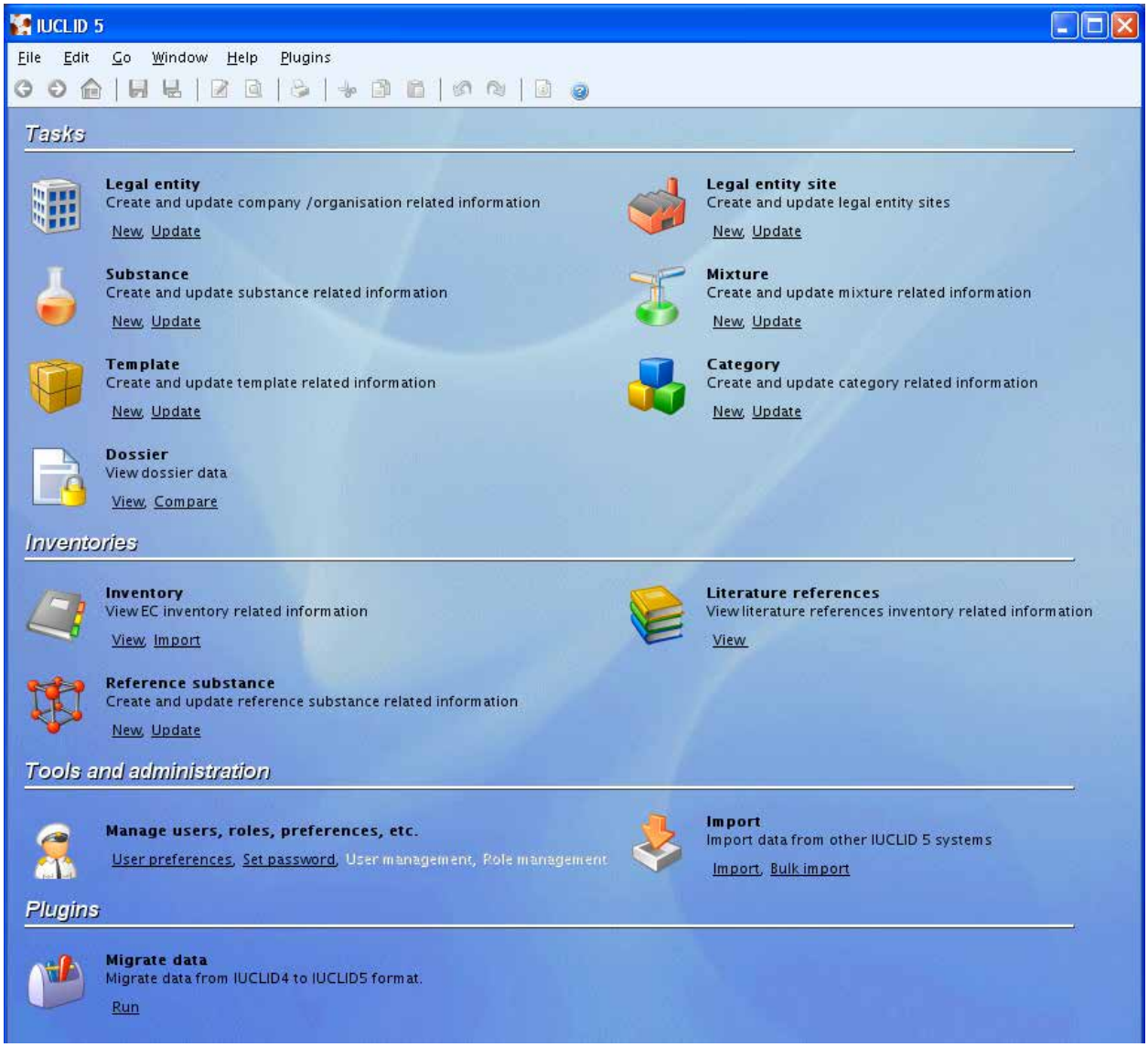
## 8.1 GENERAL PRINCIPLES

In IUCLID 5 there are three important parts related to the identification of a Substance:

the EC Inventory under "**Inventories**";

the "**Reference Substance**" inventory under "Inventories";

Section 1.1 and 1.2 of a "**Substance**" data set.



### 8.1.1 Inventories

The Inventory section contains the EC inventory (for explanations see Chapter 3.3) centrally managed and provided by the European Commission / European Chemicals Agency and the Reference Substance inventory, which is a local inventory managed and upgraded by the users on their installations as appropriate.

When the EC inventory tab is selected the user can search and display data of the inventory (i.e. EC number, CAS number, EC names, etc). This information is read-only.

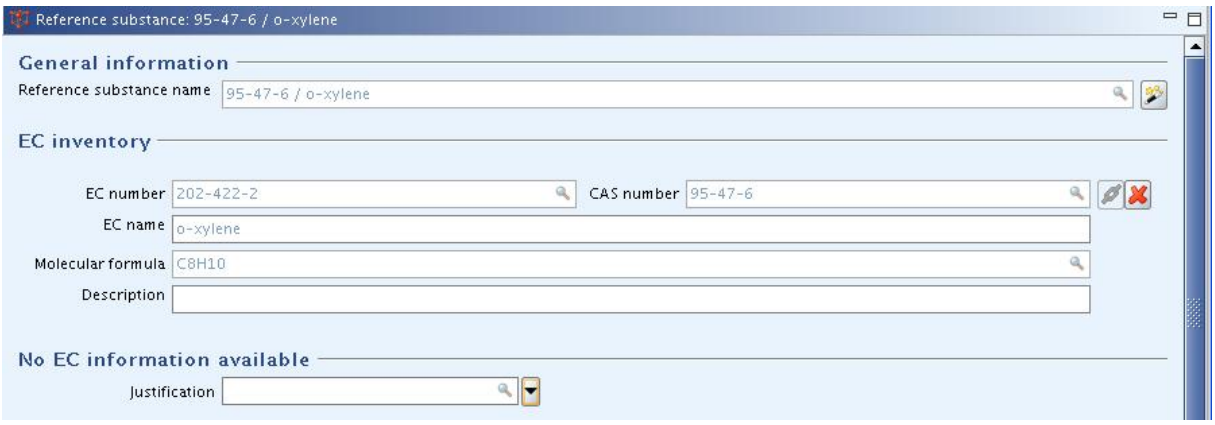
The Reference Substance tab gives the user access to his local inventory of constituents that he will use for providing the identification of his substance as manufactured, i.e. including impurities and additives.

In other words, building blocks of the substance are created and centrally maintained in the Reference Substance inventory. The Reference Substances can be re-used as appropriate for various substances.

| **Example** |
|---|
| If a substance consists out of: 91% 1,2-dimethylbenzene with 1,3-dimethylbenzene as an impurity of 5%, both the constituents 1,2-dimethylbenzene and 1,3-dimethylbenzene need to be defined in the Reference Substance inventory. The information filled in is then stored and maintained in the inventory. In case the same constituents appear in another substance at different percentages, they will already figure in the local inventory and the information can be easily re-used. |

The figures below show the Reference Substance section of IUCLID 5. They are split in separate pictures, but in IUCLID this is one screen.

**Reference substance – Part I**




The figure “Reference substance - Part I” includes:

- Reference substance name

This name can be chosen freely (in this case 95-47-6 / 1,2-dimethylbenzene).

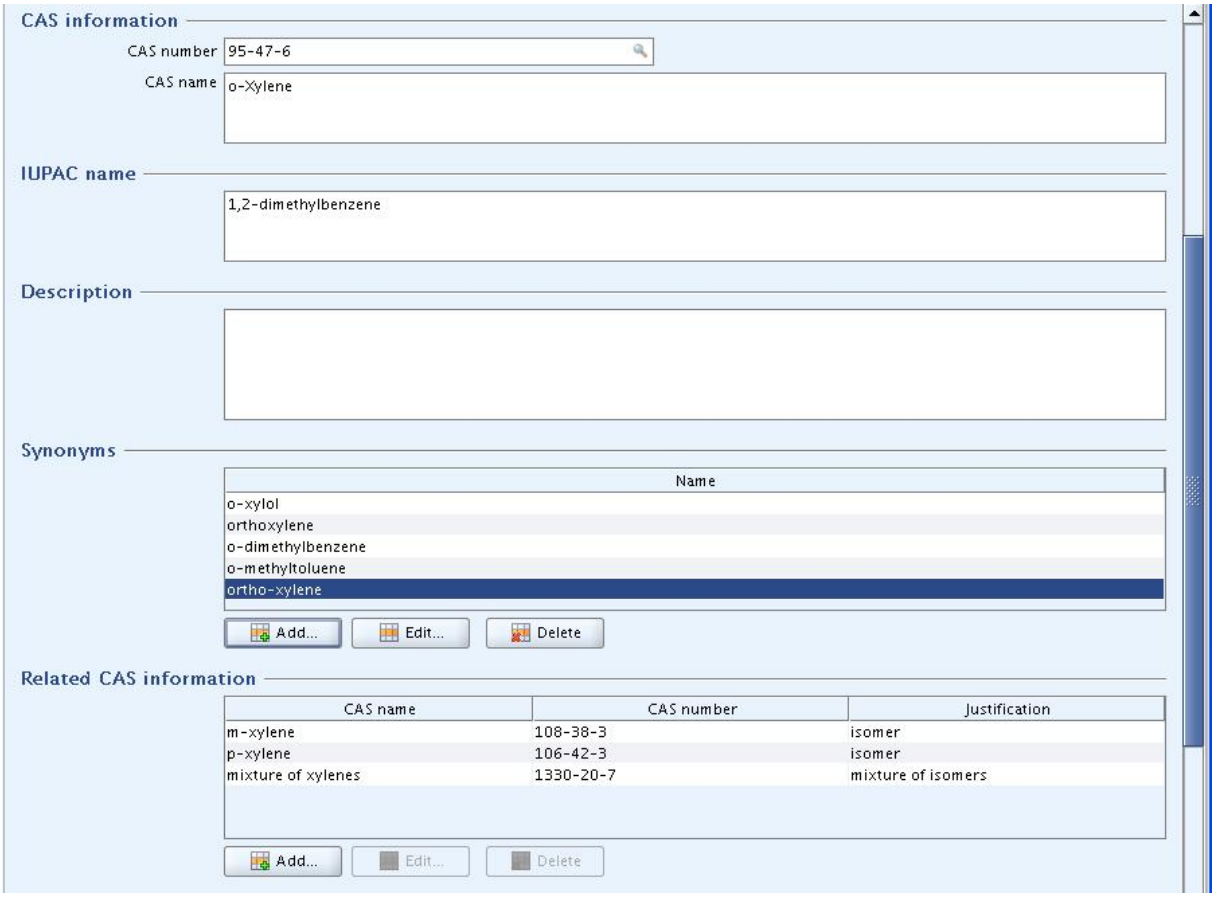
- EC inventory

The link to the read-only EC inventory, including the built-in information like the EC number.

- No EC information available

A pick up list where the reason (justification) for not having any EC inventory information can be specified (e.g. not applicable, not yet assigned).

**Reference substance – Part II**




The figure “Reference substance - Part II” includes:

- CAS information (CAS number and CAS name), including related CAS information

As a general rule, the CAS-number related to the EC-number should be given. If more than one CAS number exists (e.g. deleted CAS numbers or CAS numbers of the same substance used in different legislative systems in order to describe the substance in line with the expectation of these systems), give the other CAS number(s) as related CAS numbers;

- IUPAC name;

Note that the (chemical) name in English language of the Substances should be specified in the field “IUPAC name”. This field should also be used for UVCB substances, which are described via source and process;

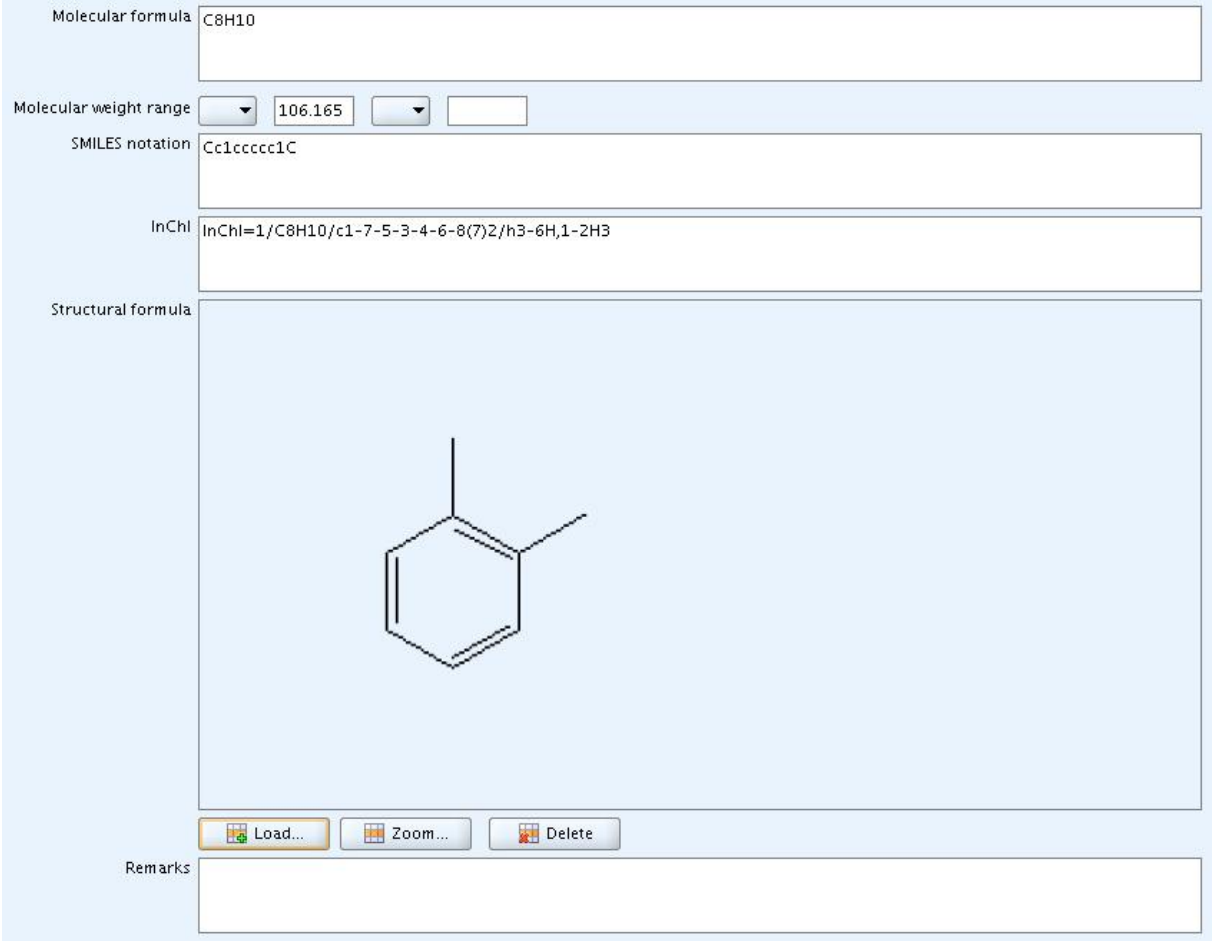
- Description field for additional information

Any additional information relevant for the description of the substance should be given in this field, e.g. for UVCB substances or minerals;

- Synonyms;

Also IUPAC names in other languages can be given here.

**Reference substance – Part III**




The figure “Reference substance - Part III” includes:

- Molecular formula;

The molecular formula shall be given in line with the Hill method.

- Molecular weight, including range;
- SMILES notation;
- InChI code;
- Structural formula as a picture.

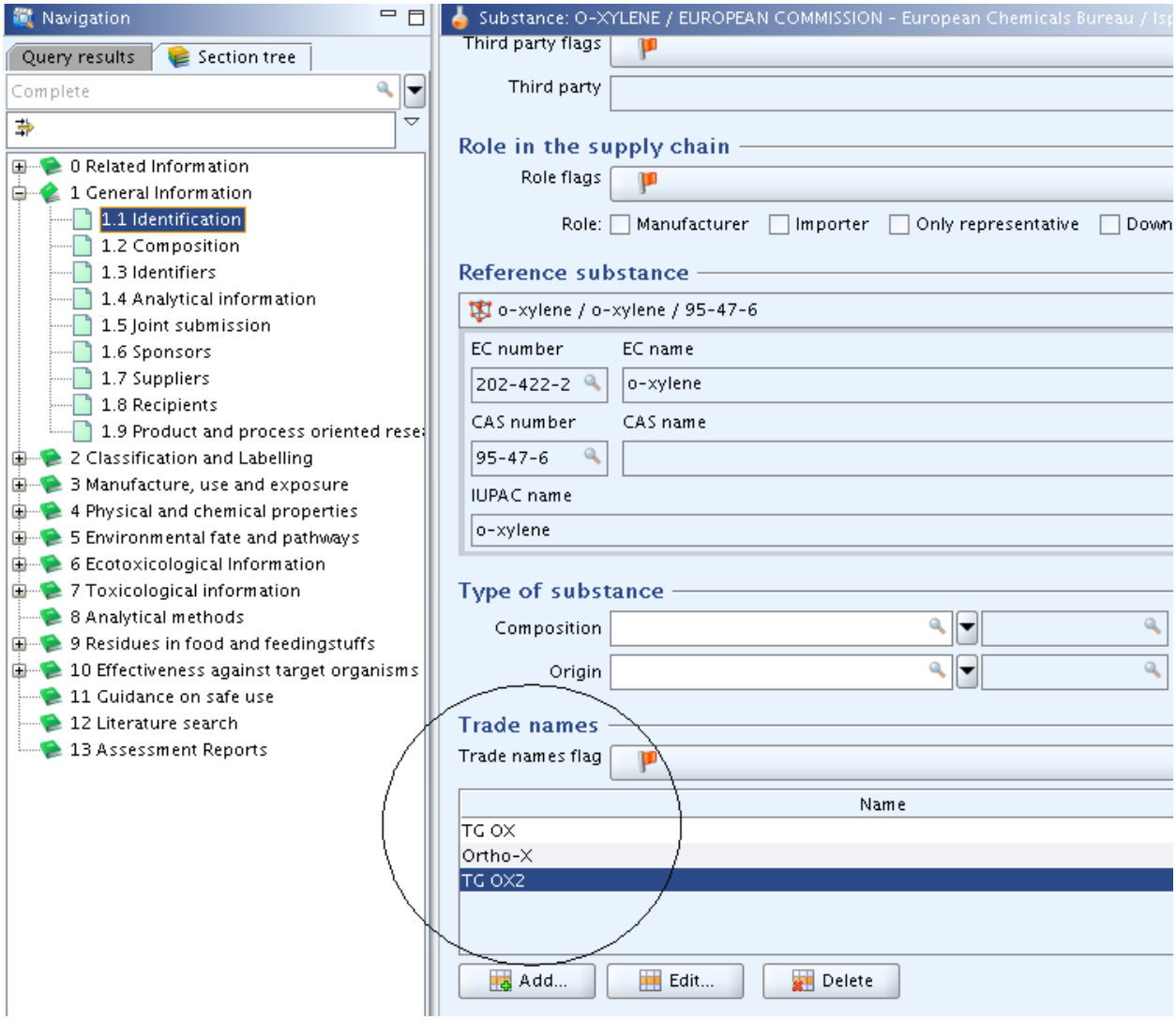
## 8.1.2 Substance data set (IUCLID Sections 1.1, 1.2, 1.3 and 1.4)

The IUCLID 5 data set contains all data for a substance like the endpoint study records, information on the classification and labelling and the chemical identity including the substance composition. Data are grouped in 11 sections.

The Substance data set can be created, searched, viewed and updated in the tab called “Substance”.

In the “substance” data set, details on the substance identification and the composition are given in Section 1.1 and 1.2.

**Substance identification – Part I**




**Section 1.1 (Substance identification)** includes

- Reference Substance

The link to the Reference Substance the substance relates to should be created here. The substance is named accordingly.

- Type of substance

From a pick list the type of substance, e.g. mono-constituent substance can be chosen.

- Trade names

All internal and external company names can be reported here

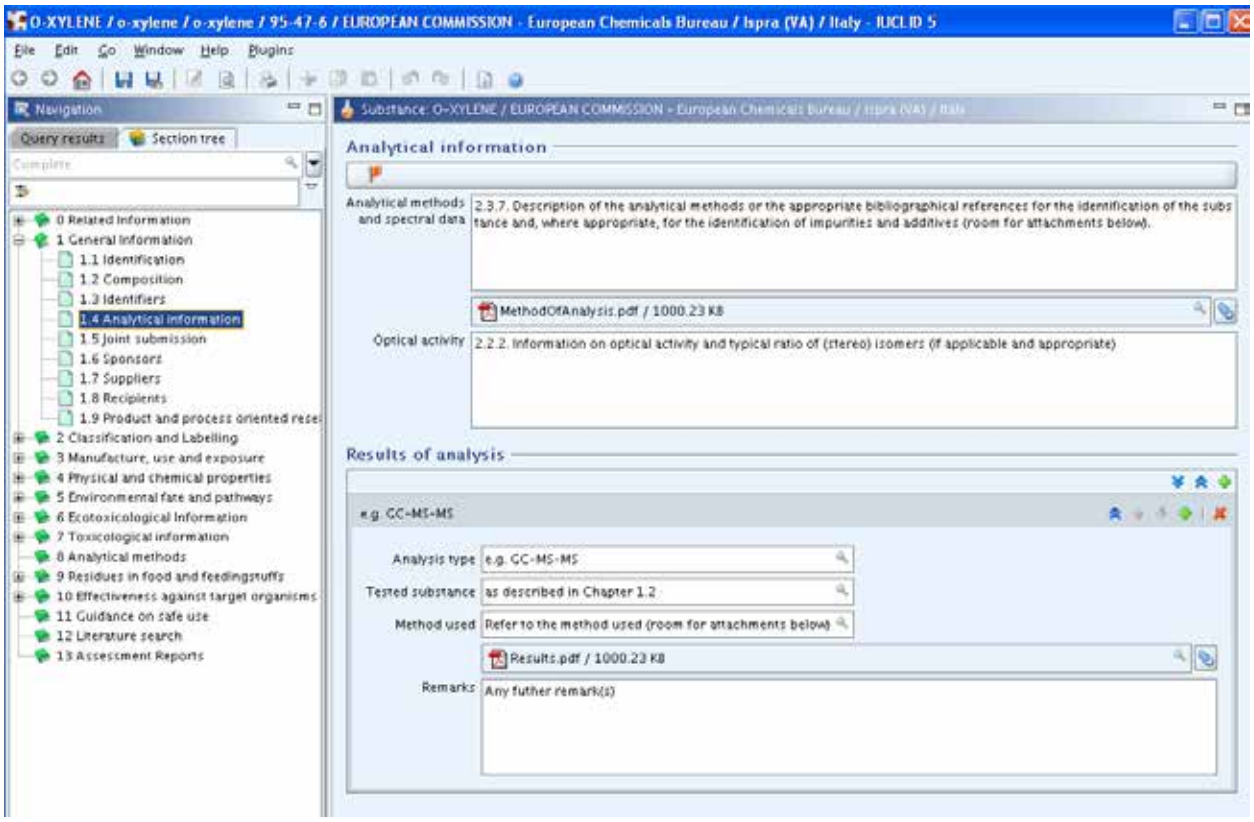
**Section 1.2 (Substance composition)** includes a description of the composition of the substance, including links to the relevant Reference Substances as building blocks. Here all constituents (e.g. main constituents, impurities) of the substances as manufactured and

additives are given. Examples including detailed guidance how to fill in Section 1.2 of IUCLID 5 are given in Chapter 8.2.

**Section 1.3 (Identifiers)** contains information to identify substances from an IT point of view, e.g. a user can specify the identifier he is using for the same substance in another IT system like a Safety Data Sheet system. This improves the data exchange between IUCLID 5 and other systems. It is not part of the identification of substances as described in this guidance document.

Section 1.3 also gives the possibility to store identification numbers which are distributed by different regulatory programmes (e.g. the REACH registration number). Also this information is not part of the identification of substances as described in this guidance document.

**Substance identification – Part II**




**Section 1.4 (Analytical information)** contains the analytical information of the substance, including information about its optical activity.

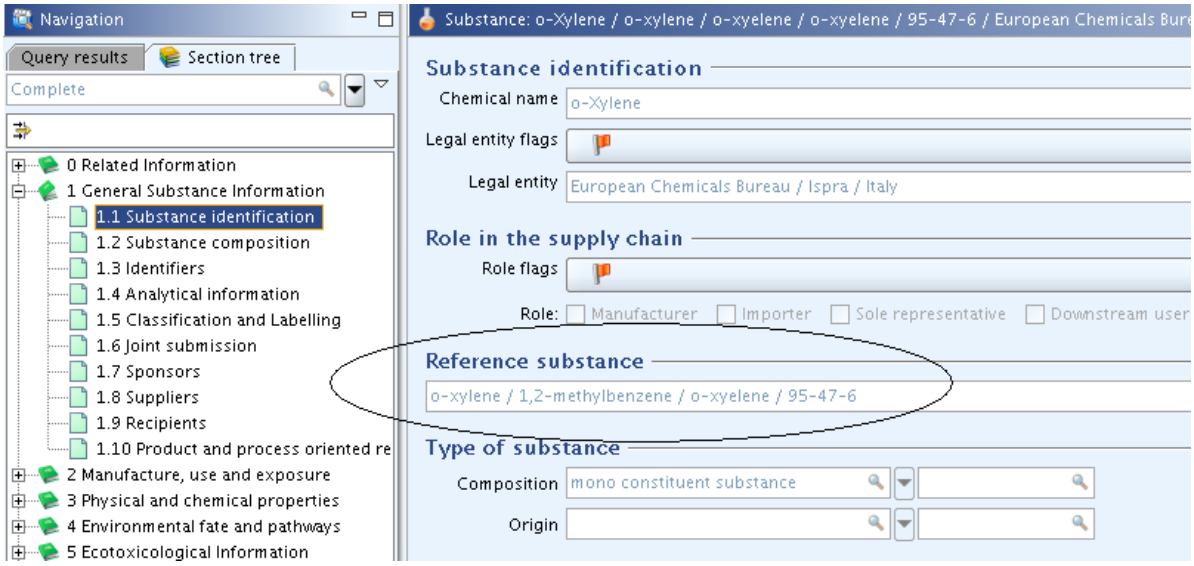
## 8.2 EXAMPLES HOW TO FILL IN IUCLID 5

An example, how to fill in IUCLID 5 for a mono-constituent substance is given in Chapter 8.2.1, an example for a multi-constituent substance in Chapter 8.2.2, an example for a substance defined by its chemical composition plus other identifiers in Chapter 8.2.3, and an example for a UVCB substance in Chapter 8.2.4.

### 8.2.1 Mono-constituent substance

| **Example:** Mono-constituent substance | | | |
|---|---|---|---|
| **Name** | 1,2-dimethylbenzene | | |
| **Main constituent** | **Typical content % (w/w)** | **Lower content % (w/w)** | **Upper content % (w/w)** |
| 1,2-dimethylbenzene | 91 | 88 | 93 |
| **Impurities** | | | |
| 1,3-dimethylbenzene | 5 | 2 | 7 |
| 1,4-dimethylbenzene | 2 | 0.5 | 3 |
| water | 2 | 0.5 | 3 |

In Section 1.1 the name of the substance is given. According to this guidance document this substance is a mono-constituent substance named as “1,2-dimethylbenzene”. In IUCLID 5, this means that the substance data set should be linked to the Reference substance 1,2-dimethylbenzene in Section 1.1.




In Section 1.2 the composition of the substance is defined:

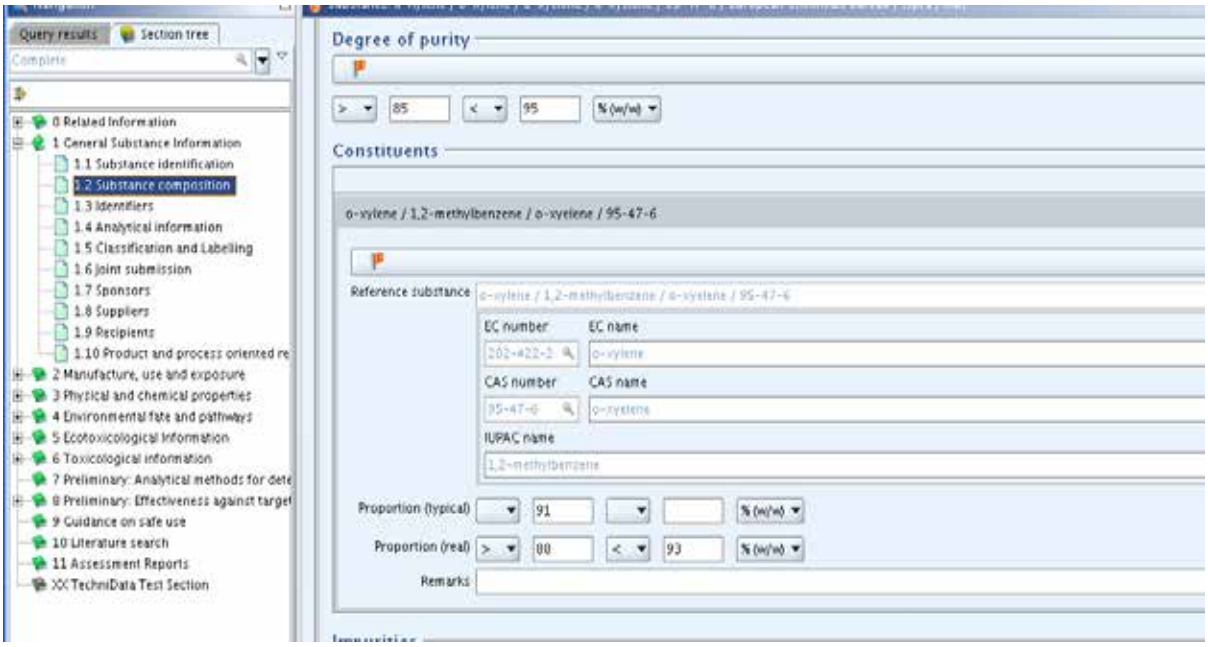
- Degree of purity

For a mono-constituent substance the degree of purity of the main constituent (normally ≥ 80%) should be given here (lower and upper limit).

- Constituents

For a mono-constituent substance the chemical identifiers (EC number and EC name, CAS number and CAS name, IUPAC name) are given here. The chemical identity is defined by the link to the Reference substance.

The field “remarks” can be used for any information. It should be used for the justification in case of deviation of the 80% rule (see Chapter 4.2.2).




- Impurities

Impurities present in a concentration ≥ 1% (or above any lower concentration limit, if relevant for the classification of the substances) should be specified by at least one of the chemical identifiers (EC number and EC name, CAS number and CAS name, IUPAC name). The chemical identity is defined by the link to the Reference substance. For each impurity the concentration (typical and range) shall be given in % (w/w).

If known, the number and total concentration of non-specified impurities shall be specified to make the total concentration complete up to 100%.

- Additives

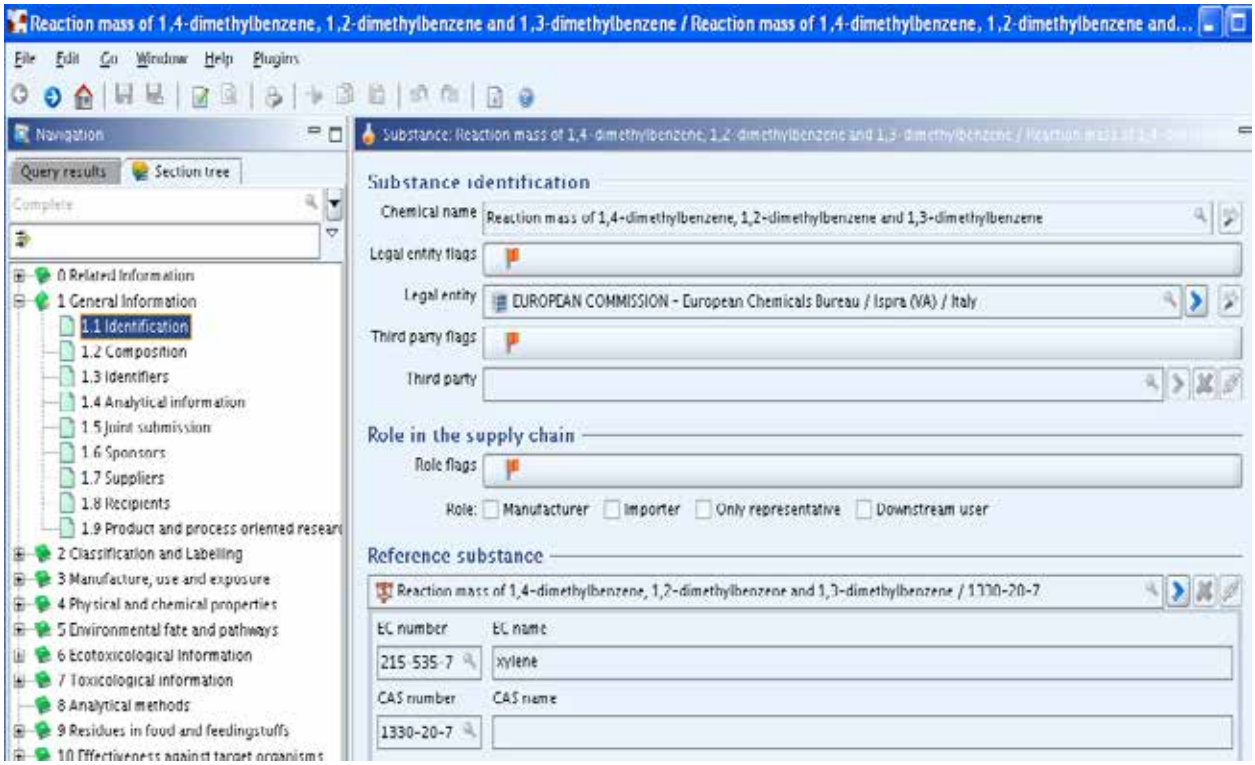
All additives (necessary to stabilise a substance) present shall be specified by the appropriate chemical identifiers (EC number and EC name, CAS number and CAS name, IUPAC name). The chemical identity is defined by the link to the Reference substance. For each additive the concentration (typical and range) shall be given in % (w/w). The stabilising function of the additive shall be specified.

### 8.2.2 Multi-constituent substance

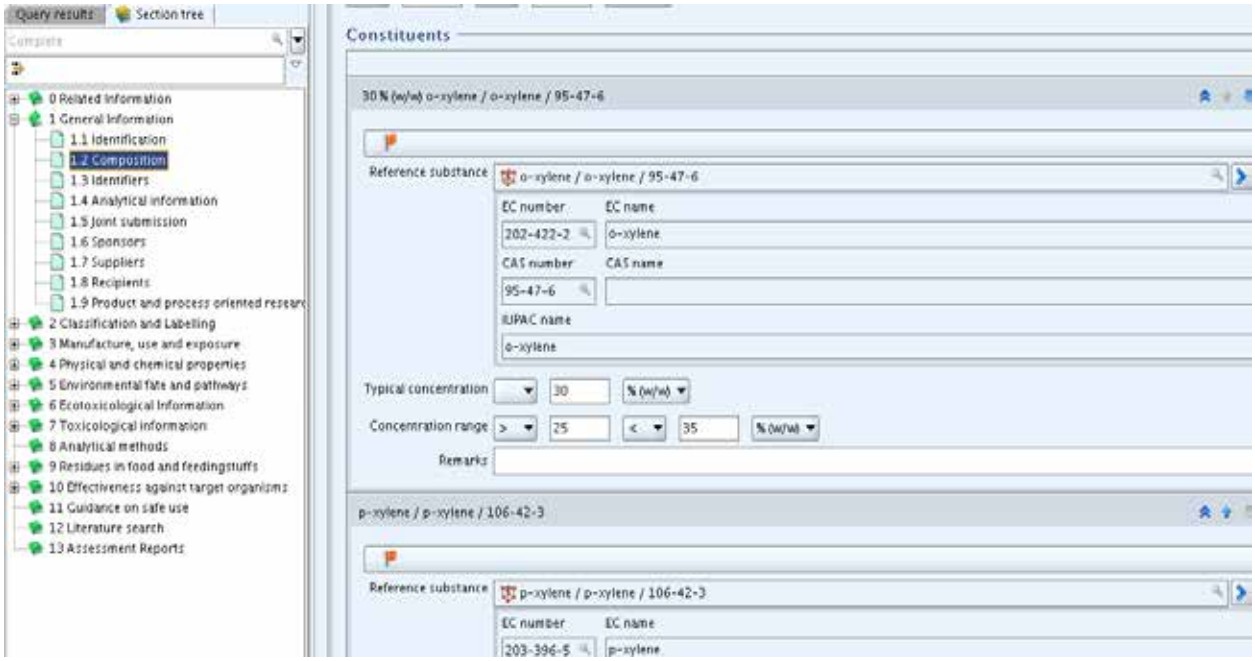
| **Example: Multi-constituent substance** | | | |
|---|---|---|---|
| **Name** | Reaction mass of 1,4-dimethylbenzene, 1,2-dimethylbenzene and 1,3-dimethylbenzene | | |
| **Main constituents** | **Typical content % (w/w)** | **Lower content % (w/w)** | **Upper content % (w/w)** |
| 1,4-dimethylbenzene | 35 | 30 | 40 |
| 1,2-dimethylbenzene | 30 | 25 | 35 |
| 1,3-dimethylbenzene | 25 | 20 | 30 |
| **Impurities** | | | |
| water | 10 | 5 | 12 |

According to this guidance document this substance is a multi-constituent substance with three main constituents and named as “Reaction mass of 1,4-dimethylbenzene, 1,2-dimethylbenzene and 1,3-dimethylbenzene”. Water is a residue solvent which cannot be separated further from the substance and should be regarded as an impurity but not as a main constituent.

In IUCLID 5, this means that the substance data set should be linked to the Reference substance “Reaction mass of 1,4-dimethylbenzene, 1,2-dimethylbenzene and 1,3-dimethylbenzene” (see Section 1.1).




For every constituent, additive and impurity, the chemical identity, the typical concentration and the concentration range are specified in Section 1.2. The degree of purity, also reported in Section 1.2, shall correspond to the overall concentration range of the main constituents. The chemical identity is defined by the link to the Reference substance.




### 8.2.3 Substance defined by its chemical composition plus other identifiers

In some cases other main identifiers are necessary in order to provide a unique identification of the substance (see Chapter 4.2.4). These additional parameters are different for each type of substance within the type. However, the additional parameter is crucial for the identification of the substance. For example, for minerals, it is important to combine the results of the elemental composition with the spectral data to identify the mineralogical composition and crystalline structure, which is then confirmed by characteristic physical and chemical properties (see also the example in Chapter 7.3).

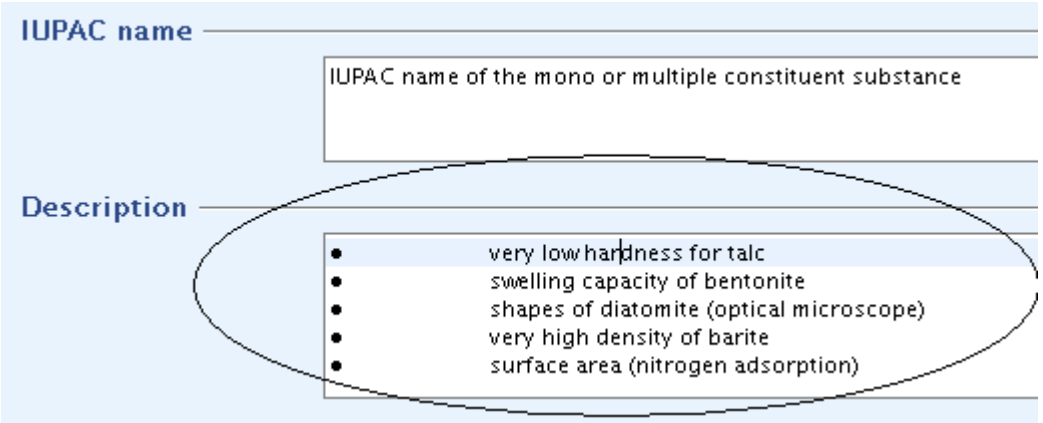
Physico-chemical properties, like:

- Crystalline structure (as revealed by X-ray diffraction)
- Shape
- Hardness
- Swelling capacity
- Density
- Surface area
- Etc.

| **Example:** Substance defined by its chemical composition plus other identifiers |
|---|
| Specific additional main identifiers can be given for specific minerals, as minerals have characteristic physico-chemical properties which enable the completion of their identification, e.g.:<br>- very low hardness for talc<br>- swelling capacity of bentonite<br>- shapes of diatomite (optical microscope) |

| | |
|---|---|
| - | very high density of barite |
| - | surface area (nitrogen adsorption) |

This kind of information should be given in the description field of the reference substance where the data set is linked to (Section 1.1 of IUCLD 5).




### 8.2.4 UVCB Substance

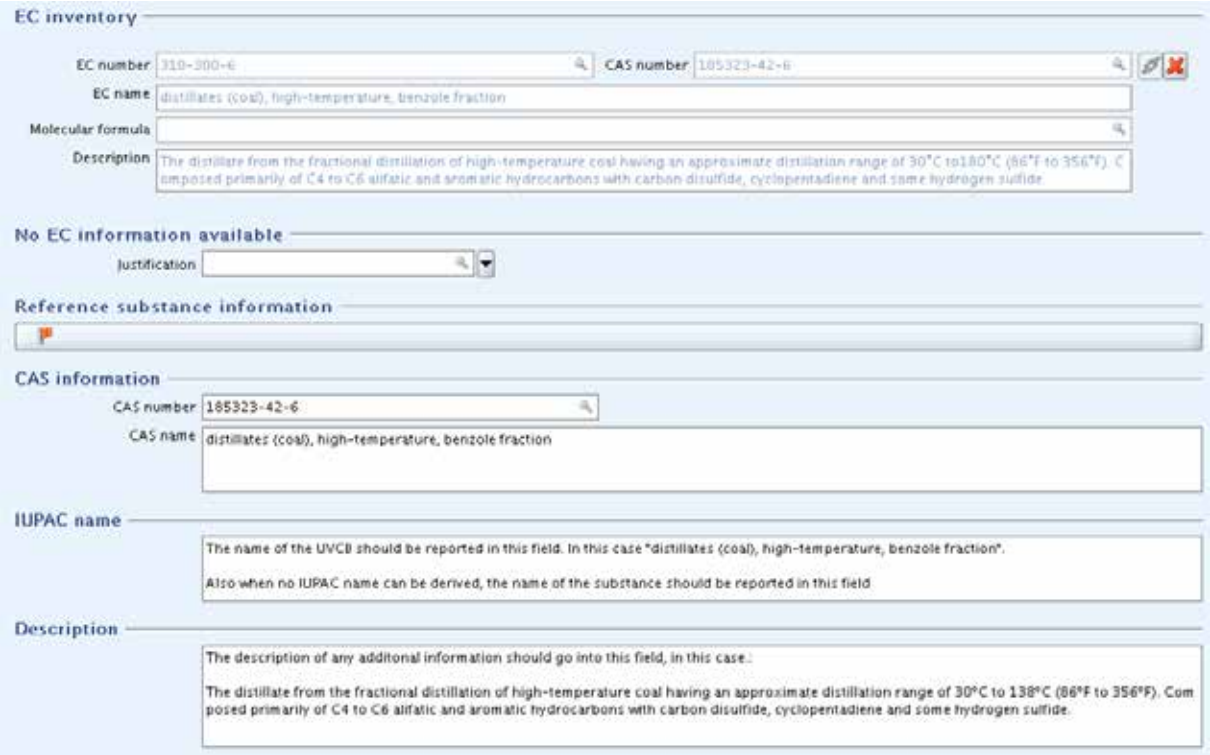
UVCB substances either cannot be uniquely specified with the IUPAC name of the constituents, as not all the constituents can be identified; or they may be generically specified but with a lack of specificity due to variability of the exact composition. The main identifiers for the UVCB substances are related to the source of the substance and the process used. Due to the lack of differentiation between constituents and impurities, the terms “main constituents” and “impurities” should not be used for UVCB substances.

However, the chemical composition and the identity of the constituents should be given as far as known. The description of the composition can often be given in a more generic way, for example “linear fatty acids C8-C16” or “alcohol ethoxylates with alcohols C10-C14 and 4-10 ethoxylate units”.

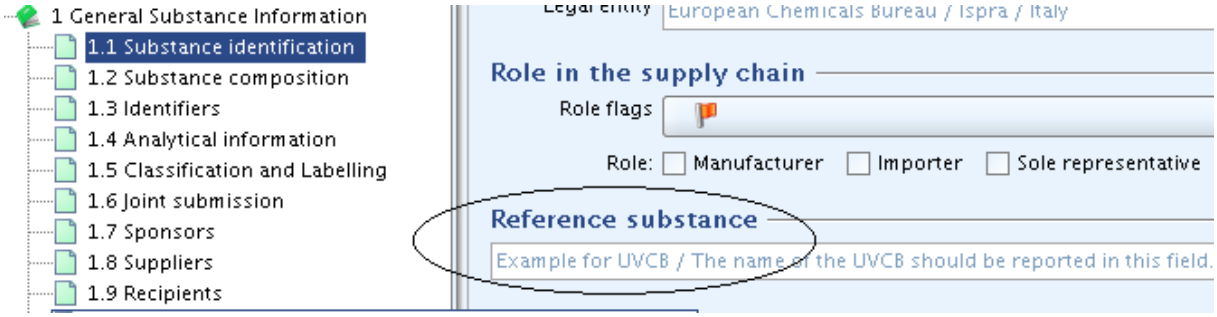
For specifying a UVCB substance the same system applies as described for mono- and multi-constituent substances. The substance itself is specified by a Reference substance as well as the known constituents.

It is important to note that when defining the substance as Reference substance, the (chemical) name of the UVCB substance should be specified in the field “IUPAC name” (although a UVCB Substance rarely has a “classical” IUPAC name). The field ‘description’ should be used for additional information (e.g. the reaction conditions, the starting materials used).

| **Example:** UVCB substance | |
|---|---|
| **Name** | distillates (coal), high-temperature, benzole fraction |
| **Description** | The distillate from the fractional distillation of high temperature coal having an approximate distillation range of 30°C to 180°C (86°F to 356°F). Composed primary of C4 to C6 aliphatic and aromatic hydrocarbons with carbon disulfide, cyclopentadiene and some hydrogen sulphide. |




For the substance data set, the same applies as described for the mono- and multi- constituent substances. The data set is linked to the Reference substance defining the substance in Section 1.1.

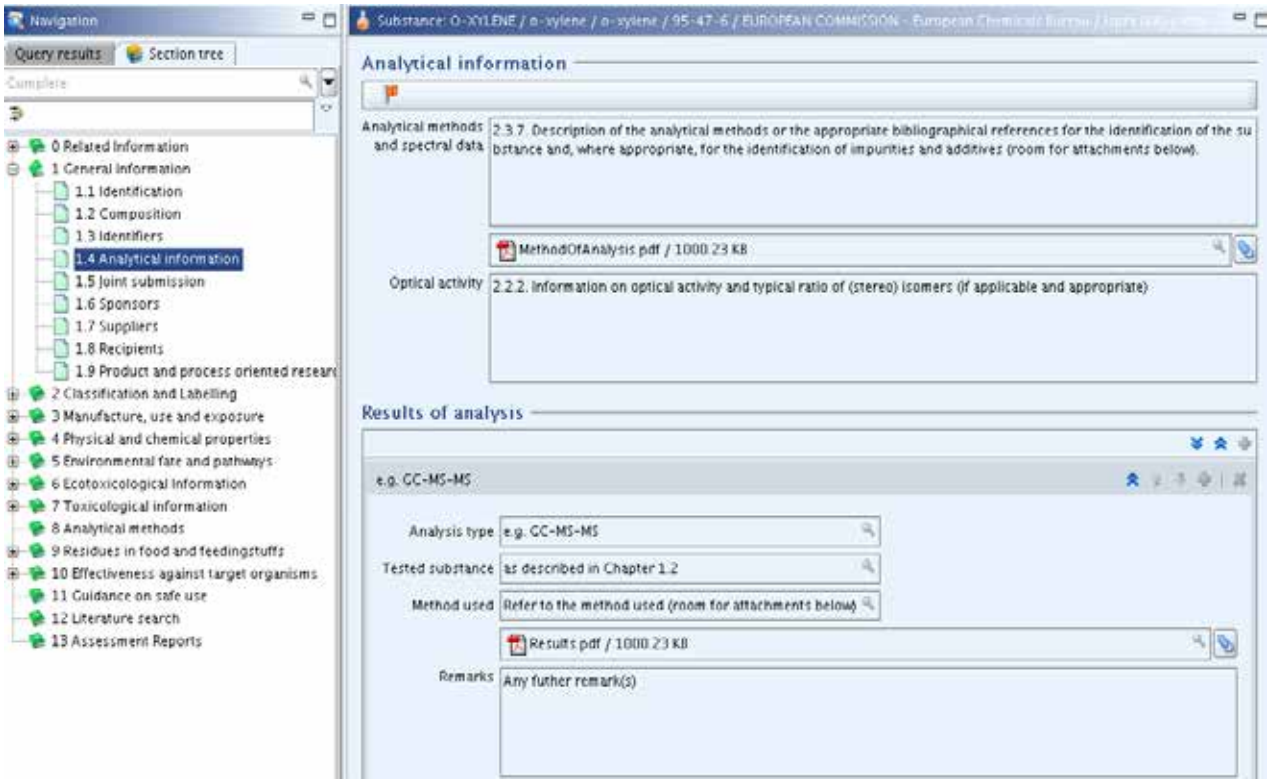



The known constituents are defined by the appropriate Reference substances as described for mono- and multi-constituent substances. No constituent should be reported under the “Impurity” header of section 1.2.

## 8.3 REPORTING OF ANALYTICAL INFORMATION

Analytical information is reported in chapter 1.4. This chapter exists out of two parts:

- Analytical information
- Results of analyis




This sub-division is directly related to requirements of REACH (*Annex VI*):

Analytical information:

- Analytical methods: in this field the description of analytical methods should be given (REACH, *Annex VI*, 2.3.7). For long text the possibility to attach documents is included.

- Optical activity: in this field information on the optical activity and typical ratio of (stereo) isomers should be given if applicable and appropriate (REACH, *Annex VI*, 2.2.2).

Results of analysis:

The result of analysis block is intended to give the user the possibility to provide information on identification related results of analysis and attach items like chromatograms. It can be used to provide spectral data (REACH, *Annex VI*, 2.3.5) or to provide chromatgraphical data (REACH, *Annex VI,* 2.3.6).

# APPENDIX I - GUIDANCE INSTRUMENTS

This Appendix includes a list of websites, databases and handbooks that can be useful for finding the appropriate IUPAC, CAS and EC names, CAS and EC numbers, molecular formulae and structure formulae, including SMILES notation, and other parameters that are required for substance identification. Commercial databases and guidance instruments have not been included.

| **General** | | |
|---|---|---|
| Substance identity parameter | **Source** | **Source description** |
| U.S. Department of Health and Human Services | http://sis.nlm.nih.gov/chemical.html | A family of databases and tools to help users to search for chemical information |
| Cambridges oft | http://chemfinder.cambridgesoft.com/ | A free database that provides chemical structures, physical properties, and hyperlinks to relevant information |
| Accelrys | http://accelrys.com/products/informatics/ | Chemical software; Accord Alphabetical Product Listing |
| Syrres | http://www.syrres.com/what-we-do/products-services.aspx | Free online searches of the following databases:<br>Environmental fate database ; KOW (online Log P) ; PHYSPROP (Physical properties) |

| **Name and other identifiers** | | |
|---|---|---|
| Substance identity parameter | **Source** | **Source description** |
| IUPAC name | http://www.iupac.org<br>or more specific:<br>http://www.iupac.org/publications/books/seriestitles/nomenclature.html#inorganic (inorganic)<br>http://www.iupac.org/publications/books/seriestitles/nomenclature.html (general) | Official website IUPAC |
| | http://www.chem.qmul.ac.uk/iupac | IUPAC chemical nomenclature and recommendations (under the authority of IUPAC) |
| | Nomenclature of Organic Chemistry (Blue Book) Pergamon, 1979 [ISBN 0-08022-3699] | Principal IUPAC nomenclature publications, update expected 2006. |
| | A Guide to IUPAC Nomenclature of Organic Compounds (recommendations 1993) (supplementary Blue Book) Blackwell Science, 1993 [ISBN 0-63203-4882] | Principal IUPAC nomenclature publications, up-date expected 2006. |

| **Name and other identifiers** | | |
|---|---|---|
| Substance identity parameter | **Source** | **Source description** |
| | Nomenclature of Inorganic Chemistry (recommendations 1990) (Red Book)<br>Blackwell Science, 1990 [ISBN 0-63202-4941] | Principal IUPAC nomenclature publications, up-date expected July 2005. |
| IUPAC name | Biochemical Nomenclature and Related Documents (White Book) Portland Press, 1992 [ISBN 1-85578-005-4] | Principal IUPAC nomenclature publications |
| | Principles of Chemical Nomenclature: a Guide to IUPAC Recommendations<br>Blackwell Science, 1998 [ISBN 0-86542-6856] | Introductory volume covering all types of compound |
| IUPAC name | http://www.acdlabs.com/products/draw_nom/ | Commercial computerised naming program that can be very helpful in naming structures of moderate complexity. Also freeware available for small molecules (IUPAC recommended) |
| | http://www.acdlabs.com/iupac/nomenclature | IUPAC nomenclature of organic chemistry (IUPAC recommenced) |
| | http://www.acdlabs.com/iupac/nomenclature/93/r93_671.htm | Complete list of approved trivial and semi-systematic root names organic compounds |
| | http://www.chemexper.com/ | The goal of the ChemExper Chemical Directory is to create a common and freely accessible database of chemicals over the internet. This database contains chemicals with their physical characteristics. Everybody can submit chemical information and retrieve information with a Web browser |
| IUBMB Nomenclature | http://www.chem.qmul.ac.uk/iubmb/ or<br>http://www.chem.qmw.ac.uk/iubmb | IUBMB biochemical nomenclature database (under authority of IUBMB) |
| Other names | http://www.colour-index.org | Colour Index Generic Names, Colour Index International, Fourth Edition Online |
| | http://online.personalcarecouncil.org/jsp/Home.jsp | INCI (International Nomenclature Cosmetic Ingredients), Official Personal Care Products Council website |
| | http://www.epa.gov/oppt/existingchemicals/pubs/tscainventory/alkyl-rg.pdf | US EPA Substances containing varying Carbon Chain Lengths (alkyl ranges using the CX-Y notation) |
| Other identifiers | http://www.cenorm.be | CE norms, official European CE-site |
| EC-number | http://esis.jrc.ec.europa.eu/ | ESIS: search on EINECS, ELINCS, NLP and *Annex I* of 67/548/EEC |
| CAS number | http://www.cas.org | Official website CAS registry service |
| | http://www.chemistry.org | Official website American Chemical Society |

| **Molecular and structural formula** | | |
|---|---|---|
| Substance identity parameter | **Source** | **Source description** |
| SMILES | http://cactus.nci.nih.gov/services/translate/ | Free SMILES generator |
| Molecular weight and SMILES | http://www.acdlabs.com/download/chemsk.html | ACDChemsketch, freeware (also commercially available) |
| Several physico-chemical parameters | http://www.epa.gov/opptintr/exposure/docs/episuite.htm | The EPI (Estimation Programs Interface) Suite™ is a Windows® based suite of physical/chemical property and environmental fate estimation models developed by the EPA's Office of Pollution Prevention Toxics and Syracuse Research Corporation (SRC). |

# APPENDIX II - TECHNICAL GUIDANCE PER SUBSTANCE IDENTIFICATION PARAMETER

The information in this appendix is intended for guidance document users who are not familiar with the technical rules for nomenclature, use of various register numbers, and notation rules for molecular and structural information, spectral data etc.

It gives some general introduction by summarising the main principles and guides the user to the original sources for complete information.

This overview is a simplified version, not complete or exhaustive, and not sufficiently detailed for the professional user. It should in no case be considered as equivalent to the official source.

## 1 Name(s) in the IUPAC- or other International nomenclature

For registration, the English IUPAC name, or another well defined internationally accepted name of the substance, shall be given.

A IUPAC name is based on the international standard chemical nomenclature set by the international organisation IUPAC, the International Union of Pure and Applied Chemistry (for suitable references see Appendix 1). The IUPAC nomenclature is a systematic way of naming chemical substances, both organic and inorganic. In IUPAC nomenclature, prefixes, suffixes and infixes are used to describe the type and position of functional groups in the substance.

**penta-1,3-dien-1-ol,** in this example:

the prefix is ***penta-1,3-***

the infix is ***-di*** and

the suffix is ***-ol***

***en-*** is the basis of the name, the root name.

The set of rules was developed over several years and is continuously changing to deal with new components of molecular diversity and possible conflicts or confusions that have been identified. The rules set by IUPAC can be used only for well defined substances.

Some general guidance is given below on the structure of an IUPAC name. For detailed support, please use the guidance provided in Chapter 4 of the guidance document text.

### 1.1 Organic substance

Step 1 Identify the number of C-atoms in the longest continuous chain of carbon atoms; This number determines the prefix, the first part, of the root name:

| Number of carbon atoms | Root |
|---|---|
| 1 | meth- |
| 2 | eth- |
| 3 | prop- |
| 4 | but- |
| 5 | pent- |
| 6 | hex- |
| 7 | hept- |
| 8 | oct- |
| N | …. |

Step 2 Determine the saturation of the chain; the saturation of the chain determines the suffix, the second part, of the root name:

| Saturation | Bonds | Suffix |
|---|---|---|
| Unsaturated | Double | -ene |
| | Triple | -yn |
| Saturated | - | -ane |

In case of multiple double or triple bonds, the number of bonds is indicated with 'mono', 'di', 'tri' etc before the suffix:

**Pentene with 2 double bonds: pentadiene**

Step 3 Combine prefix, suffix and additions to the root name

NB: For the root name, IUPAC-approved trivial and semi-systematic names may be used as well:

**Benzene, toluene, etc.**

Step 4 Use the table below:

- Identify substituents and/or functional groups: carbon or non-carbon groups attached to the chain of carbon atoms identified under 1;
- Determine the order of precedence of the substituents and/or functional groups;
- Add the suffix for the first substituent/functional group, and any subsequent ones in order of precedence;
- Add the prefix for the other substituents and functional groups in alphabetical order.

| Precedence | Group | Formula | Suffix | Prefix |
|---|---|---|---|---|
| 1 | Carboxylic acid | R-COOH | -oic acid | Carboxy |
| 2 | Ester | R-CO-O-R | -oate | - |
| 3 | Amide | R-$CONH_2$ | -amide | Carbamoyl |
| 4 | Cyanide | R-CN | -nitrile | Cyano |
| 5 | Aldehyde | R-CHO | -al | Oxo |
| 6 | Ketone | R-CO-R | -one | Oxo |
| 7 | Alcohol | R-OH | -ol | Hydroxyl |
| 8 | Thiol | R-SH | -thiol | Sulfanyl |
| 9 | Amine | R-$NH_2$ | -amine | Amino |

## 1.2 Inorganic substance

### 1.2.1 Naming of simple inorganic substances

Naming of inorganic substances is based on a set of rules (IUPAC red book, see reference in 7.1), of which the most basic are presented below:

1 Single atom anions are named with an -ide suffix:

**$O^{2-}$ is oxide**

2 Simple ionic compounds are names with the cation followed by the anion. For cations with charges >1, the charges are written using Roman numerals in parentheses immediately following the element name:

**$Cu^{2+}$ is copper(II)**

3 Hydrates are named as the ionic compound followed by a numerical prefix and -hydrate. The numerical prefixes are mono-, di-, tri-, tetra-, penta-, hexa-, hepta-, octa-, nona-, deca-:

**$CuSO_4 \cdot 5H_2O$ is "copper(II) sulphate pentahydrate"**

NB: for the purpose of registration hydrates and, where applicable, the anhydrous form, of a particular metal salt are considered to be “same substances”.

4 Inorganic molecular compounds are named with a prefix (see hydrates) before each element. The more electronegative element is written last, with an -ide suffix:

**$CO_2$ is carbon dioxide, and $CCl_4$ is carbon tetrachloride.**

5 Acids are named after the anion formed when the acid is dissolved in water. There are several possibilities:

a If, when dissolved in water, the acid dissociates into an anion with the name of “x”-ide, the acid is named hydro-“x”-ic acid:

**hydrochloric acid forms a chloride anion.**

b If, when dissolved in water, the acid dissociates into an anion with the name ”x”–ate, the acid is named “x”-ic acid:

**chloric acid dissociates to chlorate anions in water.**

c If, when dissolved in water, the acid dissociates into an anion with the name in the form of "x"–ite, the acid is named "x"-ous acid:

**chlorous acid disassociates into chlorite anions.**

### 1.2.2 Naming of mineralogical phases

Complex mineralogical phases generally contain three or more elements in combination. Most of the elements present are combined with oxygen and in order to simplify identification, the complex compounds are usually regarded by mineralogists to be built up of oxides, some of which are basic and others acidic in character. For example, in the case of silicates it has been the custom to represent them as either the sum of a number of oxides or as salts of silicic acid, or aluminosilicic acids. Accordingly, calcium orthosilicate can be represented as $2CaO.SiO_2$, a combination of separate oxides or as $Ca_2SiO_4$, as the calcium salt of orthosilicic acid $H_4SiO_4$. The same applies to other complex mineral oxides – they are named with a prefix before each oxide (e.g. $Ca_3SiO_5$ = Tricalcium silicate = $3CaO.SiO_2$). In some industrial sectors, further simplification has been introduced in order to abbreviate the compound formulae. For example, in the case of Portland cement clinker, $2CaO.SiO_2$ (calcium orthosilicate or dicalcium silicate) is shortened to $C_2S$, where C = CaO and S = $SiO_2$. Reference to standard mineralogical or industry texts is advised where complex mineralogical phases are to be named or identified.

## 1.3 Natural products and related components

For natural products IUPAC has developed several rules for systematic naming. In short it means that for substances extracted from a natural source the name is based, whenever possible, on the family, genus or species name of the organism from which the substance has been extracted:

**For a hypothetical protein, *Hypothecalia Examplare*
the names are based on hypothecalia and/or
examplare, for example Horse Examplare**

If possible, the name should reflect the known or likely distribution of the natural product. If appropriate, the class or order might also be used as the basis for the name of a substance that occurs in a number of related families. The name of natural products of unknown structure should not contain any of the prefixes, suffixes and/or infixes used in organic nomenclature:

**Condensation product of Horse examplare, Valarine
added to the N-terminus**

Many naturally occurring substances belong to well defined structural classes, each of which can be characterised by a set of parent structures that are closely related, that is, each can be derived from a fundamental structure. The systematic name for these naturally occurring substances and their derivates can be based on the name of an appropriate fundamental parent structure:

**Well known parent structures are alkaloids, steroids,
terpenoids, and vitamins**

A fundamental parent structure should reflect the basic skeleton that is common to most substances in that class. Naturally occurring substances or derivates are named after the parent structure, adding prefixes, suffixes or infixes denoting:

- modifications to the skeletal structure
- replacement of skeletal atoms
- changes in the state of hydrogenation implied by the name of the parent structure

- atoms or groups substituting hydrogen atoms of the parent structure
- configurations not already implied by the name of the parent structure, or changed from that implied

**Thiamin chloride is also known as vitamin $B_1$**

For more detailed information on systematic naming of natural products and related substances, the IUPAC should be contacted (see Appendix 1).

### 1.4 IUPAC name not possible to derive

If it is not possible to derive an IUPAC name for certain substances, other internationally recognised nomenclature, specific for those substances, can be used such as:

- Minerals and ores; mineralogical names;
- Petroleum substances
- Colour Index Generic Names [3];
- Oil additives;
- INCI (International Nomenclature Cosmetic Ingredients) [4];
- SDA (Soap and Detergent Association) names for surfactants [5];
- Etcetera.

## 2 Other names

All relevant names and/or public identifiers in all languages under which a substance is or will be marketed in the EU (e.g. trade names) are useful to include for registration under the REACH framework. This includes trade names, synonyms, abbreviations etc.

3. http://www.colour-index.org, Colour Index International, Fourth Edition Online
4. http://online.personalcarecouncil.org/jsp/Home.jsp, INCI, official Personal Care Products Council website
5. http://www.cleaning101.com, official website of SDA

## 3 EC-number from EINECS, ELINCS or NLP (EC Inventory)

The EC-number, i.e. the EINECS, ELINCS or NLP number, is the official number of the substance within the European Union. The EC-number can be obtained from the official publications of EINECS, ELINCS and NLP and of the European Chemicals Agency.

The EC-number consists of 7 digits of the type $x_1x_2x_3$-$x_4x_5x_6$-$x_7$. The first digit is defined by the list to which the substance belongs:

| List | First digit of EC-number |
|---|---|
| EINECS | 2 or 3 |
| ELINCS | 4 |
| NLP | 5 |

## 4 CAS name and CAS number

The Chemical Abstracts Service (CAS), a division of the American Chemical Society (ACS), assigns a CAS name and number to every chemical which enters the CAS registry database. The names and numbers are assigned in sequential order to unique substances identified by CAS scientists. Every substance registered at the Chemical Abstracts Service has a name according to the CAS-nomenclature, which the ACS adopts after recommendations of the ACS committee on nomenclature (see references in Appendix 1).

### 4.1 CAS name

The CAS name is the name given by the Chemical Abstract Service and is different from the IUPAC name. The CAS nomenclature is based on a limited set of criteria that are not always sufficient for deriving the name for a substance. Therefore, in general, it is recommended to contact the Chemical Abstract Service to obtain the correct CAS name.

In short, the basic nomenclature rules are:

- A ‘main’ part of the substance is selected to act as the header or parent.
- Substituents are listed after the header/parent, which is referred to as inverted order
- When more substituents are present, they are listed in alphabetical order, (including the prefixes):

**o-Xylen-3-ol is Benzene, 1,2-dimethyl, 3-hydroxy,**

### 4.2 CAS number

CAS-numbers can be obtained from the Chemical Abstract Service.

The CAS-number consists of a minimum of 5 digits, split up in three parts, separated by hyphens. The second part always consists of 2 digits, the third part of 1 digit,

$$N_i \ldots\ldots N_4\, N_3 - N_2\, N_1 - R$$

For the CAS-number checking, a “checksum” is available:

$$\frac{iN_i + \ldots\ldots 4N_4 + 3N_3 + 2N_2 + 1N_1}{10} = \frac{\sum iN_i}{10} = Q + \frac{R}{10}$$

The CAS-number must be correct according to the checksum.

## 5 Other identity codes

Other internationally recognised identity codes can be given as well, like:

- UN number;
- Colour Index Number;
- Dye number;
- Etcetera.

## 6 Molecular formula, structural formula and SMILES

### 6.1 Molecular formula

A molecular formula identifies each type of element by its chemical symbol and identifies the number of atoms of each such element found in one discrete molecule of the substance.

Molecular formulae should be given according to the (traditional) Hill system and, in addition, according to the CAS system, where this differs from the Hill system formula.

For applying the Hill method the following steps can be followed:

1. Identify the elements and list the chemical symbols;

2. Arrange the elements in the correct order:

   a. Carbon containing substances:
   Each element is mentioned by its chemical symbol, in the following sequence:

   (1) Carbon;
   (2) Hydrogen;
   (3) Other element symbols in alphabetical order:

   **Pentane: C5H12**

   **Pentene: C5H10**

   **Pentanol: C5H12O**

   b. Non carbon containing substances:
   Each element is given in alphabetical order:

   **Hydrochloric acid: ClH**

3. For each element, where the number of atoms is > 1, give the number of atoms as a subscript to the chemical symbols;

4. Add information that is not related to the main structure at the end of the molecular formula, separated by a dot or comma:

   **Sodium benzoate is C7H6O2, sodium salt**

   **Copper sulphate dihydrate is CuO4S.2H2O**

In the case that the Hill method cannot be applied for a specific substance, the molecular formula should be given in a different way, for example as an empirical formula, a simple description of the atoms and the ratio of the atoms available, or the formula given by the Chemical Abstract Service (see Chapter 4 of the guidance document text).

### 6.2 Structural formula

A structural formula is needed for the visualisation of the disposition of the molecules within the substance and their relationships to each other. The structural formula should indicate the location of the atoms, ions or groups and the nature of the bonds joining them. This includes also isomerism, i.e. cis/trans, chirality, enantiomers etc.

The structural formula can be given in different formats: in the form of a molecular formula and/or in the form of a structural diagram.

- *Structural formula in the form of a molecular formula*

  1. Write down all elements group wise and in order of appearance:

**n-pentane: CH3CH2CH2CH2CH3**

2. Each substituent is written down between brackets, directly after the atom to which it is connected:

**2-methylbutane: CH3CH(CH2)CH2CH3**

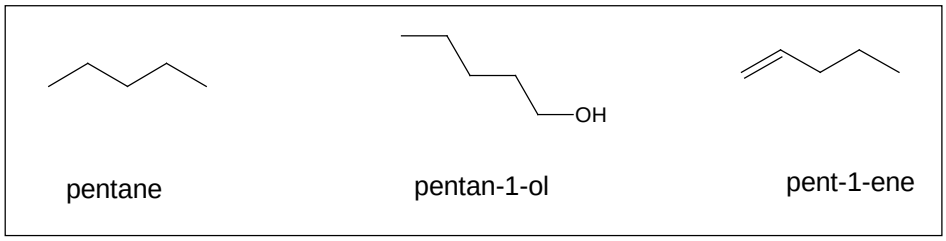
3. In case of double or triple bonds, show them between the groups of elements affected:

**pent-1-ene: CH2=CHCH2CH2CH3**

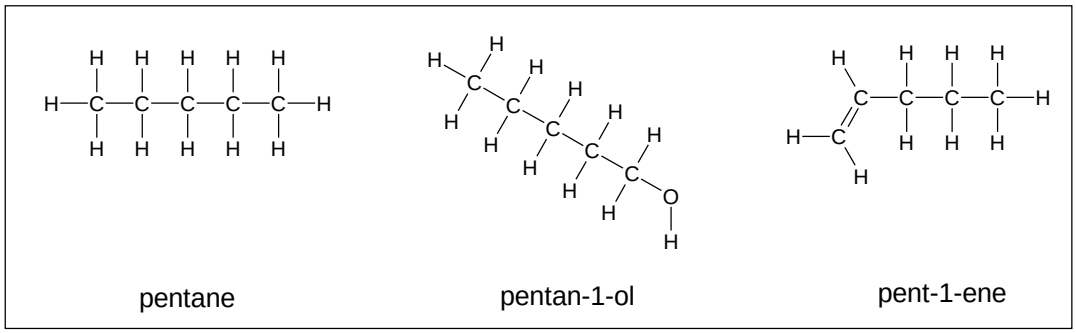
- *Structural formula in the form of a structural diagram*

For a structural diagram, the elements and the bonds between the elements are visualised in a 2D or 3D picture. Several methods exist:

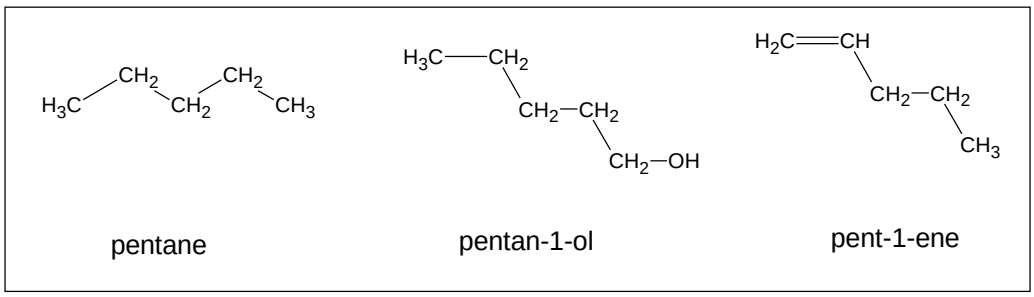
1. Showing all non-carbon elements and hydrogen attached to non-carbon elements.




2. Showing all elements by name




3. Showing carbon and hydrogen as groups (e.g. CH3), all non-carbon elements and all hydrogens not bonded to carbon.




## 6.3 SMILES notation

SMILES is the acronym for Simplified Molecular Input Line Entry Specification.[31] It is a chemical notation system used to represent a molecular structure by a linear string of symbols.

[31] Weininger (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and

With standard SMILES, the name of a molecule is synonymous with its structure: it shows indirectly a two dimensional picture of the molecular structure. Since a two dimensional chemical structure can be drawn in various ways, there are several correct SMILES notations for one molecule. The basis of SMILES is the representation of a valence model of a molecule; therefore, it is not suitable to describe molecules which cannot be represented by a valence model.

SMILES notations are comprised of atoms, designated by elemental symbols, bonds, parentheses, used to show branching, and numbers, used for cyclic structures. A SMILES notation denotes a molecular structure as a graph with optional chiral indications. A SMILES notation describing the structure only in terms of bonds and atoms is called a generic SMILES; a SMILES notation written with isotopic and chiral specifications is known as an isomeric SMILES.

In short the SMILES notation is based on several basic rules:

1. Atoms are represented by their atomic symbols;

2. Each atom, except for hydrogen, is specified independently;

   a. Elements in the “organic subset” B, C, N, O, P, S, F, Cl, Br and I are written without brackets and without attached H, as long as the number of H conforms to the lowest normal valence(s) consistent with explicit bonds:

| Element in “organic subset” | “Lowest normal valence(s)” |
|---|---|
| B | 3 |
| C | 4 |
| N | 3 and 5 |
| O | 2 |
| P | 3 and 5 |
| S | 2, 4 and 6 |
| F | 1 |
| Cl | 1 |
| Br | 1 |
| I | 1 |

   b. Elements in the “organic subset” are written with brackets as soon as the number of H does not conform to the lowest normal valence:

      **Ammonium cation is NH4+**

   c. Elements other than those in the “organic subset” are written between brackets with any attached hydrogen shown.

3. Aliphatic atoms are entered in upper case; aromatic atoms are entered in lower case:

   **benzene is c1ccccc1 and cyclohexane is C1CCCCC1**

4. Hydrogen is only included in the following situations:

   a. Charged hydrogen, i.e. a proton, [H+];

encoding rules; J. Chem. Inf. Comput. Sci.; 1988; 28(1); 31-36.

b. Hydrogens connected to other hydrogens, i.e. molecular hydrogen, [H][H];

c. Hydrogens connected to other than one other atom, e.g. bridging hydrogens;

d. Isotopic hydrogen specifications, e.g. deuterium ([2H]);

e. If the hydrogen is connected to a chiral atom.

5. The four basic bonds are shown as follows:

| Type of bond | SMILES notation |
|---|---|
| Single | - (no need to show) |
| Double | = |
| Triple | # |
| Aromatic | Lower case letters |

6. Substituents are shown by enclosure in parentheses, and immediately after the atoms to which they are connected:

**2-methylbutane is CC(C)CC**

a Substituents are always shown directly after the relevant atoms; they cannot follow a double or triple bond symbol:

**Pentanoic acid is CCCCC(=O)O**

b. Substituents within substituents are allowed:

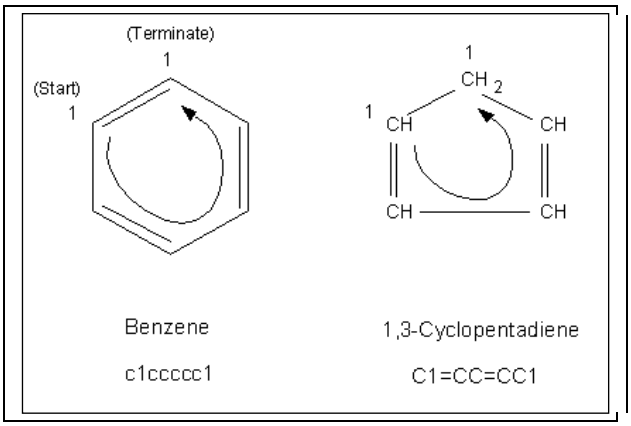
**2-(1-methylethyl)butane is CC(C(C)C)CC**

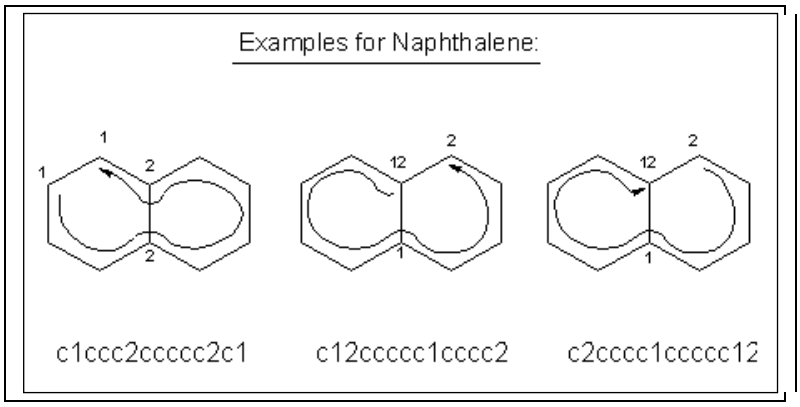
7. For cyclic structures, the numbers 1 through 9 are used to indicate the starting and terminating atom of the cycle.

a. The same number is used to indicate the starting and terminating atom for each ring. The starting and terminating atom must be connected to each other.

b. Numbers are entered immediately following the atoms used to indicate the starting and terminating positions.

c. A starting or terminating atom can be associated with two consecutive numbers.






8. Disconnected compounds are designated as individual structures or ions separated by a dot (".").Adjacent atoms separated by dot (“.”) are not directly bonded to each other, e.g. Van der Waals bonding:

**Aminopropene hydrochloride is C=CC(N).HCl**

9. Isomeric configuration is specified by the "slash" characters "\" and "/". These symbols indicate the relative direction between two isomeric bonds. (cis =”/ \”, trans = “/ /”). SMILES uses local chirality, which means that the chirality must be completely specified:

**cis-1,2-dibromoethene is Br/C=C\Br**
**trans-1,2-dibromoethene is Br/C=C/Br**

10. Enantiomers or chirality are specified by the ”@“ symbol. The symbol ”@” indicates that the following neighbours of the chiral atom are listed anticlockwise. If the symbol ”@@” is used, the atoms are listed clockwise. The chiral atom and the “@” are shown between brackets:

**2-chloro-2-hydroxypropanoic acid with**
**specified chirality is C[C@](Cl)(O)C(=O)(O)**

11. Isotopic specifications are indicated by preceding the atomic symbol with a number equal to the relevant integral atomic mass. An atomic mass can only be specified inside brackets:

**Carbon-13 is [13C] and Oxygen-18 is [18O]**

For the determination of the SMILES notation, several tools (SMILES generators) are available (see Appendix 1)

## 7 Information on optical activity

Optical activity is the ability of asymmetric substances to rotate the orientation of planar polarized light. Such substances, and their mirror images, are known as enantiomers and have one or more chiral centres. Although differing in geometric arrangement, enantiomers possess identical chemical and physical properties. Since each type of enantiomer affects polarized light differently, optical activity can be used to identify which enantiomer is present in a sample and therefore, also the purity of the substance. The magnitude of the rotation is an intrinsic property of the molecule.

Enantiomers always have opposite rotations: they polarize light to the same extent, but in opposite directions. The optical activity of an enantiomer mixture is therefore an indication of the ratio between the two enantiomers. A 50-50 mixture of enantiomers has an optical activity of 0.

The observed rotation depends on the concentration, the length of the sample tube, the temperature and the wave length of the light source.

Optical activity is, therefore, the defining parameter to identify an asymmetric substance; and it is the only parameter to distinguish the substance from its mirror image. Therefore, if applicable, the optical activity of the substance should be given.

The standard for optical activity is called the specific rotation. The specific rotation is defined as the observed rotation of light at 5896 angstrom, with a path length of 1 dm and at a sample concentration of 1 g/ml. The specific rotation is the observed rotation divided by the path length (dm) times the concentration (g/ml).

Optical activity can be measured with several different methods. The most common are:

- Optical rotation, in which the rotation of the plane of polarisation of a beam of light passed through the sample is measured;

- Circular dichroism, in which the absorption of right and left polarised light by a sample is measured.

If the substance rotates the light to the right (clockwise) it is called dextrorotatory and is designated with a + sign. If it rotates light to the left (counter clockwise) it is called laevorotatory and it is designated with a - sign.

## 8 Molecular weight or molecular weight range

The molecular weight is the weight of a molecule of a substance expressed in atomic mass units (amu) or as the molar mass (g/mole). The molecular weight may be calculated from the molecular formula of the substance: it is the sum of the atomic weights of the atoms making up the molecule. For molecules like certain proteins or undefined reaction mixtures, for which a single molecular weight cannot be determined, a molecular weight range can be given.

Several methods can be used to determine the molecular weight of substances:

- For determining the molecular weights of gaseous substances, Avogadro's law can be used, which states that under given conditions of temperature and pressure a given volume of any gas contains a specific number of molecules of the gas

$$PV = nRT = NkT$$

n = number of moles

R = universal gas constant = 8.3145 J/mol K

N = number of molecules

k = Boltzmann constant = 1.38066 x 10-23 J/K = 8.617385 x 10-5 eV/K

k = R/NA

NA = Avogadro's number = 6.0221 x 1023 /mol

- For liquids and solid substances the molecular weight can be determined by determination of their effects on the melting point, boiling point, vapour pressure, or osmotic pressure of some solvent;

- Mass spectrometry, a highly accurate measurement method;
- For molecules of complex substances with high molecular weights, like proteins or viruses, the molecular weights may be determined by measurement of, for example, sedimentation rate in an ultracentrifuge or by light-scattering photometry;
- Several tools are available which can calculate the molecular weight on the basis of a structural diagram or a molecular formula of the substance (see Appendix 1).

## 9 Substance composition

For each substance the substance composition as a combination of the main constituents, additives and impurities shall be reported in line with the rules and criteria described in Chapter 4 of the guidance document text.

Each constituent, additive or impurity needs to be properly identified by:

- Name (IUPAC name or other internationally accepted name);
- CAS number (if available);
- EC number (if available).

For each constituent, additive or impurity, its percentage should be given (preferably by weight, or by volume),giving, where possible, the range in the commercial substance.

For the constituent(s), the typical percentage purity with the upper and lower limits for typical commercial batches should be given; for additives and impurities the typical percentage purity or the upper and lower limits should be given. Normally, the given values should add up to 100%.

## 10 Spectral data

Spectral data are needed to confirm the structure given for a mono-constituent substance or to confirm that a reaction mixture is not a preparation. Several methods can be used for spectra (ultra-violet, infra-red, nuclear magnetic resonance or mass spectrum). Not all methods are suitable for all types of substances. Where possible, the guidance document will give guidance for the appropriate spectra to be included for different substance types (ECB, 2004; ECB, 2005).

For several of the well-known methods the following information should be indicated on the spectrum itself or in annexes:

*Ultraviolet-Visible (UV-VIS) spectrum*

- The identity of the substance;
- Solvent and concentration;
- Range;
- Position (and epsilon values) of main peaks;
- Effect of acid;
- Effect of alkali.

*Infrared Spectroscopy (IR) spectrum*

- The identity of the substance;
- Medium;
- Range;
- Results (indicate the main peaks important for the identification e.g. interpretation of the fingerprint area).

*Nuclear Magnetic Resonance Spectroscopy (NMR) spectrum*

- The identity of the substance;
- Nucleus and frequency;
- Solvent;
- If appropriate, internal or external reference;
- Results (indicate the signals important for substance identification and the signals corresponding to the solvent and the impurities);
- For 1H NMR spectra the integration curve should be provided;
- The intensity of weak NMR peaks should be increased vertically and complex patterns expanded.

*Mass Spectroscopy (MS) Spectrum*

- The identity of the substance;
- Accelerating voltage;
- Method of loading (direct insertion, via GC, etc.);
- Ionisation mode (Electron Impact, Chemical Ionisation, Field Desorption, etc.);
- The molecular ion (M);
- Significant fragments for the identification of the substance;
- M/z values or assignments of the peaks important for the identification of the structure;
- Complex patterns should be expanded.

Other internationally recognised methods can be used as well if the spectral data will confirm the identification of the substance, e.g. the internal structure. Examples include XRD to identify the constituents of complex mineral oxides and XRF to analyse their chemical composition.

The following general requirements are needed for a clear understanding and/or interpretation of the spectra:

- Note significant wavelengths or other data as appropriate;

- Provide extra information, e.g. spectra of starting materials;
- Give solvent used and/or other essential details as indicated above for some methods;
- Provide clear copies (rather than originals) with scales properly marked;
- Provide information on the substance concentrations used;
- Ensure the most intense substance-related peaks approach the full-scale mark.

## 11 High performance liquid chromatography, gas chromatography

Where appropriate to the type of substance, a chromatogram needs to be provided to confirm its composition. For example, an appropriate chromatogram will confirm the existence of impurities, additives and the constituents of a reaction mixture. The two best known methods for separation and identification of mixtures are gas chromatography (GC) and high performance liquid chromatography (HPLC). The two methods are based on the interaction of a mobile phase with a stationary phase, leading to separation of the constituents of a mixture.

For GC/HPLC chromatograms the following information should be indicated on the chromatogram itself or in annexes (ECB, 2004; ECB, 2005):

- *HPLC*
  - The identity of the substance;
  - Column properties, such as diameter, packing, length;
  - Temperature, also temperature range if used;
  - Composition of the mobile phase ,also range if used;
  - Concentration range of the substance;
  - Visualisation method, e.g. UV-VIS;
  - Results (indicate the main peaks important for substance identification);
- *GC*
  - The identity of the substance;
  - Column properties, such as diameter, packing, length;
  - Temperature, also temperature range if used;
  - Injection temperature;
  - Carrier gas and pressure of carrier gas;
  - Concentration range of substance;
  - Visualisation method, e.g. MS;
  - Peak identification;

- Results (indicate the main peaks important for substance identification).

## 12 Description of the analytical methods

*Annex VI* of REACH requires the registrant to describe the analytical methods and/or to provide the bibliographical references for the methods used for identification of the substance and, where appropriate, for the identification of impurities and additives. This information should be sufficient to allow the methods to be reproduced.

# APPENDIX III - UPDATE OF THE DOCUMENT

All the changes introduced to the document are listed in the table below except minor changes such as corrections of typographical errors, slight amendments to wording of a sentence for improved English or addition of link to another guidance document.

Chapter 9 References has been deleted and the references have been moved to the relevant Chapters of this guidance document as foot notes.

**I. Changes from Version 1 to Version 1.1 (corrigendum, November 2011)**

| Section | Change made |
|---|---|
| Title | Reference to the CLP Regulation (Regulation (EC) No 1272/2008 of 16 December 2008) has been added in the title and in the chapters' titles. |
| General | References to the CLP Regulation have been included throughout the text as appropriate. |
| General | Additional text added to clarify the scope of the guidance document. Redundant text has been removed. |
| 1.1 | The term "TGD" has been changed to "guidance document" throughout the document. |
| 1.2 | The term "preparation" has been changed to "mixture" throughout the document following the amendment to REACH by the CLP Regulation (Regulation (EC) No 1272/2008 of 16 December 2008). |
| 1.2 | The term "item" has been changed to "section" throughout the document. |
| 1.3 | Redundant text has been removed. |
| 1.3 | The term "pre-registration" has been changed to "(late) pre-registration" throughout the document. |
| 2.1 | The abbreviations AAS and CLP have been inserted while RIP and TGD have been removed. |
| 2.2 | Descriptions of Alloy, EC Inventory, IUCLID have been amended. Definitions of EC Number, List Number, Mixture and Notified Substance have been introduced. The definition of "preparation" has been deleted. |
| 3.2 | The section has been revised to clarify the content. |
| 3.3 | The section has been revised to clarify the content with respect to the CLP obligations. |
| 4.2.2.1 | The way to present the constituents has been changed from concentration percentage to alphabetical order, so that the relative composition cannot be deduced from the list order. |
| 4.2.3.1 | The term lattice has been changed to crystal. |
| 4.3.1.2.3 | The section has been revised to clarify the content. |
| 5 | Redundant text has been removed. |
| 5 | Reference to the Data Submission Manuals Part 18 "How to report the substance identity in IUCLID 5 for registration under REACH" has been included. |
| 5 | The section has been revised to clarify the content. |
| 6.1 | Description of pre-registration has been changed to (late) pre-registration. |
| 6.1 | Redundant text has been removed. |
| 7.2 | Redundant text has been removed to clarify the content. |
| Appendix 1 | Hyperlinks have been updated or changed where not working properly. |

| | |
|---|---|
| Appendix 2 | Section 4.3 has been removed since its content can be found in the relevant website. |

**II. Changes from Version 1.1 to Version 1.2 (corrigendum, March 2012)**

| | |
|---|---|
| 2.2 | The definition of phase-in substance has been aligned with the defifition in Regulation (EC) No 1907/2006 as introduced by Council Regulation (EC) No 1354/2007 and by Corrigendum, OJ L 36, 5.2.2009, p. 84 (1907/2006). |

**European Chemicals Agency**
P.O. Box 400, FI-00121 Helsinki
http://echa.europa.eu