TECHNICAL SPECIFICATION

ISO/TS 19844

First edition
2015-12-15

Health informatics — Identification of medicinal products — Implementation guidelines for data elements and structures for the unique identification and exchange of regulated information on substances

Informatique de santé — Identification des médicaments — Lignes directrices pour la mise en oeuvre des éléments de données et structures pour l'identification unique et l'échange d'informations réglementées sur les substances

Reference number
ISO/TS 19844:2015(E)

ISO copyright office
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the WTO principles in the Technical Barriers to Trade (TBT) see the following URL: Foreword - Supplementary information

The committee responsible for this document is ISO/TC 215, *Health informatics*.

Introduction

This Technical Specification is a guide for implementing ISO 11238, *Health informatics — Identification of medicinal products — Data elements and structures for the unique identification and exchange of regulated information on substances*. This Technical Specification was developed in response to a worldwide demand for guidance on the implementation of internationally harmonised specifications for medicinal products. It is one of a group of four implementation guides for a total of five ISO standards which together provide the basis for the unique identification of medicinal products. The other standards in this group are:

- ISO 11615, *Health informatics — Identification of medicinal products — Data elements and structures for the unique identification and exchange of regulated medicinal product information*

- ISO 11616, *Health informatics — Identification of medicinal products — Data elements and structures for the unique identification and exchange of regulated pharmaceutical product information*

- ISO 11239, *Health informatics — Identification of medicinal products — Data elements and structures for the unique identification and exchange of regulated information on pharmaceutical dose forms, units of presentation, routes of administration and packaging*

- ISO 11240, *Health informatics — Identification of medicinal products — Data elements and structures for the unique identification and exchange of units of measurement*

The standards for the Identification of Medicinal Products (IDMP) support the activities of medicines regulatory agencies worldwide by jurisdiction. These include a variety of regulatory activities related to development, registration and life cycle management of medicinal products as well as pharmacovigilance and risk management.

The business objective of this implementation guide is to provide a means for exchanging regulatory substance information. To meet the primary objectives of the regulation of medicines and pharmacovigilance, it is necessary to exchange medicinal product information in a robust and reliable manner.

For the purposes of this Technical Specification, all conditions (e.g. mandatory, conditional, optional) correspond to the necessary requirements to uniquely and unambiguously identify a substance. Implementation of the ISO IDMP standards may dictate that mandatory elements for identification be tagged as conditional or optional, based on regional requirements. If a section is identified as 'optional' but is implemented in a specific region, conformance described within that section is applicable. The scope of this Technical Specification is to identify the scientifically necessary elements for the unique identification of substances/specified substances.

Health informatics — Identification of medicinal products — Implementation guidelines for data elements and structures for the unique identification and exchange of regulated information on substances

1 Scope

This Technical Specification is used in the implementation of ISO 11238. This Technical Specification defines substances based on their scientific identity (i.e. what they are) rather than on their use or method of production.

ISO 11238 provides the conceptual framework for defining substances and specified substances and for assigning unique identifiers in the context of the ISO IDMP standards. ISO 11238 describes general concepts for defining and distinguishing substances and a high level model for the structuring of information for substances. This Technical Specification provides detailed explanations of each type or grouping of substance information, an element-by-element description for implementation of ISO 11238, and examples for a variety of substances and specified substances.

This first edition of the Technical Specification will only address substances, and Groups 1 to 3 of the specified substances as defined in ISO 11238 and Annexes A, B, C, and D. It is anticipated that specified substances Group 4, as defined in ISO 11238, will be addressed in a subsequent edition of this Technical Specification. Some information that would typically fall under specified substances Group 4 may be covered in the Annexes of this Technical Specification. This information, although not defining of either a substance or a specified substance Group 1, may be essential to distinguishing substances.

This Technical Specification addresses the following:

- Data elements necessary for defining substances and specified substances Groups 1 to 3;
- The logical use of data elements as defined in ISO 11238;
- Substances and specified substances Groups 1 to 3 business rules for
  - determining necessary data elements,
  - distinguishing and defining materials according to ISO 11238,
  - triggering the assignment of identifiers.

This Technical Specification does not address the following:

- Business processes for data management;
- Implementation of a specific data information system (e.g. a relational database schema);
- Normative messaging standards for substances;
- The maintenance of controlled vocabularies;
- The specific global identifier system that should be used;
- Nomenclature standards for substances.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 8601, *Data elements and interchange formats — Information interchange — Representation of dates and times*

ISO 11238, *Health informatics — Identification of medicinal products — Data elements and structures for the unique identification and exchange of regulated information on substances*

3 General background and history

Due to the lack of a common and harmonised approach to define substances, regulators and pharmaceutical industry are faced with the inability to:

1) effectively exchange medicinal substance information in a structured and efficient way;

2) ensure data consistency and evaluate/compare information across regions, which especially impairs pharmacovigilance and compliance activities;

3) develop consistent terminology for use throughout the healthcare community.

The objectives of the IDMP standards are to address the issues outlined above by developing harmonised standards that build on the regulatory and technical processes already established and to support the population and maintenance of existing systems/applications with fully reliable regulatory medicinal product information.

Harmonised standards will stimulate vendors to develop "off-the-shelf" tools (that are interoperable due to the standard itself). Harmonised standards will also help to maximise forward compatibility of data and minimise the complexities of backward compatibility.

This implementation guide is intended to assist reporters (including pharmaceutical companies, regulatory authorities and non-commercial sponsors) in constructing messages or transmitting information that allows substances to be defined unambiguously and assigned unique IDs. It also provides guidance to help choose the correct Substance ID from a public data source that provides unique substance and specified substance identifiers. It is anticipated that an extensive list of substance identifiers as well as the definitional elements upon which the ID was based will be provided. This Technical Specification is not intended to be a guide for a maintenance organisation. The maintenance organisation may also create alternative methods to submit information consistent with the ISO model.

Table 1 is an example table for class and elements description.

Table 1 — Example table for class and element description

User Guidance	
Example(s)	
Conformance	
Data Type	
Values Allowed	
Business Rule(s)	

In contrast to other parts of the guide, conformance will refer to whether an element is required for a given substance type or a specified substance group. Conformance is not meant to be applied globally to all types of messages.

Definition: Conformance will be expressed based on the following terminology: Mandatory, Conditional and Optional.

Mandatory: refers to data elements that are required and shall therefore be implemented.

Conditional: refers to data elements that are subject to business rules and may became required by:
- data rules;
- process rules;
- regional rules.

Optional: refers to data elements that are informative but not definitional.

The description on whether a data element is conditional by data, by process or by regional rule is out of scope of this Technical Specification and will be defined within regional implementation guides.

The information provided in the table refers to the global guidance. When there is no information in the conformance table row (e.g. information on business rule is not provided), please refer to the regional implementation guide.

4 Substance (Mandatory)

4.1 General

All medicinal products consist of substances; these substances can be active ingredients, excipients, or packaging materials. There are two fundamental levels of information described in ISO 11238, a “substance level” and a “specified substance” level. Both levels are included in the more generic concept of an ingredient. At the substance level, substances are defined based on inherent attributes rather than use or method of manufacture. At the specified substance level, four separate groups of elements provide additional information.

In order to define or distinguish material either at a substance or specified substance level, a number of attributes should be taken into consideration:

- For chemicals, the molecular structure is captured at the substance level;

- For proteins, the amino acid sequence, sites and type of glycosylation, and the presence and position of disulfide bonds will be captured at the substance level;

- For nucleic acids, the sequence, type of sugar and linkage will be captured at the substance level;

- For other polymers, the monomers used to synthesize the polymer, the structural repeating units, the molecular weight and/or a property related to molecular weight (e.g. viscosity), the source of naturally derived polymers and any modifications that irreversibly alter the molecular structure will be captured at the substance level;

- For structurally diverse material, taxonomic, anatomical and fractionation information, properties related to the underlying molecular structure of the material, and modifications that alter the underlying molecular structure will be captured at the substance level;

- Mixture substance consists of a simple combination of single substances that are either isolated together or are the result of the same synthetic process. The biological source of the mixture is also captured where relevant at the substance level. Proportions are not captured at the substance level. It should be noted that a mixture

substance description should only include the substances that are generally or consistently present in the material. This excludes impurities and degradants.

Other attributes will be specific to the specified substance levels:

- Constituent substances in a multi-substance material;
- Proportions of constituent substances in a multi-substance material;
- Physical state;
- Grade or purity of material;
- Manufacturing information;
- Analytical data.

There are four groups of elements that are used to further define and specify substances. Specified substances are always composed of at least one substance.

Specified substance Group 1 is typically used to define:

- Multi-substance materials consisting of multiple substances, which are not defined as mixture substances;
- Additional information regarding herbal and allergenic extracts;
- Physical state, including polymorphic forms;
- Detailed glycosylation information.

Specified substance Group 2 is typically used to define:

- Manufacturer and the overall manufacturing process and critical process version number;
- In addition there is the possibility to make use of the reference source document class to store Release specifications of the intended manufactured substance.

Specified substance Group 3 is typically used to define:

- Grade or level of purity (Pharmacopoeial Specifications) and In house specification used to cover a set of specifications of all approved manufacturers for the substance.

Specified substance Group 4 is typically used to define:

- Detailed Analytical Data;
- Detailed Manufacturing Information.

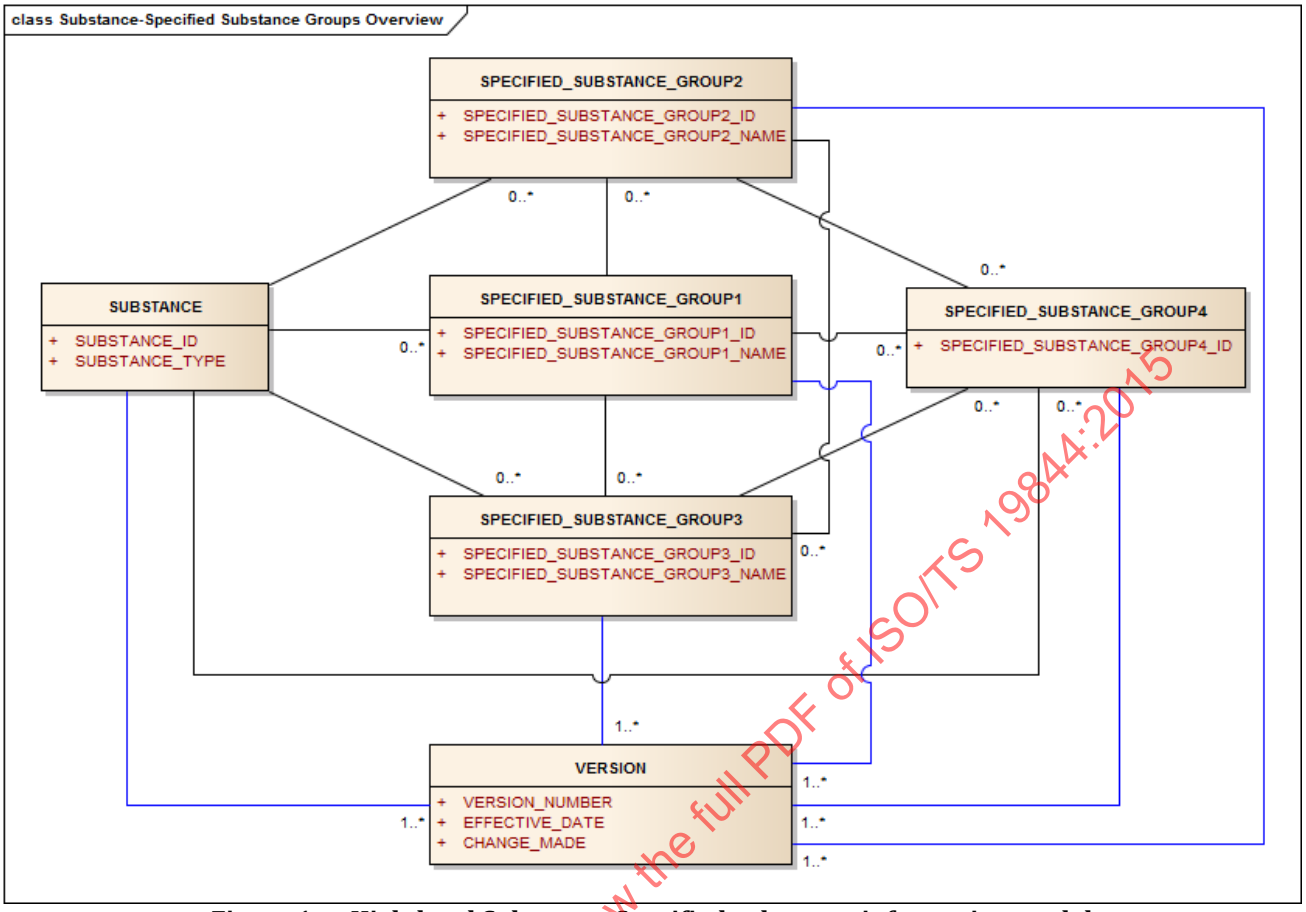


Figure 1 — High-level Substance-Specified substance information model

4.2 Defining substances

A substance is any matter that has a discrete existence, irrespective of origin, which may be biological or chemical. Substances can be single well-defined chemical entities containing a definite molecular structure, synthetic (e.g. isomeric mixtures) or naturally-occurring (e.g. conjugated oestrogens) mixtures of chemicals, or materials derived from plants, animals, microorganisms or inorganic matrices that are not definable by a single or limited number of molecular structures. Substances may be active moieties, salts, solvates, stoichiometric complexes or mixtures of compounds that are isolated or synthesised/ obtained or produced in a process together. Materials that are combined from multiple sources to form a product are not considered substances.

A substance is generally defined by what it is, and not by how it is made or used. Substance definitions are typically based on the immutable properties of a given material. These properties include the molecular structure, or structures of a given material, taxonomic, anatomical or fractionation information for material that cannot be represented by molecular structures. Purity, physical form, and method of production are typically not considered when defining substances.

In addition to defining information there is also information that is essential to validate the defining information and this information should be submitted if available. Validation may be performed based on relevant provided documents (e.g. regulatory dossier) or information available in recognised source (e.g. pharmacopoeias).

The primary goal of the ISO IDMP standard - *Data Elements and Structures for the Unique Identification and Exchange of Regulated Information on Substances* is to define unambiguously all substances present in regulated products. Once a substance has been defined, a unique identifier that is permanently associated with that substance will be assigned. This Technical Specification describes the necessary information for this registration process. Reference information, names, codes and IDs that can be associated with a substance are also described. The document is not comprehensive in regards to reference and definitional information. Other substance

attributes may be associated for both reference and definitional information as new scientific standards or information becomes available.

Substances that are mixed together to form a product, even if there are physical interactions between the substances will be described as a specified substance or a product.

EXAMPLE 1 FD&C Blue No. 1 Aluminium Lake is not defined as a mixture substance but rather as two separate substances FD&C Blue No. 1 and aluminium oxide. Also, simethicone is not a substance because it contains two substances, namely dimethicone and silicon dioxide. These materials would be described as Group 1 specified substances.

Conversely, if new covalent bonds and/or simple salts or solvates are formed, separate Substance IDs for each molecular entity are assigned.

EXAMPLE 2 Fluticasone propionate would have a separate ID from fluticasone; atorvastatin calcium trihydrate would have a separate ID from atorvastatin calcium anhydrous.

Differences at the specified substance level should trigger a new specified substance ID.

EXAMPLE 3 Insulin Human Crystalline and Insulin Human Amorphous will have different Group 1 specified substances IDs and map to the substance Human Insulin.

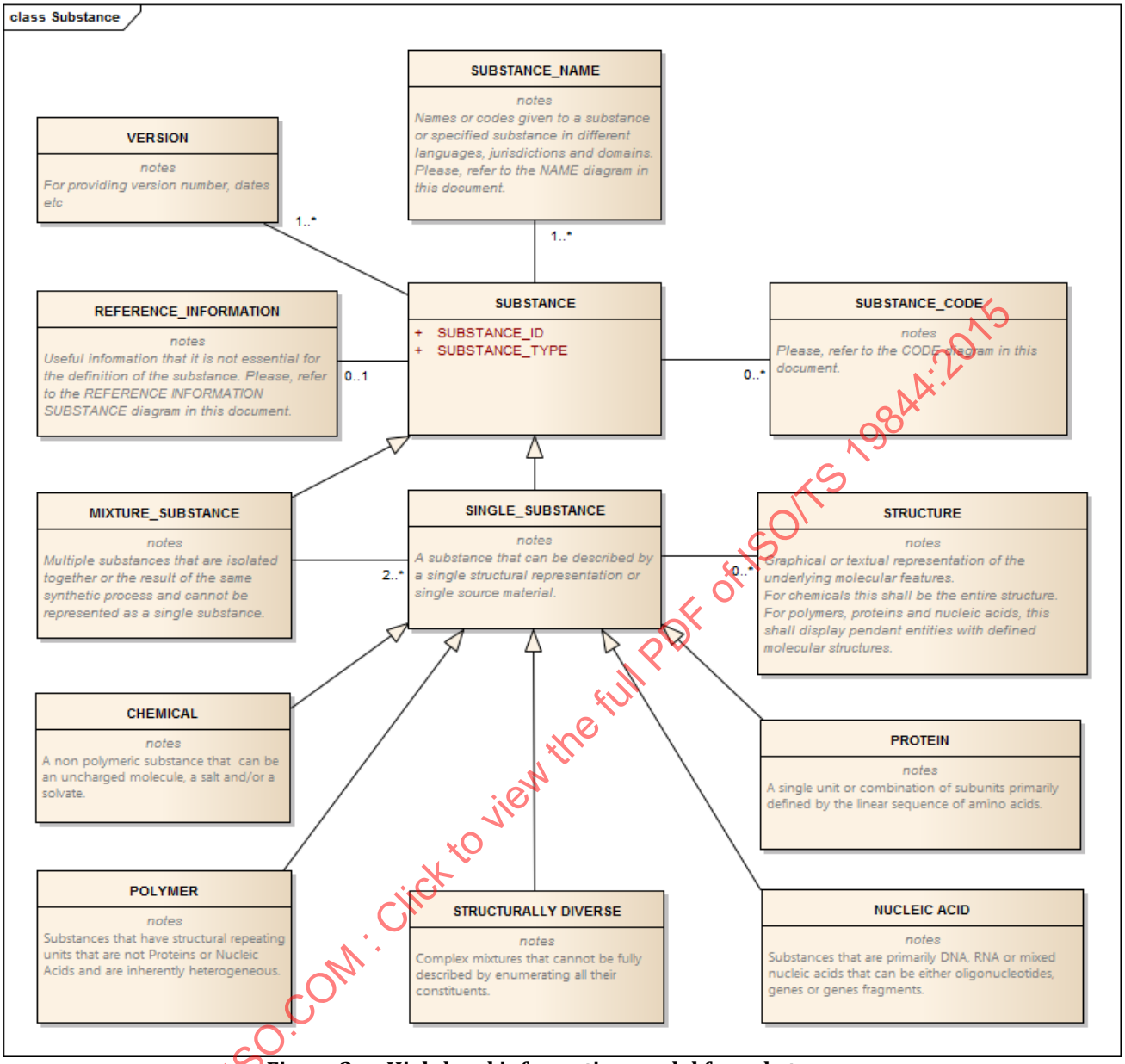


Figure 2 — High-level information model for substances

4.3 Substance types (Mandatory)

At the basic level, all material present in medicinal products will either be described as a substance and/or a specified substance Group 1. Additional information, such as manufacturer, grade and manufacturing information on a given material will be associated with specified substance Groups 2 to 4.

Substances are either one of five types of single substances or a mixture. Mixtures consist of a combination of related single substances where the source material and essential properties are also captured but proportions are not. For the purposes of definition, all single substances are of one of the following types:

- (simple) chemical;
- protein;
- nucleic acid;
- polymer;
- structurally-diverse.

Each of these different types of substances has distinct elements essential for definition. Substances can also be a **mixture of single substances,** typically related and of the same type. Some materials may appear to have elements from multiple types of substances e.g. PEGylated proteins, drugs covalently linked to polymeric matrices or proteins, etc., to fully define the substance. Each of these substances will be described as a single type, although it is possible for the same to be defined as two types one of which would be an alternative definition.

EXAMPLE 1: Proteins modified with a polymer or drug will be described as a protein with the polymer or drug captured as a modification. Drugs covalently linked to a polymer will be described as a polymer with the drug either captured as an end group or as structural modification. Detailed examples will be provided in the Annex to address these complex materials. Although every substance definition will be defined with a single type it is possible that an alternative definition of a substance will be of a different type.

EXAMPLE 2: Desmopressin can either be defined as a modified peptide or as a chemical substance.

Mixtures consist of a combination of related single substances where the source material and essential properties are also captured but proportions are not. At the specified substance Group 1 level proportions can be captured.

Additional information, such as manufacturer, grade, and manufacturing information on a given material will be associated with specified substances Groups 2-4. The first step in defining a material is to determine whether the material can be defined as a single substance, a mixture of single substances or a specified substance Group 1. A mixture substance is a substance that cannot be defined using a single set of substance elements or structural representations. Racemic mixtures, most polymers and structurally diverse substances, although inherently mixtures, can be defined using a single set of elements and therefore are not described as mixtures. Material that contains multiple substances that are not related and not synthesised or isolated together will be described as specified substance Group 1. Specified substance Group 1 would also include crystalline polymorphs, substances in varied physical states along with the capture of additional properties regarding sterility or the presence of endotoxin.

Examples of specified substance Group 1 include simethicone, adjuvanted vaccine antigens, isophane insulins and additional herbal and allergenic extract information not captured at the Substance level. Sterile water, water for injection, ice, and steam would be different specified substances Group 1 that would map to the substance water.

Mixtures are substances like Gentamicin sulfate, USP or Gentamicin sulfate, EP which consists of mixtures of distinct but related individual aminoglycosides. In USP Gentamicin sulfate is described as a mixture of Gentamicin C1A sulfate, Gentamicin C1 sulfate, and Gentamicin C2 sulfate. In EP it is described as a mixture of five components Gentamicin C1, C1a, C2, C2a, and C2b respectively which are typically isolated together from a bacterial culture. At the substance level all related substances typically present at a level greater than 1% would be listed as components of the mixture regardless of whether the substance is listed in a given pharmacopoeia. Mixture Substances are defined in the ISO 11238 Substance standard as a simple combination of single substances that are either isolated together of are the result of the same synthetic process.

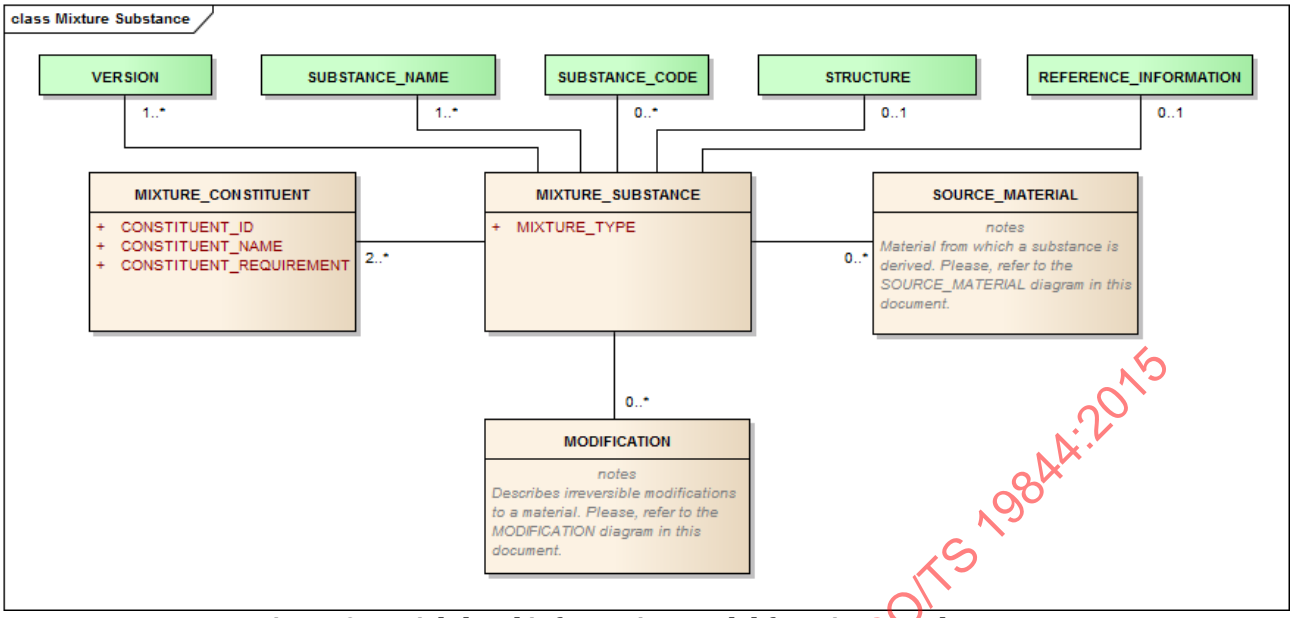
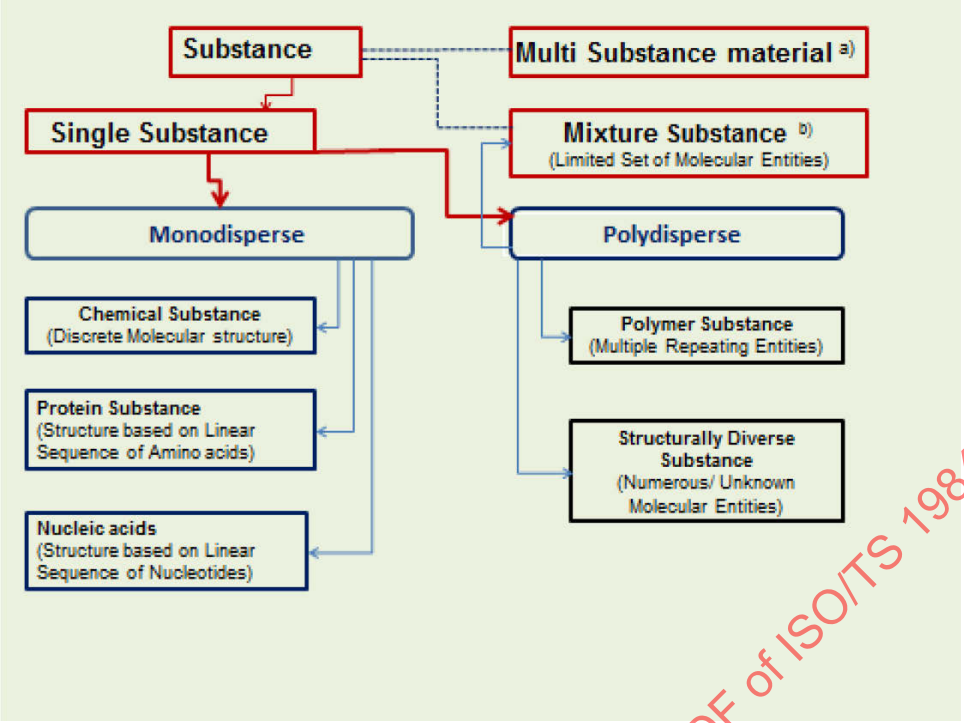


Figure 3 — High-level information model for mixture substance

For mixture substances in which multiple related substances are isolated together from a single source, the source will be a defining element. Starches are typically a mixture substance of the two distinct polymers amylose and amylopectin. Corn, wheat, potato and tapioca starches are all distinct substances distinguished by the source.

Most other polymers will be described as single substances with the definition based on the structural repeating unit and a representation of the average molecular weight of the polymer as well as the biological source for biopolymers or polymers derived from biological matrices.

Figure 4 represents a decision tree or process by which the type of substance is determined. Monodisperse substances are substances that can be described as a single molecular entity. Polydisperse substances are substances that typically have multiple molecular entities that are too numerous or too diverse to be captured as a mixture (e.g. cells or tissues) or where the production of the substance inherently results in polydispersity (e.g. polymers).



a) Multi-substance material: Multi substances and/or specified substances of diverse origin used as a component in the formulation in a medicinal product.

b) Mixture substance: Type of polydisperse substance that is a combination of single substances isolated together or synthesised/obtained or produced in the same process.

Figure 4 — Decision tree or process by which the type of substance is determined

Table 2 — Substance types

User Guidance	Information about the substance type as described in and in accordance with ISO 11238 definitions
Example(s)	"Chemical" "Protein" "Nucleic acid, "Polymer", "Structurally Diverse"
Conformance	MANDATORY
Data Type	CD
Values Allowed	Chemical, Protein, Nucleic Acid, Polymer, Structurally Diverse, Mixture, specified substance Group 1, specified substance Group 2, specified substance Group 3, specified substance Group 4

4.4 Substance ID (Mandatory)

Every substance will be identified by an ID. An ID will be generated following the initial submission of substance data. The ID should be unique to each substance, non-semantic, non-chronological, and of fixed length with an integrity check. Once an ID is assigned to a given physical or conceptual material the ID will not change. Changes or additions to defining information result in a new version of the substance definition but would not result in a new Substance ID. Although an effort should always be made to obtain complete information prior to assignment there will be instances when additional definitional information on a given substance is transmitted after assignment. When incorrect or incomplete information may be transmitted and used to define a substance the information will be changed or added but the ID will not change. In rare cases when the same substance is found to have two IDs one of the IDs will remain a primary ID and the use of the other ID will be deprecated and captured in the Substance Code class For some materials there may be multiple IDs that a material could map to. A submitter should always submit the ID that is the most descriptive of the material.

EXAMPLE Glucose, when a solid crystalline material can exist with an alpha or beta pyranose structure in an anhydrous state, or as a monohydrate with an alpha pyranose structure. If a solid crystalline form of glucose exists

in a medicinal product the ID specific to the structure should be used. Glucose in a liquid or an amorphous solid state would use the Substance ID that does not specify the specific form.

Both substances and specified substances will be identified with an ID. IDs will be released to the public if the substance is in a licensed medicinal product. The IDs for substances in the investigational stage will only be released if an official name exists or a company code is associated with defining information and is found together in at least one single reference that is from a reputable source in the public domain (i.e. scientific journal, presentations or posters at scientific conferences, company publication, patent or published patent application, public databases such as STN from CAS). Substance IDs will always be issued to the requester if sufficient information is provided to unambiguously define the substance.

Table 3 — Substance ID

User Guidance	ID to be used in all electronic submissions to identify a substance. Generated when sufficient information is available to unambiguously define a substance. ID will be permanently associated with a given substance and each substance at the substance level shall have one and only one ID. NOTE If a unique "Substance ID" has been assigned, this "Substance ID" shall be specified based on the Substance Name controlled vocabulary.
Example(s)	UNII Code, EV-Code, Global Substance ID
Conformance	MANDATORY
Data Type	II
Values Allowed	Value could be a code associated with a preferred term, or a specific type of data. If a code is transmitted the preferred term associated with that code should also be transmitted.
Business Rule(s)	All substances will be identified by a single ID. NOTE: The ID will only be released to the public if the defining information is in the public domain or if a company that provides the defining information requests public release or releases the code in public marketing materials. An ID will always be issued to an organisation that requests an ID and supplies the information necessary to define a substance. A flag to control the release of the ID to the general public is part of the Reference Source information (4.6).

4.5 Substance names (Mandatory)

Although ISO 11238 does not provide any guidance on substance nomenclature, it does provide a structure for the capture of names and codes that are used to refer to a substance. For the purpose of ISO 11238 there are basically five types of names, Official names, systematic names, other names, brand names and company codes.

At least one name or company code should be associated with each substance. Names for a given substance and/or standardised names submitted by another party may be available in the substance terminology. Indication that a name is the preferred name or term by which a substance will be referred to in a given jurisdiction and domain may be presented in the terminology. This preferred term may vary depending on jurisdiction or domain or the state of development of a given substance.

EXAMPLE During the early stages of clinical development, the preferred term for a substance may be its company code. In later stages an INN name may be associated with the substance and become the preferred term for that substance.

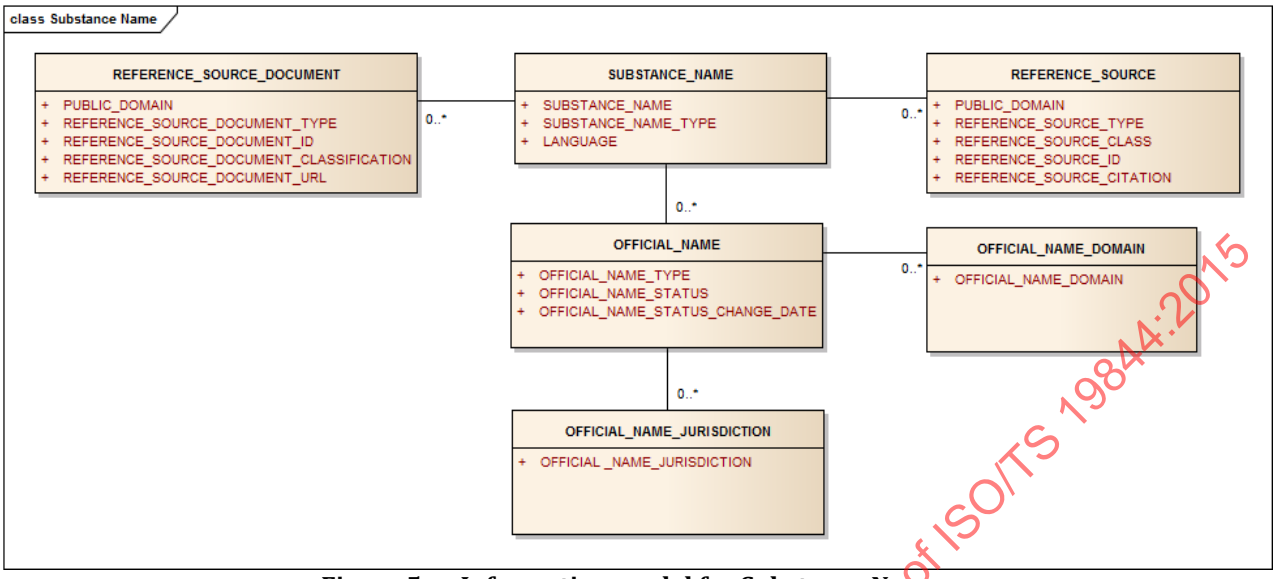


Figure 5 — Information model for Substance Names

4.5.1 Substance name

When requesting an ID a submitter should submit all names and company codes which have been associated with the substance. A submitter may also want to add a name to a substance that is already in the database. Defining information for a given substance should also be transmitted when requesting that a name is added to the database to ensure the correct substance identity. A submitter may also submit names in multiple languages. Although there is no requirement that names are unique, other or common names should also explicitly distinguish related substances from one another. Distinguishing characteristics for many polymers, such as molecular weight, degree of polymerisation, and viscosity, should be indicated in the names of these substances. Standardisation on such names to facilitate searching and identification of substances may be adopted within regions.

EXAMPLE Polyethylene glycol is a name for a family of polymers, so it would not be an acceptable name for an individual substance. Polyethylene Glycol 1000 is an official name in the US for a polyethylene glycol polymer with an average molecular weight of 1000. In Europe the INN name Macrogol 1000 would be an official name for the same substance. In the personal care products domain PEG-20 is the official name for the same substance and 20 refers to the average degree of polymerisation.

Table 4 — Substance name

User Guidance	Name or company code associated with the Substance.
Example(s)	"Amoxicillin"
Conformance	MANDATORY
Data Type	ST
Values Allowed	Free text, multiple names allowed, each in its own <asNamedEntity> element.

4.5.2 Substance name type

Every name shall have one and only one name type. Official names are typically non-proprietary names used in a given jurisdiction and domain to refer to a specific substance. The domain, jurisdiction, and authority that assigned the name (USAN, INN, JAN etc.) and the language of the name are also captured.

Systematic names will typically be used for simple chemicals and structurally diverse materials where the definitions are based on a chemical structure or systematic taxonomic information. For chemicals these names are typically derived from IUPAC or CAS systems of nomenclature. From the systematic names of chemicals a molecular structure can typically be derived and the name can be checked by a number of chemical drawing programmes that convert a given name to a molecular structure. Brand names are names by which a company identifies a given substance typically for marketing purposes. Other names associated with a given substance are for instance common names. Company codes (also known as lab/laboratory codes or clinical trial codes) are also treated as names. They are assigned by a given company to substances in clinical or preclinical development.

Table 5 — Substance name type

User Guidance	Each name shall be associated with a type.
Example(s)	Official name, systematic name, other name, brand name, company code
Conformance	MANDATORY
Data Type	CD
Values Allowed	OFFICIAL, SYSTEMATIC, COMPANY CODE, OTHER, BRAND
Business Rule(s)	A name shall have one and only one name type.

4.5.3 Language

All names should be associated with at least one language. Often the name, particularly INN names, will be identical in multiple languages. The language for company codes that use a Latin character set will be described as English (en). The language for the systematic name of plant, microorganism or animal derived substances will be also be described as English since many of these will be a hybrid of a Latin Name for the species and an English name for the part and/or the subspecies information.

EXAMPLE The INN name alisertib is identical in English, French and Spanish.

Table 6 — Language

User Guidance	If the name is language dependent, that language shall be specified. The language used to provide the structured substance information shall be specified according to the XML standard's xml:lang attribute specification, which is based on ISO 639-1, alpha-2 codes
Example(s)	en – English, de – German, fr –French
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	ISO 639-1, alpha-2 codes
Business Rule(s)	While xml:lang allows country-specification of language, e.g. en_US vs. en_UK or pt_PT vs. pt_BR, such specificity is to be avoided except if the country variant of the language exceptionally does give rise to an alternative spelling. In that case, only the exceptional spelling needs to be tagged with the country-specific language code, not the regular spelling for that language in all other countries. If the same name spelling exists for multiple languages of interest, the entire name is repeated each time with its own language code.

4.5.4 Official name (repeat as necessary)

Each official name shall have one or more official name types. The official name type reflects the organisation that assigns or recognises the name associated with the substance. These names are typically non-proprietary names that are used in the labelling of pharmaceuticals. The domains and jurisdictions in which the official name is used are also captured, tracked and maintained within the terminology. Although most official names will refer to

material at the substance level there will be instances where an official name will refer to material at the specified substances Group 1 level [e.g. Simeticone (BAN, INN) or Simethicone (USAN)]. The official name class is optional and shall be repeated as necessary.

4.5.4.1 Official name type

Any authority which assigns official names will typically do so in a publication, and as such, referencing the publication is sufficient indication for the name being "official" and sufficient representation for the "official name type". When new official names are issued by official naming bodies, the naming bodies will be reflected in this terminology.

Table 7 — Official name type

User Guidance	Designation of which authority assigned the Official Name. All official names need to have at least one such designation. The name type is the name of the authority or authorities that have assigned or have adopted the name.
Example(s)	BAN, COSING, EP, FCC, INCI, INN, JAN, JECFA, MARTINDALE, USAN, USP.
Conformance	MANDATORY — all official names require a designation of the naming authority
Data Type	CD
Values Allowed	All values in the name assigning authority identifier system.
Business Rule(s)	All official names must have a naming authority designated. If multiple authorities have assigned the same name, a separate <asNamedEntity> is specified for each authority that needs to be listed.

4.5.4.2 Official name status

The status of the official name is associated with a given jurisdiction. It is possible that an official name could be official in one jurisdiction and superseded in another. With some official name types an initial name will be proposed and the name will later become recommended. Recommended INN names would be considered to be the Primary Official name in many jurisdictions.

Table 8 — Official name status

User Guidance	The status of the official name.
Example(s)	Current, Alternate, Superseded, Proposed
Conformance	OPTIONAL
Data Type	CD
Values Allowed	Current, Alternate, Superseded, Proposed
Business Rule(s)	Each official name shall have a single status. The fact that a name is alternate is evident by having more than one name assigned to the same substance by the same authority for the same language, domain, etc.

4.5.4.3 Official name status change date

The change date of the status of the official name shall be specified according to the ISO 8601 date format (the variant without any delimiters).

Table 9 — Official name status change date

User Guidance	Date of official name change if known.
Example(s)	20110601
Conformance	OPTIONAL
Data Type	DATE
Business Rule(s)	If official name status changes the date of the official change will be captured if known.

4.5.4.4 Official name domain (repeat as necessary)

Official names will also be associated with a domain. The domain and the actual name of a domain will vary according to jurisdiction.

EXAMPLE The terms DRUG and BIOLOGIC could be used in Japan and the US as official name domain. In the EU the term MEDICINE may cover the name of many substances that would be either DRUGS or BIOLOGICS in the US and Japan.

Table 10 — Official name domain

User Guidance	Specifies when and for what purpose the official name is to be used. All official names can have at least one domain. This is useful to differentiate different names for the same substance as used for a drug active ingredient as opposed to a food colour additive.
Example(s)	BIOLOGIC, COSMETIC, DRUG, FOOD, MEDICINE, DIETARY SUPPLEMENT
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	Used if different names exist that are not appropriate to be used across all domains.

4.5.4.5 Official name jurisdiction (repeat as necessary)

When a product is on the market and contains pharmaceutically active substances at least one jurisdiction should be associated with the name of the active substance. For inactive substances or excipients the jurisdiction may not be apparent or captured. Pharmacopoeias will often be consulted to determine the jurisdiction of an official name. The Official Name Jurisdiction class is optional; when necessary the following conformances apply:

Table 11 — Official name jurisdiction

User Guidance	The jurisdiction of the official name shall be provided.
Example(s)	US, EU, JA
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	ISO 3166-1 alpha-2 codes For the European Union EU shall be used.
Business Rule(s)	All the jurisdictions in which the Official Name is to be used should be indicated.

4.6 Reference sources (Mandatory)

All data elements associated with a substance or specified substance should have a reference source. These sources could be regulatory submissions, publications by naming authorities such as INN or USAN, pharmacopoeias, books, journals, web sites, meeting materials of public databases such as chemical abstracts. A source can refer to multiple data elements and a given element can have multiple sources.

4.6.1 Public domain

Most reference sources other than regulatory submissions will be in the public domain. All regulatory submissions will typically be considered to not be in the public domain. If the reference source is in the public domain the data elements referenced by the source could be made public. There may be a separate determination whether any portion of the record can be made public. This determination may vary depending on type of substance. Regardless submitters of information should always explicitly indicate whether data should be considered confidential.

EXAMPLE Particular care should be taken in the release of company codes. Although the company codes are frequently mentioned in the public domain the release of a company code associated with a Substance ID and the defining information should only be made if all of the defining elements are associated with the company code in a single source in the public domain. For chemicals, the chemical structure should be associated with the company code in the public domain. For proteins such as monoclonal antibodies, names or company codes could be released to the public only if the target of the antibody is mentioned in the same public source as the name or code. A separate determination will be made as to whether the protein sequence is in the public domain.

Table 12 — Public domain reference source

User Guidance	Indication of whether the source is in the public domain. Company codes will only be released if the defining information is associated with the code in a public source such as chemical abstracts (CAS), a journal article, or a poster at scientific meeting attributable to the company.
Example(s)	Yes, No
Conformance	OPTIONAL
Data Type	BL
Values Allowed	Yes, No
Business Rule(s)	All regulatory submissions are considered not to be in the public domain. There may be an additional flag that determines whether any information can be released on a given substance.

4.6.2 Reference source type

All reference sources will be associated with a type. The type may actually identify the source of the information if it is a public database or naming authority. For official names and defining information, the primary source should be used if available. A single source can be used for many data elements. Public databases particularly those of official naming bodies, regulatory submissions, manufacturer's technical specifications, scientific journal articles and pharmacopoeias are all sources that can be used.

Table 13 — Reference source type

User Guidance	The reference source type in which the data elements were actually found.
Example(s)	BP, CHEMID (NLM), EP, IND, INN, ITIS, JAN, JOURNAL, JP, KEGG, MARTINDALE, NDA, PERSONAL CARE PRODUCTS COUNCIL (PCPC), PUBCHEM, USAN, USP, WEB PAGE
Conformance	MANDATORY
Data Type	CD
Business Rule(s)	All names should have at least one source and hence source type.

4.6.3 Reference source class

Each reference information type will be associated with a class. The class will be automatically associated with each type of information.

Table 14 — Reference source class

User Guidance	Useful for classification of sources.
Example(s)	Regulatory submission
Conformance	OPTIONAL
Data Type	CD
Values Allowed	LITERATURE, OFFICIAL NAME SOURCE, REGULATORY SUBMISSION, WEB.
Business Rule(s)	Need not be entered as the value is dependent on the Reference Source Type.

4.6.4 Reference source ID

Many sources such as regulatory submissions, INN lists and others will be associated with a number or ID. The ID should be captured.

Table 15 — Reference source ID

User Guidance	An ID associated with reference source.
Example(s)	IND number, INN list; PMID (PubMed Identifier).
Conformance	CONDITIONAL
Data Type	ST

4.6.5 Reference source citation

A citation should be captured for a literature reference. The specific format for a given citation may be defined at regional level.

Table 16 — Reference source citation

User Guidance	The actual reference source or citation in which the name was found.
Example(s)	Literature reference: Encycl. 3: 429 (1792)
Conformance	CONDITIONAL
Data Type	ST

4.6.6 Reference source document (new class to be included in the second edition of ISO 11238)

The document should be concise and contain information specific for given data elements.

Table 17 — Reference source document

User Guidance	Documents that contain relevant information on data elements and are referenced should be stored if available.
Conformance	OPTIONAL
Data Type	ED
Values Allowed	The documents can be saved as PDF files or TEXT files. If the documents or data source contains structured data such as spectra, it can be transmitted in a standard format that can be visualised and searched.

4.6.7 Reference source document type (new class to be included in the second edition of ISO 11238)

Table 18 — Reference source document type

User Guidance	The type of the document
Example(s)	Regulatory submission, Journal article , Release specification, In House specification
Conformance	OPTIONAL
Data Type	CD

4.6.8 Reference source document classification (new class to be included in the second edition of ISO 11238)

A classification system will be also developed for documents. The classification system will indicate the type of information contained within the reference document based on a terminology for the classification of reference source documents.

Table 19 — Reference source document classification

User Guidance	The reference document should be classified according to the information contained within
Example(s)	Chemical structure, Name, NMR spectra, Mass Spectra,
Conformance	OPTIONAL
Data Type	CD

4.6.9 Reference source URL (new class to be included in the second edition of ISO 11238)

Table 20 — Reference source URL

User Guidance	The reference source URL in which the name was found. The URL can be captured here.
Conformance	CONDITIONAL
Data Type	ST

4.7 Substance code (Conditional)

This section of reference information allows the capture of codes that are typically associated with a substance. The capture of these codes typically facilitates the defining and mapping of substances and linking of substances to a variety of information sources. Codes should be communicated if known but are not necessary or mandatory. The terminology should have the ability to capture, store and distribute codes as necessary. All the codes that are captured should be associated with a publicly recognised code system and map directly to a given substance. It should be noted that company codes are not captured in this section but are considered a type of name for a given substance. Commonly used public codes such as CAS registry numbers with a given substance should be provided and associated to the substance.

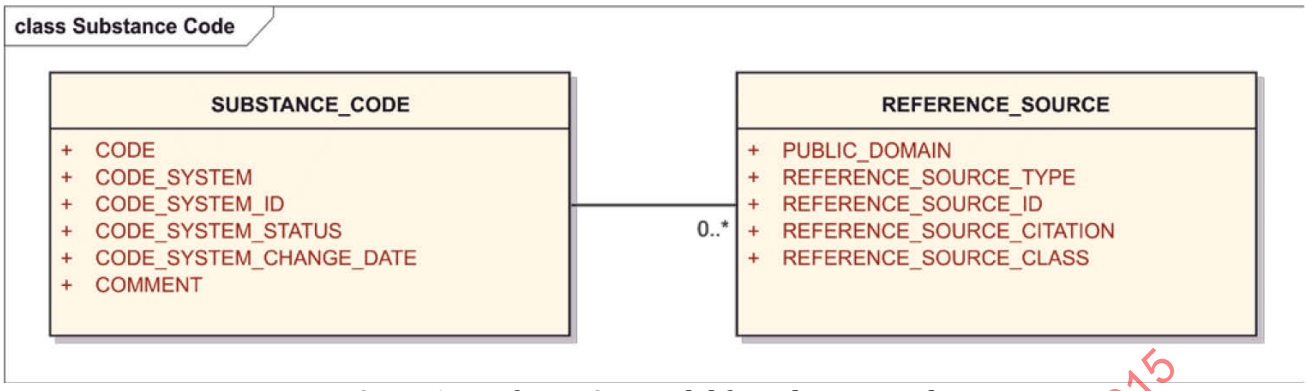


Figure 6 — Information model for substance codes

4.7.1 Code

The actual code should be captured using the same format that is used in the code system. Only codes associated with a code system will be captured. The code should be specifically associated with a given substance. Many public and non-public databases identify substances with a code and these codes can be very helpful in mapping substances to various systems. Codes should always be verified against the source system. Different jurisdictions may require a code from a code system or multiple code systems to be associated and submitted with a substance.

Table 21 — Code

User Guidance	The specific code (the value) with regard to a referenced code system. NOTE: InChI is not considered a code and shall be specified within the structure description.
Example(s)	11-22-356
Conformance	CONDITIONAL
Data Type	ST
Business Rule(s)	Only codes from recognised code systems will be captured. The code should specifically link to a substance but need not uniquely link to substance. This code can also be a deprecated Substance ID generated based on ISO 11238.

4.7.2 Code system

Every code should be associated with a code system and a list of names of code systems expected to be submitted should be maintained by the maintenance organisation.

Examples CAS registry numbers, EC numbers, FDA UNII codes, EMA XEVMPD codes, ASK numbers, EPA Pesticide codes.

Table 22 — Code system

User Guidance	Name of the code system.
Example(s)	CAS, EINECS, NSC, ASK, ATC, NDF-RT; RX-CUI ETC.
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	If the Code class applies, the code system name is implicit and derived from the Code System ID.

4.7.3 Code system ID

The code system ID may be an OID or an ID as applicable. The ID is associated with code system and need not be submitted.

Table 23 — Code system ID

User Guidance	The ID of the code system.
Example(s)	CAS Registry code: 0049; FDA Substance Registration System (UNII) 0050
Conformance	CONDITIONAL
Data Type	II
Business Rule(s)	If the Code class applies, the code system ID is required as all codes must be linked to a code system.

4.7.4 Code system status

The code system status is typically the status provided by the system for a given code. Many code systems can end up with multiple codes representing the same substance. Often one code will be the preferred code and others will either be deleted or deprecated. Ad-hoc terms may also be developed to indicate a code status.

Table 24 — Code system status

User Guidance	The status of the code assignment
Example(s)	current, proposed,
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	“active” — Current, “active” — Alternate, “superseded” — Terminated, “pending” — Proposed

4.7.5 Code system status change date

Table 25 — Code system status change date

User Guidance	The date at which the code status is changed as part of the terminology maintenance.
Example(s)	20110601
Conformance	OPTIONAL
Data Type	TS.DATE

4.7.6 Comment

Table 26 — Comment

User Guidance	Any comment can be provided in this field, if necessary.
Conformance	OPTIONAL
Data Type	ST

4.7.7 Reference source

The reference source will be captured according to 4.6.

4.7.7.1 Reference information

Where applicable, the reference information shall be provided.

4.7.8 Substance classification (repeat as necessary)

The standard should have the ability to capture a variety of classification systems. Classification systems are typically based on molecular structure, chemical properties, pharmacological effects, mechanism of action, therapeutic targets or indication.

The standard shall have the ability to capture multiple classifications and variable levels of classification. Although most classification will be associated with an external classification system, ad-hoc classification of substances may be developed within this terminology as needed.

The information in Figure 7 shall be provided:

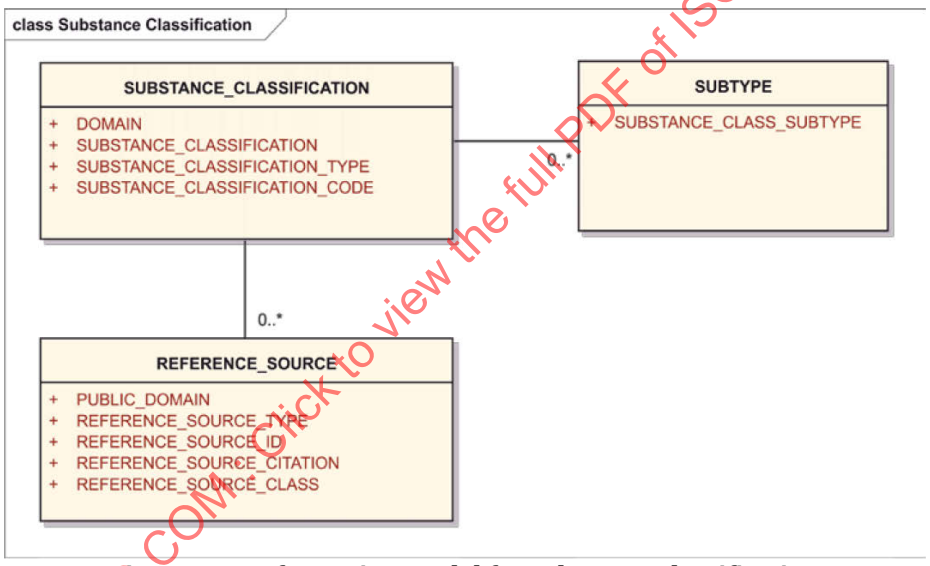


Figure 7 — Information model for substance classification

4.7.8.1 Domain

Classification systems will often only be used within a specific domain. The domain will be associated with the classification system.

Table 27 — Domain

User Guidance	The domain of the substance classification shall be provided.
Example(s)	Human pharmaceuticals, human vaccine, animal drug, animal vaccine, food, food additive, colorant, pesticide; tobacco additive; flavour, excipient
Conformance	OPTIONAL
Data Type	CD

4.7.8.2 Substance classification (code system)

There are a number of systems that classify substances based on chemical structure, pharmacological effect, organ system affected, mechanism of action or molecular targets. These classification systems are useful and this information, if available, should be provided and maintained as applicable.

Table 28 — Substance classification (code system)

User Guidance	Substance Classification terms based on a controlled vocabulary that is controlled by the system or type of substance classification. The system should be capable of capturing a variety of classification systems and each system should be identified. All active ingredients are typically classified in multiple systems. Ad-hoc classification may be created as necessary.
Example(s)	ATC, NDF-RT, NCI, AHPA, MESH; Therapeutic Category of Drugs in Japan
Conformance	OPTIONAL
Data Type	CD

4.7.8.3 Substance classification type

Nearly all classification systems contain multiple levels. Type is the highest level or the initial level of the classification system.

Table 29 — Substance classification type

User Guidance	The initial type or highest level of a classification system.
Example(s)	Organ system in ATC.
Conformance	OPTIONAL
Data Type	CD
Business Rule(s)	The value is IMPLICIT and can be derived by the “Substance Classification — Subtype “if available.

4.7.8.3.1 Substance classification subtype (repeat as necessary)

Table 30 — Substance classification subtype

User Guidance	Many classification systems will have multiple subtypes of classification this field will attempt to capture this information.
Example(s)	NDF-RT Mechanism of Action (MOA) or Physiological Effect.
Conformance	OPTIONAL
Data Type	CD

4.7.8.3.2 Substance classification code

Classification code with map to the active moiety of a given substance.

Table 31 — Substance classification code

User Guidance	Code from the classifying code system. The code should be tied to the substance...
Example(s)	N0000175684 NDF-RT code for Oxycodone that maps to the mechanism of action classification. N02AA05 ATC code that maps to Oxycodone.
Conformance	OPTIONAL
Data Type	II
Business Rule(s)	The classification code shall be part of an internationally recognised classification system.

4.7.8.3.3 Reference source

The reference source will be captured according to 4.6.

4.7.9 Target

The target section allows the capture of information related to the targets that a substance is known to interact with and the interaction results in a biological effect or transformation. These targets can be therapeutic, metabolic, interactions postulated to result in toxicity or interactions of unknown effect. Targets are typically proteins, or complexes of proteins. The target may be a substance in its own right and ID shall be assigned and/or existing ID of a known system used (i.e. GENE, Uniprot) to refer to a target.

For monoclonal antibodies the target shall always be provided. For other types substances the target may be provided if known. Ideally the target is a substance with its own Substance ID. However, in specific circumstances, the target may be handled as a relationship between two substances.

4.7.9.1 Target ID

Typically the Target ID is either a Substance or specified substance ID. ISO 11238 will be updated to rename Target Gene ID element as Target ID.

Table 32 — Target ID

User Guidance	ID of the target relation molecular entity upon which the substance acts. One Target ID should be specified. The target should be a molecular entity or a family of molecular entities. The target is usually a protein entity that the drug actually interacts with. Both therapeutic and metabolic targets and targets associated with toxicity should be captured.
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	ISO 11238 Substance ID
Business Rule(s)	The ISO 11238 Substance ID shall be referenced.

4.7.9.2 Target name

At least one name of the target shall be provided when target information is submitted. Target names will typically be handled no differently than other substance names. If an active ingredient targets more than one target each target should be listed. In addition to names a list of codes associated with the target may also be used. Gene bank and Uniprot codes can be particularly useful for identifying and tracking targets. The ISO 11238 standard will be updated to rename Target Gene Name element as Target Name.

Table 33 — Target name

User Guidance	Name of the target molecular entity upon which the substance acts. Each target should be associated with a target name.
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	ISO 11238 substance name
Business Rule(s)	ISO 11238 substance name implicitly derived from Target ID. A variety of names can be captured. A single name will be a preferred name and associated with the Target ID.

4.7.9.3 Interaction type

The type of interactions between a substance and its target should also be captured. A controlled vocabulary should be developed to capture these types of interactions.

Table 34 — Interaction type

User Guidance	Type of interaction between the substance and the target. The interaction type will be captured for all substrate-target interactions. A substance can have multiple types of interactions.
Example(s)	AGONIST, INDUCER, INHIBITOR, PARTIAL ANTAGONIST, SUBSTRATE, PRODUCT
Conformance	CONDITIONAL
Data Type	CD

4.7.9.4 Target organism (new class to be included in the second edition of ISO 11238)

When drugs are targeted against a specific organism the organism should be captured. For antivirals and antibacterials there is typically no need to give the strain level unless the organism is resistant to more commonly used drugs.

Table 35 — Target organism

User Guidance	The organism for which the active substance is targeted.
Example(s)	HOMO SAPIENS; M. TB, HIV-1 etc.
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	At least one organism type should be associated.

4.7.9.5 Target organism type

Every organism should have a type that is maintained within this terminology. The organism type may also be used independent of the organism particularly for substances such as antibacterials which may be targeted against a wide range of organisms.

Table 36 — Target organism type

User Guidance	Every target organism should have a type associated with the organism.
Example(s)	HUMAN, VIRAL, BACTERIAL, FUNGAL, INSECT, HELMINTH, MALARIAL
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	At least one organism type should be associated.

4.7.9.6 Target type

Each target-substance interaction should be assigned a type. The type is dependent on the substance and also to some extent the use of the substance. A given target-substance interaction can be of more than one type.

EXAMPLE The therapeutic target would be HIV-1 protease when ritonavir is used in HIV-related therapy. The metabolic target is Cyp 3A4 and ritonavir is an inhibitor of this target. Ritonavir is also an inhibitor of P-glycoprotein a drug transporter. In many combination products ritonavir is used strictly to inhibit Cyp 3A4/5 and transporter enzymes.

Table 37 — Target type

User Guidance	Type is a limited form of classification associated with a target of an active substance. Some targets can be both toxic and therapeutic; the choice is dependent on the dose of a given substance.
Example(s)	THERAPEUTIC, METABOLIC, TOXIC; TRANPORTER
Conformance	OPTIONAL
Data Type	CD
Business Rule(s)	Each target should be associated with a target type when the target can be classified,

4.7.9.6.1 Reference source

The reference source will be captured according to 4.6 when necessary.

4.7.10 Gene

The gene section will only be captured for proteins and nucleic acids. It does not apply for chemicals, polymers or structurally diverse substances. It is only used as reference information and refers to the gene that is the origin of the substance (typically a protein). Genes may also be substances and would be described as nucleic acids.

4.7.10.1 Gene sequence origin

For monoclonal antibodies the gene sequence origin will be used to capture species from which the antibody was derived.

EXAMPLES For monoclonal antibodies the typical values can be HUMAN, HUMANISED MOUSE; HUMANISED RAT; CHIMERIC MOUSE-HUMAN; CHIMERIC RAT-HUMAN.

Table 38 — Gene sequence origin

User Guidance	Common name for the organism from which the final gene sequence originated.
Example(s)	Human, Bovine, Human; Humanised; Mouse-Human Chimeric; Mouse
Conformance	OPTIONAL
Data Type	CD
Values Allowed	Typically from the NCBI Gene database.
Business Rule(s)	NOTE: it applies primarily to proteins.

4.7.10.2 Gene ID

Table 39 — Gene ID

User Guidance	ID associated with the gene of origin
Example(s)	YP_299723.1
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	Typically from the NCBI Gene database.

4.7.10.3 Gene name

Table 40 — Gene name

User Guidance	Complete name given for a gene. Every gene which has an ID should also have a name.
Example(s)	hIL-10 gene
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	Implicit from the Gene ID from the applicable Gene database.

4.7.10.3.1 Reference source

The reference source and related information will be captured according to 4.6.

4.7.11 Gene elements

Gene elements will be captured for genes that are used in gene therapy.

When the gene element section applies, the following information shall be provided:

4.7.11.1 Gene element type

Table 41 — Gene element type

User Guidance	Type of the gene element. Should be captured for all genes that are transduced into cells and are intended to be expressed.
Example(s)	enhancer, promoter, silencer, terminator, repressor
Conformance	CONDITIONAL
Data Type	CD

4.7.11.2 Gene element ID

Table 42 — Gene element ID

User Guidance	Unique identifier for gene element.
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	A unique identifier will be associated with each gene element.

4.7.11.3 Gene element name

Table 43 — Gene element name

User Guidance	Specific gene element name.
Example(s)	SV40 Enhancer
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	A unique identifier will be associated with each gene element.
Business Rule(s)	Implicit from the Gene element ID from the applicable Gene element controlled vocabulary.

4.7.11.4 Reference source

The reference source and related information will be captured according to 4.6. One or more reference sources could be cited.

4.7.12 Substance relationship

Relationships between substances such as the relationship between a salt form and its active moiety or parent substance should be captured. These relationships are often essential to the description of medicinal products, the basis of strength and the classification of substances. These relationships are often obvious and rules will be developed for specifying substances involved in each type of relationship. For example the active moiety of all sodium salts would be the free acid, conversely the active moiety of a hydrochloride salt would be a free base.

4.7.12.1 Relationship

The relationship is not a mere code but the actual connection between the two substances which may be different for different kinds of relationships:

There are a variety of different relationships that are possible but for the identification of medicinal products the most important relationships are between a substance, its active moieties, the basis of strength and constituents

in herbal and other complex substances. Parent and salt-solvate relationships and relationship of drug substance to its metabolites and impurities are also important although not essential for the description of a medicinal product. If there are known relationships between substances the relationships should be provided and maintained. **The parent-active moiety relationship** may be essential in describing medicinal products.

Table 44 — Relationship

User Guidance	Relationship between substances. NOTE: Target and gene elements are not to be specified in this section.
Example(s)	parent, active moiety, salt, radiolabelled, prodrug, metabolite, enantiomer
Conformance	CONDITIONAL
Data Type	CD

4.7.12.1 Interaction type

Table 45 — Interaction type

User Guidance	The Type of interaction between the substances.
Conformance	OPTIONAL
Data Type	CD

4.7.12.2 Related substance ID

Table 46 — Related substance ID

User Guidance	Substance ID associated with the related substance. All related substances should have a Substance ID.
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	ISO 11238 Substance ID

4.7.12.3 Related substance name

Table 47 — Related substance name

User Guidance	Preferred term or primary name of related substance.
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	Based on ISO 11238 substance name
Business Rule(s)	The preferred term of the related substance is captured; this information is required when there is a "Substance Relationship"; The substance name is implicit and derived from the Substance ID.

4.7.12.3.1 Amount type

Table 48 — Amount type

User Guidance	Most elements that require a quantitative value will also have a field called amount type. Amount type should always be specified because the actual value of the amount is often dependent on it. EXAMPLE: In capturing the actual relative amounts of substances or molecular fragments it is essential to indicate whether the amount refers to a mole ratio or weight ratio. For any given element an effort should be made to use same the amount type for all related definitional elements.
Example(s)	Mole percent; Weight percent; Mole ratio; Weight ratio, Weight, Moles
Conformance	CONDITIONAL
Data Type	CD

4.7.12.4 Amount

For some substance relationships it may be desirable to capture a quantitative or semi-quantitative amount that may further define the relationship (i.e. major metabolite, degradants etc.). Amount shall be captured according to 4.9.

Substance Relationships which refer to reactions (e.g. metabolism, prodrug) are quantified by specifying the quantity of the reactants:

4.7.12.4.1 Reference source

The reference source will be captured according to 4.6.

4.8 Structure (repeat as necessary) (Conditional)

Structural information is an essential element for all chemical substances and for other types of substances that have structurally definable elements or modifications. The representation should contain structural information in one or more of the standardised formats as indicated below. A graphical image of the molecular structure should also be provided if available. A complete representation of the structure should be provided and is usually sufficient to define chemical substances. Structural representations can be in a variety of formats or even multiple formats. The commonly used formats include molfile, SMILES, InChI, and cdx. Indication of a preferred structural representation may be defined at regional level. Although most chemical substances may be represented by a single unambiguous structural representation, there are also many substances that can be represented by multiple structural representations or as a mixture of single substances. For the initial implementation and the transmittal of data for substance definitions, an approach that limits submissions to a single or a limited number of unambiguous structural representations is recommended. There are several examples below that illustrate the process in choosing a defining structural representation to submit. A more detailed structure representation guide is provided as part of the annex B (Chemical substance).

For Amine Salts, with the exception of ammonium and quarternium, salts shall be represented as a free base.

EXAMPLE: DIFLOXACIN HYDROCHLORIDE

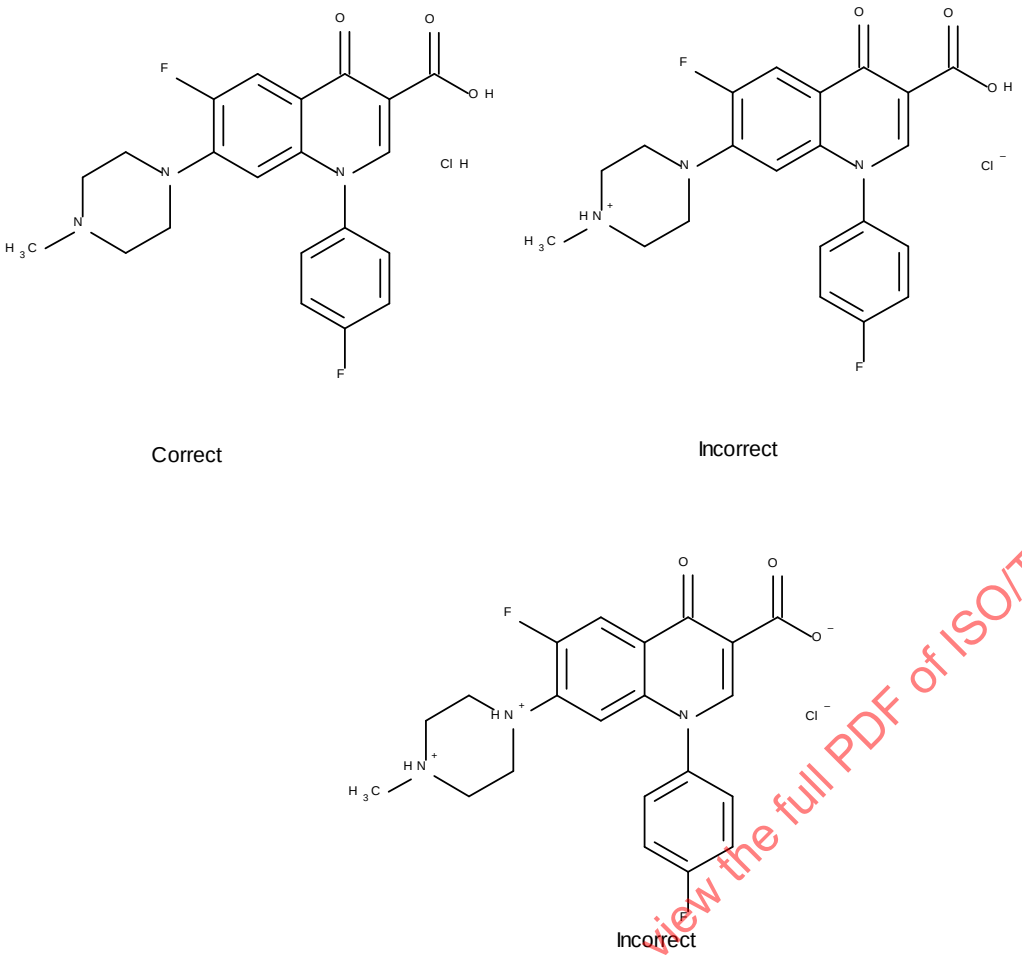


Figure 8 — Correct and incorrect graphical images of the molecular structure of difloxacin hydrochloride

Metal and ammonium salts: All metal and ammonium salts (NH4+) shall be represented in a charged configuration with all equivalent functional groups ionised. Charge balance will be achieved when necessary by adding hydrogen ions (protons).

EXAMPLE: MONOSODIUM CITRATE

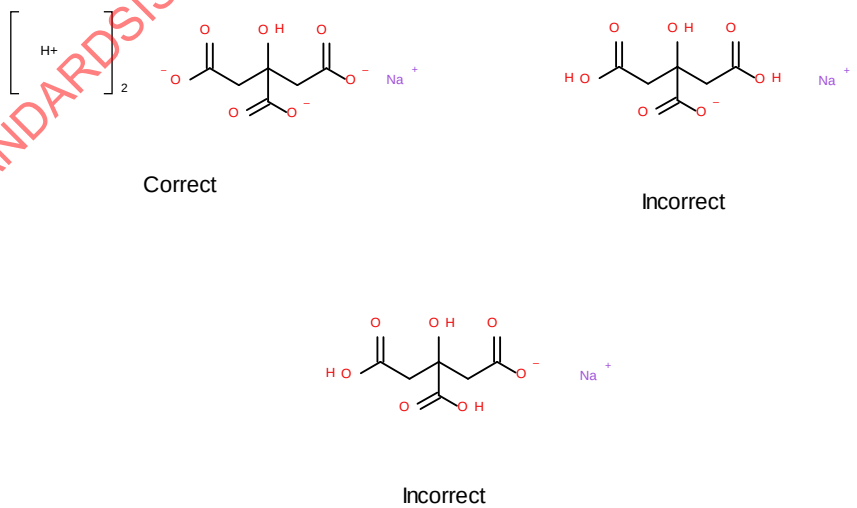


Figure 9 — Correct and incorrect graphical images of the molecular structure of monosodium citrate

Racemic mixtures can be represented as mixtures of enantiomers or by a single structural representation. When using a single structural representation and the substance only contains one chiral centre, a single structure will be drawn that does not contain a chiral bond. If more than one chiral centre exists, relative stereochemistry shall be indicated. The stereochemistry of each moiety of a given substance will be captured separately (4.9).

EXAMPLE: OFLOXACIN

A racemic mixture of the R and S isomers. A single structural representation can be used in this case.

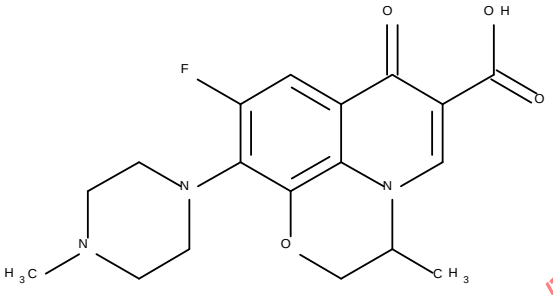


Figure 10 — Graphical image of the molecular structure of difloxacin hydrochloride

There are also substances that are mixtures of multiple enantiomers. These substances can also be transmitted in a single structural representation.

Substances can exist in interconverting forms or with a different defined structure depending on the physical state. In cases where the actual structure is a single configuration in a solid physical state, that structure should be captured. When a substance exists in a liquid state and can interconvert to a variety of structures, the least ambiguous single structure should be chosen.

EXAMPLE DEXTROSE (GLUCOSE): Exists as three separate substances in the crystalline state. For a crystalline substance one of the following representations should be chosen.

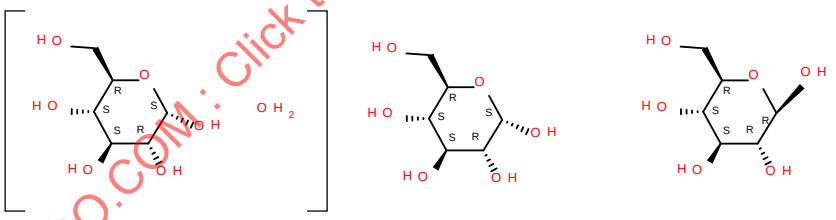
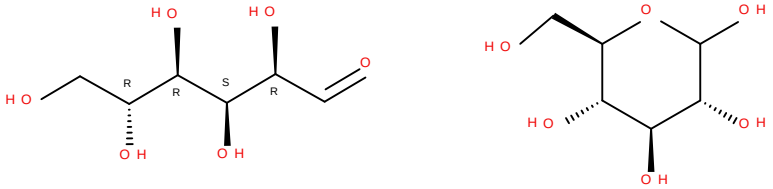


Figure 11 — Graphical images of the molecular structure of the three crystalline states of dextrose

In the liquid state dextrose and many other monosaccharides, disaccharides and higher homologs that end in an aldehyde are often referred to as reducing sugars and exist as mixtures of interconverting substances as illustrated in the structures of glucose below.

In this case the linear single representative structure is chosen to represent dextrose in the liquid or amorphous state. If pyranose or furanose forms were chosen the representative structure would have one site of stereochemical ambiguity. One general rule is that if there are multiple interconverting structural representations for a substance the preferred representation is the one with the least ambiguity. In figure 12 the pyranose and furanose structures have one site of stereochemical ambiguity.



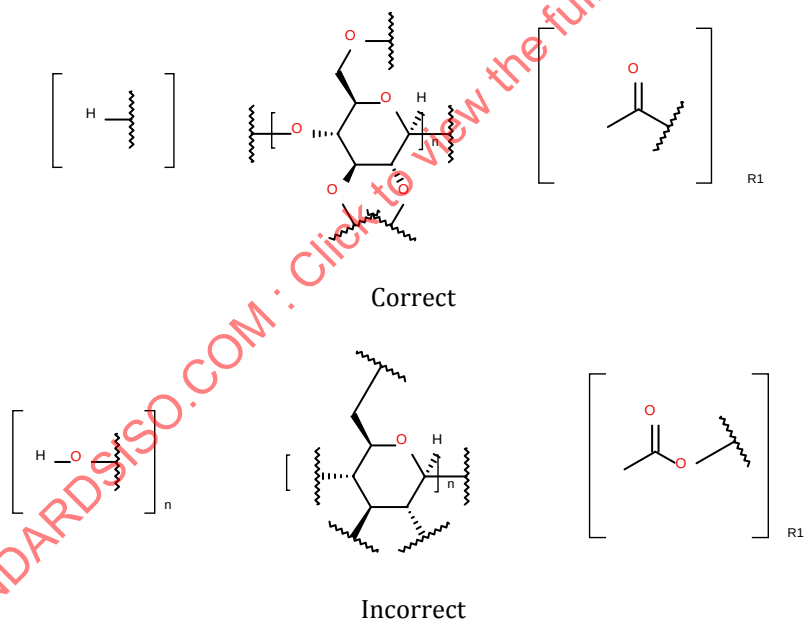
Correct Incorrect

Figure 12 — Correct and incorrect graphical images of the molecular structure of (liquid) dextrose

In this case the linear single representative structure is chosen to represent dextrose in the liquid or amorphous state. If pyranose or furanose forms were chosen the representative structure would have one site of stereochemical ambiguity. One general rule is that if there are multiple interconverting structural representations for a substance the preferred representation is the one with the least ambiguity. In the example below the pyranose and furanose structures have one site of stereochemical ambiguity.

Structural representations will also be used to represent fragments of molecules. Fragments will be used in describing partially modified polymers and are typically generated to describe partially acylated polymers, proteins or even nucleic acids. In order to consistently describe substances a fragment will be generated in a defined manner. Ester and amides will always be broke such the oxygen or nitrogen with stays with the alcohol or amine. Ethers will be broken such that the oxygen stays with polymeric or larger structural unit.

EXAMPLE: Cellulose Acetate

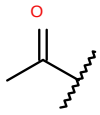


Correct

Incorrect

Figure 13 — Correct and incorrect graphical images of the molecular structure of cellulose acetate

The acetyl fragment is represented in Figure 14.

Correct: 

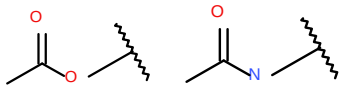
Incorrect: 

Figure 14 — Correct and incorrect graphical images of the molecular structure of the acetyl fragment

In fragments derived from ethers, the oxygen will be retained by the polymer or largest portion of a molecule.

EXAMPLE: Hypromelloses

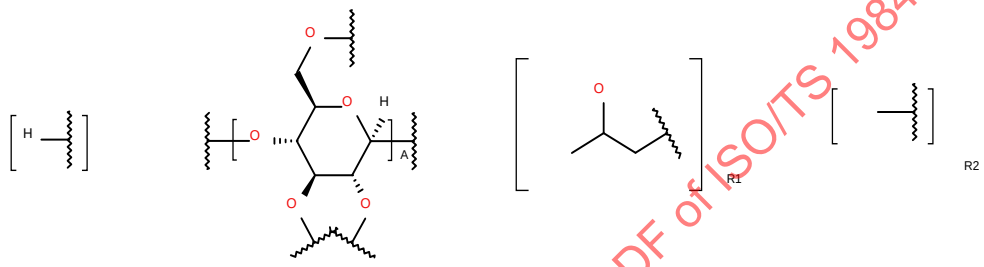


Figure 15 — Graphical images of the molecular structure of hypromelloses

Figure 16 provides the high level information for the Class Structure.

In the Class Structure the element group molecular formula is represented by molecular formula by moieties and by the total of moieties in accordance with the Hill system captured by the term molecular formula. In addition the element group molecular weight is connected to the Class Structure also for stoichiometric substances. This addition will be proposed in the next version of the ISO 1238 Standard.

See also the high level information at the substance type chemical.

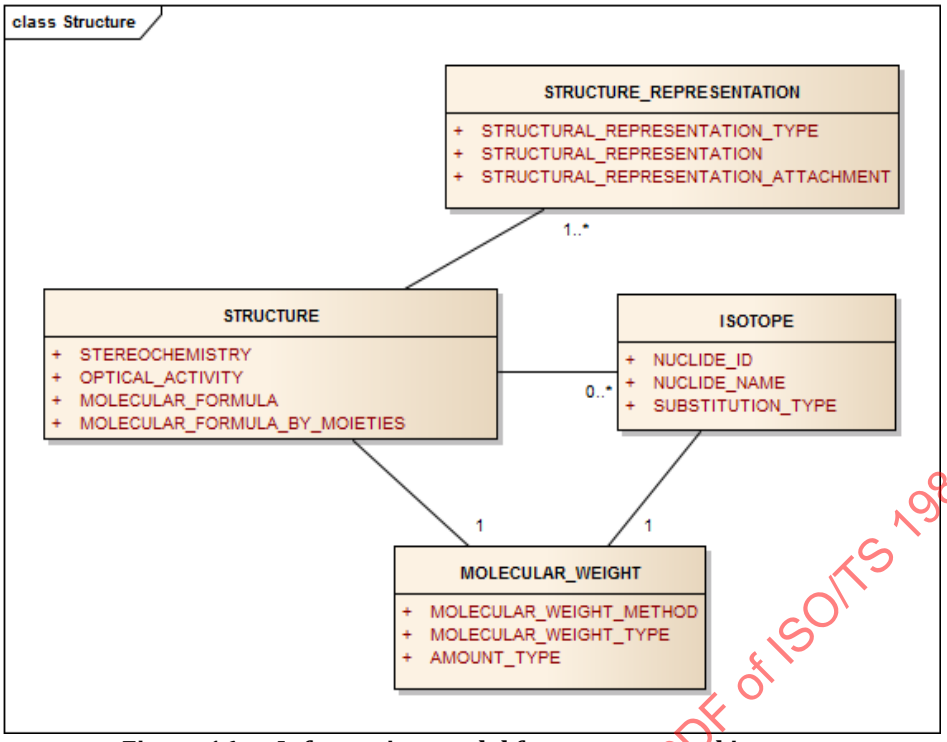


Figure 16 — Information model for structure and isotope

A structural representation is an essential element for defining chemical substances and is often sufficient for defining most chemical substances.

Four common structural representation types are listed below. Each of the types can unambiguously describe substances with a definite molecular structure. Stereochemistry should be completely defined in each structural representation. Only MOLFILES and CDX have the ability to describe polymeric repeating units and molecular fragments. There are a number of tools available, both free and commercial, that allow the creation of molecular structural representations. Multiple structural representation types can be submitted.

This class is CONDITIONAL. When the Substance type (as defined in 4.3) is either chemical, or polymer this class becomes MANDATORY and the following data elements should be described:

4.8.1 Structural representation type

Table 49 — Structural representation type

User Guidance	This field indicates the type of structure. A FULL structure is a structure in which the complete connectivity of the substance is known and the substance is monodisperse. A PARTIAL structure is a structure in which either complete connectivity or stereochemistry is not defined. A REPRESENTATIVE structure is used when the connectivity of the underlying material is diverse and a single structure is needed to represent that material
Example(s)	Full, partial, representative, fragment, ionic moiety, structural repeating unit
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	A structure can only be of one type. Field is mandatory when structural elements are present. When the Substance type (as defined in 4.3) is either chemical, protein, polymer or nucleic acid, this class becomes MANDATORY

Structural representation type is specified as a MIME Media Type code:

Type	MIME Media Type
InChI	application/x-InChI
MOLFILE	application/x-mdl-molfile
SMILES	application/x-smiles
CHIME	application/x-mdl-chime

4.8.2 Structural representation

Table 50 — Structural representation

User Guidance	The structural representation shall be provided as text string, the following formats are acceptable: InChI, SMILES, MOLFILE, CDX.
Example(s)	InChI, MOLFILE SMILES, CDX
Conformance	CONDITIONAL
Data Type	ST
Business Rule(s)	Each structural representation will have one and only one type. When the Substance type (as defined in 4.3) is either chemical, protein, polymer or nucleic acid, this class becomes MANDATORY

4.8.3 Structural representation attachment

Table 51 — Structural representation attachment

User Guidance	An attached file that may contain the structural representation shall be provided for complex data formats, e.g. CDX, MOLFILE.
Conformance	OPTIONAL
Data Type	ED

4.8.4 Stereochemistry

The overall stereochemistry of the substance should be indicated in this field.

Table 52 — Stereochemistry

User Guidance	The stereochemistry of the substance shall be indicated in the structure and should be captured here. Special cases of stereochemistry that can't be indicated in the structure shall be described based on a controlled vocabulary. Mixtures of stereoisomers shall be represented explicitly as a mixture of substances with absolute stereochemistry. In case the absolute stereochemistry is unknown the substance definition should be marked as "Incomplete".
Example(s)	AXIAL R, AXIAL S, SQUARE PLANAR 1, SQUARE PLANAR 2, SQUARE PLANAR 3, SQUARE PLANAR 4, TETRAHEDRAL, OCTAHEDRAL 12, OCTAHEDRAL 22, OCTAHEDRAL 21, CHIRAL, ACHIRAL
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	AXIAL R, AXIAL S, SQUARE PLANAR 1, SQUARE PLANAR 2, SQUARE PLANAR 3, SQUARE PLANAR 4, TETRAHEDRAL, OCTAHEDRAL 12, OCTAHEDRAL 22, OCTAHEDRAL 21; RACEMIC, MIXED, CHIRAL, ACHIRAL, ABSOLUTE, AXIAL, EPIMERIC, MESO, UNKNOWN, CIS, TRANS
Business Rule(s)	When the Substance type (as defined in 4.3) is either chemical, protein, polymer or nucleic acid, this class becomes MANDATORY

4.8.5 Optical activity

The optical activity of a chiral substance should be captured if known. The extent of optical rotation is not mandatory at the substance level.

Table 53 — Optical activity

User Guidance	The optical activity shall be described based on a controlled vocabulary.
Example(s)	(+/-), (+), (-)
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	(+/-), (+), (-)
Business Rule(s)	Shall be entered if known for substances that have at least one moiety with stereochemistry defined as chiral. Can be a defining element when the absolute stereo configuration is not known.

4.8.6 Molecular Formula

Table 54 — Molecular formula

User Guidance	Molecular formula is written as a sequence of pairs of atom symbol and stoichiometric number without any text formatting or special characters. This formal notation is essentially a code system. Specified according to the Hill system, i.e. first C, then H, then alphabetical.
Example(s)	$C_{26}H_{33}ClN_2O_9S$
Conformance	CONDITIONAL
Data Type	ST
Business Rule(s)	Can be implicitly derived from the structure; Validity check for identity and structural representation

4.8.7 Molecular Formula by Moieties (new class to be included in the second edition of ISO 11238)

Table 55 — Molecular formula by moieties

User Guidance	Specified per moiety according to the Hill system, i.e. first C, then H, then alphabetical. and each moiety separated by a dot.
Example(s)	$C_{20}H_{25}ClN_2O_5 \cdot C_6H_6O_3S \cdot H_2O$
Conformance	CONDITIONAL
Data Type	ST
Business Rule(s)	Can be implicitly derived from the structure; Validity check for identity and structural representation

4.8.8 Isotope (repeat as necessary)

Applicable for single substances that contain a radionuclide or a non-natural isotopic ratio, e.g. C-13 enriched material. All radionuclide and non-natural isotopes will also be represented in the structure representation for chemical substances. For other types of substances a structural representation could be created to indicate the position of substitution.

4.8.8.1 Nuclide ID

Table 56 — Nuclide ID

User Guidance	A Substance ID for each non-natural or radioisotope is created. The ID is linked to a single atom zero valence element.
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	ISO 11238 Substance ID
Business Rule(s)	Mandatory when substance contains non-natural or radioisotope.

4.8.8.2 Nuclide name

Table 57 — Nuclide name

User Guidance	The name for each isotope shall be provided.
Example(s)	CARBON C-13
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	ISO 11238 Substance names
Business Rule(s)	Mandatory when substance contains non-natural or radioisotope. Implicitly derived from Nuclide ID.

4.8.8.3 Substitution type

Applicable for single substances that contain a radionuclide or a non-natural isotopic ratio, e.g. C-13 enriched material. All radionuclide and non-natural isotopes will also be represented in the structure representation.

Table 58 — Substitution type

User Guidance	The type of isotopic substitution present in a single substance: - specific if the site of attachment/substitution indicated in structure; - non-specific if the nuclide is distributed throughout molecule or substance; - Unknown if site unknown. Substitution refers to the relationship between the nuclide and the rest of the substance.
Example(s)	specific; non-specific; unknown
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	Mandatory when substance contains non-natural or radioisotope.

4.9 Amount (Conditional)

The amount section serves to provide the quantitative or qualitative values that are associated with a variety of elements. The same format will be used for all quantitative, semi-quantitative and qualitative values. This amount section will be used throughout the standard for any properties or elements where quantitative, semi-quantitative or qualitative values apply. An amount type will usually be associated with each amount and the same amount type will typically apply to all of the quantitative values in a given section (i.e. for multiple fragment substitution either mole percent, weight percent or degree of substitution should be applied to each fragment they should not be mixed).

4.9.1 Average

Table 59 — Average

User Guidance	Used to capture quantitative values for a variety of elements. If only limits are given, the arithmetic mean would be the average. If only a single definite value for a given element is given, it would be captured in this field.
Example(s)	Average, Exact, Numeric Value
Conformance	CONDITIONAL
Data Type	PQ
Values Allowed	Single quantitative values or an average value for a given element.

4.9.2 Low limit

It should be noted that there can be two types of low limit, one is a limit on actual values and the other is the lower limit of the average of values. Both these limits can be captured and type of lower limit should be indicated either in the amount type for a given element or in the units indicated for the amount. The low limit can be used alone to express one-sided ranges greater than or equal to this limit.

Table 60 — Low limit

User Guidance	Lower limit for a given quantitative value.
Conformance	CONDITIONAL
Data Type	PQ
Business Rule(s)	To be used to express the lower limit of a range of values, if applicable

4.9.3 High limit

Similar to the low limit, the high limit should indicate whether the value refers to a range of values or a range of averages. The high limit can also be used to express one-sided limits.

Table 61 — High limit

User Guidance	Higher limit for a given quantitative value.
Conformance	CONDITIONAL
Data Type	PQ
Business Rule(s)	To be used to express the higher limit of a range of values

4.9.4 Unit

Most units should be consistent with those described in UCUM. There may be a need for additional units to describe particular properties and the maintenance organisation should maintain them. Units may be dependent on the method of measurement or calculation to determine the property.

EXAMPLE The hydrophobic-lipophilic balance can be a defining property for amphiphilic polymers particularly when the exact ratio of monomers or structural repeating units is not provided. There are at least two methods of calculation for this property and the method may be indicated in the unit (i.e. Davies Method; Griffin's Method).

Table 62 — Unit

User Guidance	The unit of measure
Example(s)	1, g, mg, mol, mmol, L, [IU], mg/L, mol/L, g/mol, [IU]/g
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	UCUM / ISO 11240 units
Business Rule(s)	The Unit is mandatory for amount expressed as numeric value. When used with limits, both limits must have the same dimension (and usually will have the same unit.)

4.9.5 Non-numeric Value

Table 63 — Non-numeric value

User Guidance	Qualitative, coded property value. To be used to capture qualitative values.
Example(s)	Positive, negative, Yes, No, complete, partial
Conformance	CONDITIONAL
Data Type	CD

4.10 Source material (Conditional)

Source material shall capture information on the taxonomic and anatomical origins as well as the fraction of a material that can result in or can be modified to form a substance. This set of data elements shall be used to define polymer substances isolated from biological matrices. Taxonomic and anatomical origins shall be described using

controlled vocabulary as required. This information is captured for naturally derived polymers (e.g. starch) and structurally diverse substances. For Organism belonging to the Kindom Plantae the Substance is defined referring to the fresh material of a single species or infraspecies. For (herbal) substance the fraction information will be captured at the Substance information and additional information for herbal extracts will be captured at the specified substance Group 1 information level. See for further explanation the Substance Class: Structurally Diverse and the herbal annex.

The information model for the source material is shown in Figure 17.

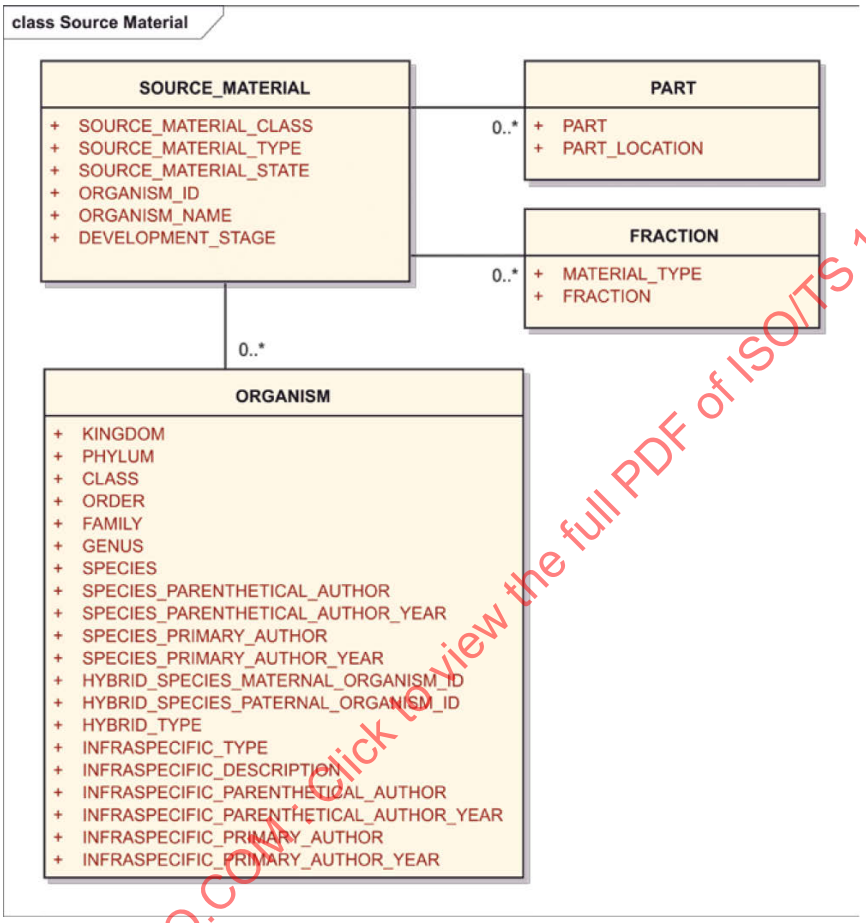


Figure 17 — Information model for source material

Specifically, the information on the source material shall be provided by means of the following set of data elements.

4.10.1 Source material class

Table 64 — Source material class

User Guidance	General high level classification of the source material specific to the origin of the material.
Example(s)	MINERAL, BIOLOGIC
Conformance	CONDITIONAL
Data Type	CD

4.10.2 Source material type

Table 65 — Source material type

User Guidance	The type of the source material shall be specified based on a controlled vocabulary. For vaccines this section refers to the class of infectious agent.
Example(s)	bacterium, human, mammal, fungus, virus, plant
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	The "Source_Material_Type" values are presented based on the "source material class" values above. Because organic / inorganic is just a class best captured in terminology of the source material types, only one data element is needed for Type which implies Class.

4.10.3 Source material state

Table 66 — Source material state

User Guidance	The state of the source material when extracted.
Example(s)	Live, activated, inactivated, attenuated, conjugated, live (attenuated).
Conformance	CONDITIONAL
Data Type	CD

4.10.4 Organism ID

Table 67 — Organism ID

User Guidance	The unique identifier associated with the source material parent organism shall be specified.
Conformance	CONDITIONAL
Data Type	II

4.10.5 Organism name

Table 68 — Organism name

User Guidance	The organism accepted Latin name shall be provided based on the binomial Latin nomenclature.
Example(s)	*Ginkgo biloba* L.
Conformance	CONDITIONAL
Data Type	CD

4.10.6 Development stage

Table 69 — Development stage

User Guidance	Stage of life for animals, plants, insects and microorganisms. This information shall be provided only when the substance is significantly different in these stages (e.g. foetal bovine serum).
Example(s)	foetus, infant, juvenile, adult, senescent, leafing, pre-flowering, flowering, fruiting, etc.
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	If it is a distinguishing factor, even if it is implicitly understood by experts.

4.10.7 Part description (repeat as necessary)

Information of the part of the material used to produce the substance shall be provided by means of the following data elements:

4.10.7.1 Part

Table 70 — Part

User Guidance	The portion of an organism with a definable anatomical location.
Example(s)	Cartilage, Root and Stolon
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	Depending on the Source_Material_Type values (e.g. human/mammalian or plant) the relevant controlled vocabulary applies (e.g. organism or plant).

4.10.7.2 Part location

Table 71 — Part location

User Guidance	The detailed anatomic location when the part can be extracted from different anatomical locations of the organism. Multiple alternative locations may apply.
Example(s)	For cartilage: knee, elbow , stomach, shoulder
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	Applicable only if Source_Material_Type is human or mammalian

4.10.8 Fraction (repeat as necessary)

Many complex materials are fractions of parts of plants, animals, or minerals. Fraction elements are often necessary to define both Substances and specified substances Group 1. For substances derived from Plants, fraction information will be captured at the Substance information level. Additional information for Extracts, such as extraction solvent composition, will be captured at the specified substance Group 1 information level. For other types of substances fractions will be captured at the substance level (e.g. blood products).

4.10.8.1 Material type

Table 72 — Material type

User Guidance	The specific type of the material constituting the component.
Example(s)	LIPIDS; PHOSPHOLIPIDS; PROTEINS; NUCLEIC ACID
Conformance	CONDITIONAL
Data Type	CD

4.10.8.2 Fraction

A controlled vocabulary for specific fractions may be developed.

Table 73 — Fraction

Example(s)	OILS; JUICE; POLYCLONAL ANTIBODIES; SERUM; PLASMA
Conformance	CONDITIONAL
Data Type	CD

4.10.9 Organism

This subclause describes the organism which the substance is derived from. For vaccines, the parent organism shall be specified based on these section elements. As and Example full taxonomy will be described for the Substance Name: *Ginkgo biloba* L., Leaf

Organism Taxonomy

A significant challenge in ISO 11238 implementation surrounds the concept of organism because varying opinions exist among taxonomic experts about how individuals should be grouped. This is often referred to as the species problem. The species problem is a mixture of difficult related questions that often come up when biologists define the word "species". Definitions are usually based on how individual organisms reproduce and which other individuals they are capable of reproducing with to produce fertile offspring. This definition works well for vertebrates and most invertebrates but is less useful for plants and useless for bacteria, viruses, and many protists. The practical approach to “species” that allows implementation of the standard is that each “species” is a kind of organism and each “species” is based on a set of characteristics that are shared by all the organisms in the “species”.

This common usage of "species" is as a category, type, or natural kind. These “species” share common morphological, genetic and signature constituent characteristics. Traditional taxonomy may not be able to provide sufficient type or kind distinctions so other characteristics may be required to properly differentiate organism groups. From this point of view, bitter fennel is distinct from sweet fennel based on consistently different chemical profiles even though they are only marginally recognized in taxonomic trees as distinct. The same is true for hot peppers versus sweet peppers and collards versus kale. Similarly, microorganism host and serotype distinctions are often important and necessary. Detailed strain information is also necessary to describe viruses because their genomes change rapidly but the virus strains should be grouped at the species or serotype level.

The standard recognizes the traditional seven levels of taxonomy as well as several intraspecific types and hybrids. While these are all useful grouping concepts, they have not been implemented consistently for the entire taxonomic tree. The general approach to management of organism groupings is to defer to scientific taxonomy experts. These experts review the scientific literature and determine the appropriate grouping concepts. These concepts are managed as accepted or valid scientific names. For higher plants and fungi, Kew Gardens has become the primary providers of this expert opinion. That information is available publicly through the World Catalogue of Plant Species and Index Fungorum as represented publicly in the Medical Plants Name Service (gold standard

data, http://apps.kew.org/mpns-portal/), the Plant List (includes a zero to three star data quality rating http://www.theplantlist.org/) and Species Fungorum (http://www.speciesfungorum.org/). Taxonomic information for most other organisms is available through Catalogue of Life (COL, currently 1,600,000+ organsims, http://www.catalogueoflife.org/col/) and its 151 contributing source databases. The most notable contributers are Kew Gardens and the Integrated Taxonomic Information System (www.itis.gov). COL has integrated the source datasets into a common seven level taxonomic tree using an eight kingdom model and includes intraspecifics where relevant. For organisms omitted in the above sources, NCBI taxonomy (http://www.ncbi.nlm.nih.gov/taxonomy) is the most useful source database.

The higher levels of taxonomy are useful but all levels above family are not required for organism definition. The primary definition of organism is the combination of genus and species with author and family in that order providing validation. Intraspecific type and description and hybrid characteristics are also definitional where appropriate. The organism should be described as whole with no other characteristics. Because taxonomy is continuously changing, new organism identifiers should only be generated after analyzing potential collisions with existing identifiers using the above public datasets and/or conferring with appropriate taxonomic experts. Developmental stage, part and viral strain characteristics should be associated with separate identifiers that refer back to the whole organism parent substance.

4.10.9.1 Kingdom

Table 74 — Kingdom

User Guidance	The kingdom of an organism shall be specified.
Example(s)	Plantae
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	This is derived from the species and is referenced from an appropriate terminology / taxonomy (e.g. the NCBI Entrez Taxonomy ID implies all higher levels.)

4.10.9.2 Phylum

Table 75 — Phylum

User Guidance	The phylum of an organism shall be specified.
Example(s)	Tracheophyta
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	This is derived from the species and is referenced from an appropriate terminology / taxonomy (e.g. the NCBI Entrez Taxonomy ID implies all higher levels.)

4.10.9.3 Class

Table 76 — Class

User Guidance	The class of an organism shall be specified.
Example(s)	Mammalia, Maxillopoda, Sauropsida. Ginkgoopsida
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	This is implied by the species and is referenced from an appropriate terminology / taxonomy (e.g. the NCBI Entrez Taxonomy ID implies all higher levels.)

4.10.9.4 Order

Table 77 — Order

User Guidance	The order of an organism shall be specified,
Example(s)	Ginkgoales
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	This is implied by the species and is referenced from an appropriate terminology / taxonomy (e.g. the NCBI Entrez Taxonomy ID implies all higher levels.)

4.10.9.5 Family

Table 78 — Family

User Guidance	The family of an organism shall be specified.
Example(s)	Ginkgoaceae
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	This is implied by the species and is referenced from an appropriate terminology / taxonomy (e.g. the NCBI Entrez Taxonomy ID implies all higher levels.)

4.10.9.6 Genus

Table 79 — Genus

User Guidance	The genus of an organism shall be specified; refers to the Latin epithet of the genus of the plant/animal; it is present in names for genera, species and infraspecies.
Example(s)	Ginkgo
Conformance	CONDITIONAL
Data Type	CD

4.10.9.7 Species

Table 80 — Species

User Guidance	The species of an organism shall be specified; refers to the Latin epithet of the species of the plant/animal; it is present in names for species and infraspecies.
Example(s)	Biloba
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	Must be specified as a code from an appropriate taxonomy terminology (e.g. NCBI Taxonomy ID)

4.10.9.7.1 Species literature reference

Species are described by original articles, and the taxonomy may vary over time, so the reference to the original article is critical. A reference to an article should be complete and not just use author name abbreviations.

The structure is the same as reference source, with the addition of adding structured authors and years:

The reference source will be captured according to 4.6.

4.10.9.7.2 Species parenthetical author

The species parenthetical author should be captured as reported. For herbal substances the author will be captured according to the MPNS of Kew Gardens.

Table 81 — Species parenthetical author

User Guidance	The parenthetical author of an organism species shall be specified; refers to the first author who published the plant/animal name (of any rank).
Example(s)	L.
Conformance	CONDITIONAL
Data Type	ST

4.10.9.7.3 Species parenthetical author year

Table 82 — Species parenthetical author year

User Guidance	The parenthetical author year of an organism species an organism shall be specified; refers to the year in which the first author published the plant/animal name (of any rank).
Example(s)	2010, 2011
Conformance	OPTIONAL
Data Type	ST

4.10.9.7.4 Species primary author

Table 83 — Species primary author

User Guidance	The primary author of an organism species shall be specified; refers to the first author, who published the plant/animal name (of any rank).
Example(s)	Sch.Bip
Conformance	OPTIONAL
Data Type	ST

4.10.9.7.5 Species primary author year

Table 84 — Species primary author year

User Guidance	The primary author year of an organism species shall be specified; refers to the year in which the first author published the plant/animal name (of any rank).
Example(s)	2010
Conformance	OPTIONAL
Data Type	ST

4.10.9.7.6 Hybrid species maternal organism ID

Table 85 — Hybrid species maternal organism ID

User Guidance	The identifier of the maternal species constituting the hybrid organism shall be specified based on a controlled vocabulary.
Conformance	CONDITIONAL
Data Type	II

4.10.9.7.7 Hybrid species paternal organism ID

Table 86 — Hybrid species paternal organism ID

User Guidance	The identifier of the paternal species constituting the hybrid organism shall be specified based on a controlled vocabulary.
Conformance	CONDITIONAL
Data Type	II

4.10.9.7.8 Hybrid type

Table 87 — Hybrid type

User Guidance	The hybrid type of an organism shall be defined.
Example(s)	Yes, No
Conformance	OPTIONAL
Data Type	BL
Business Rule(s)	If the hybridisation step is present, the organism is a hybrid.

4.10.9.7.9 Infraspecific type

Table 88 — Infraspecific type

User Guidance	The infraspecific type of an organism shall be specified.
Example(s)	cultivar, variety, serotype, strain, subspecies; cell line; year of isolation
Conformance	OPTIONAL
Data Type	CD

4.10.9.7.10 Infraspecific description

Infraspecific information will be used to describe strains, serotypes, varieties, cultivars and cell lines. Often there is not a naming or taxonomic authority for this information. Standardised formats for this information may be developed. The actual infraspecific information should also be captured in the naming of these substances and synonyms should be maintained to help to identify infraspecific information and avoid duplication.

Table 89 — Infraspecific description

User Guidance	The infraspecific description of an organism shall be specified based on a controlled vocabulary. For Influenza Vaccine, the infraspecific description shall contain the syntax of the antigen in line with the WHO convention.
Example(s)	A/BRISBANE/59/2007(H1N1)
Conformance	CONDITIONAL
Data Type	CD

4.10.9.7.11 Infraspecific literature reference

The reference source will be captured according to 4.6.

4.10.9.7.12 Infraspecific parenthetical author

Table 90 — Infraspecific parenthetical author

User Guidance	The infraspecific parenthetical author of an organism shall be specified; refers to the first author who published the infraspecific plant/animal name (of any rank).
Conformance	OPTIONAL
Data Type	ST

4.10.9.7.13 Infraspecific parenthetical author year

Table 91 — Infraspecific parenthetical author year

User Guidance	The infraspecific parenthetical author year of an organism shall be specified; refers to the year in which the first author published the infraspecific plant/animal name (of any rank).
Example(s)	2011
Conformance	OPTIONAL
Data Type	ST

4.10.9.7.14 Infraspecific primary author

Table 92 — Infraspecific primary author

User Guidance	The infraspecific primary author of an organism shall be specified; refers to the first author who published the infraspecific plant/animal name (of any infraspecific rank).
Conformance	OPTIONAL
Data Type	ST

4.10.9.7.15 Infraspecific primary author year

Table 93 — Infraspecific primary author year

User Guidance	The infraspecific primary author year of an organism shall be specified; refers to the year in which the first author published the infraspecific plant/animal name (of any infraspecific rank).
Example(s)	2009
Conformance	OPTIONAL
Data Type	ST

4.11 Modification (repeat as necessary) (Conditional)

The Modification section is to be used to describe irreversible modifications to material (e.g. PEGylation, phosphorylation, hydrogenation). The modifications may be physical, chemical, enzymatic, etc. Modifications may be described by their structural result (substitution of moieties to residues, etc.) or by the process, reagents, or processing time if a specific structural modification cannot be determined (e.g. aggregated albumin).

A minimal description of the modification process shall be generated when a definitive structural modification cannot be determined.

This section applies to:

- Nucleic Acids
- Proteins
- Polymer
- Structurally Diverse Substances
- Mixture
- specified substance Group 1

This section shall be repeatable to describe each modification occurred on each residue modified.

The information model for the class modification group is shown in Figure 18.

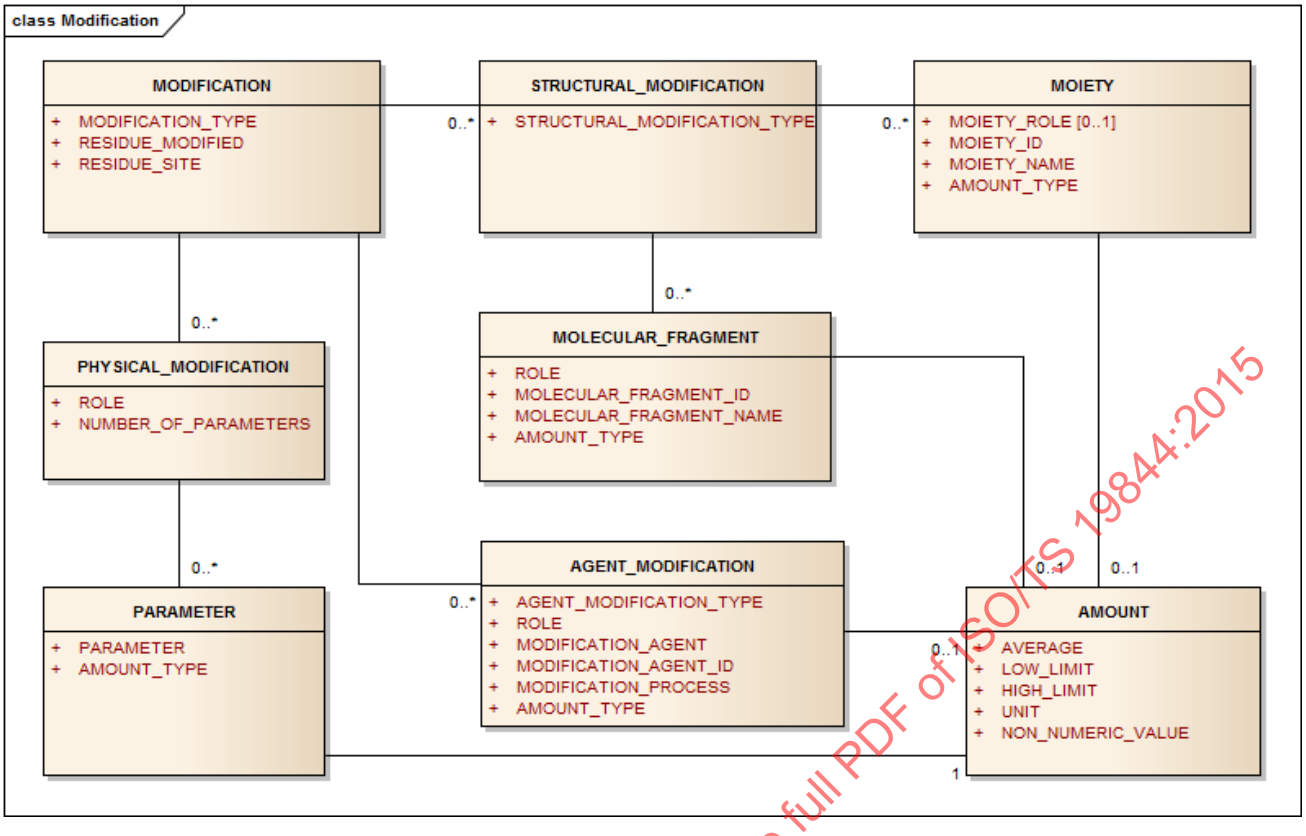


Figure 18 — Information model for modifications

Modifications will be divided into *structural modifications* where the existing molecular structure is discretely modified, *physical modifications* where no new covalent bonds are formed and *agent modifications* where the chemical modifications are diverse or unclear.

Specifically, information on the modification shall be provided by means of the following descriptors:

4.11.1 Modification type

The modification type becomes apparent by its description. Structural modification is described using the structural elements. Agent and physical modifications are described by a series of elements. Agents are also substances in their own right.

Table 94 — Modification type

User Guidance	The type of the modification should be specified. Modifications will be divided into structural modifications where the existing molecular structure is discretely modified, physical modifications where no new covalent bonds are formed and agent modifications where the chemical modifications are diverse or unclear.
Example(s)	structural modifications, agent modifications, moiety, Physical Modification
Conformance	CONDITIONAL
Data Type	CD

4.11.2 Residue modified

Table 95 — Residue modified

User Guidance	The amino acid or nucleic acid base that is modified. It may be known that lysine residues are modified but the particular lysine in the sequence may not be known. It should be captured for both specific and non-specific modifications.
Example(s)	Lysine
Conformance	CONDITIONAL
Data Type	CD

4.11.3 Residue site

Table 96 — Residue site

User Guidance	The position of specific residue undergone to modifications shall be described.
Example(s)	1_20 ; Protein unit 1, amino acid at position 20; C-Terminal removal
Conformance	CONDITIONAL
Data Type	ST

4.11.4 Structural modification

All structurally specific post-translational modifications are expressed as amino acid substitutions. In order to limit ambiguity, all structural substitutions will be represented with a single structural element that contains the amino acid, a linker if present and the conjugate as a single entity. The substituted amino acid is assumed to connect through the alpha amino and the carboxylic acid group. If there is potential ambiguity or if the connectivity is not through the alpha amino acid groups, connection points or atoms should be used. The substituted amino acid will be a substance in its own right. For informational purposes the linker and conjugate can also be indicated as separate elements. When the sites of modifications are known they should be listed. For a modification that is not site specific, the extent of modification of should be captured. New substances would be created if there are consistent differences in the extent of modification. For example a monoclonal antibody with an average of two toxins conjugated to it would be a different substance than the same monoclonal antibody with an average of four toxin conjugates.

It should be noted that post-translational modifications, such as glycosylation, phosphorylation, or sulfation that result in microheterogeneity will not be described at the substance level, but could be described at the specified substance level. The microheterogeneity of proteins can be captured at the specified substance Group 1 the level and type of glycosylation is necessary to distinguish the material.

EXAMPLE Substitution of N terminal amino acid by pyroglutamic acid.

Structural modification is conditional and becomes mandatory when the type of modification as described in 4.11.1 refers to a structural modification, then the following specifications apply:

4.11.4.1 Structural modification type

Table 97 — Structural modification type

User Guidance	The Type of structural modification should be described.
Example(s)	Amino Acid Substitution
Conformance	CONDITIONAL
Data Type	CD

4.11.4.1.1 Moiety

For structural modifications, the moiety shall be provided based on the specifications provided in 4.10.

4.11.4.1.2 Amount type

Amount type should be captured according to 4.7.12.3.1.

4.11.4.1.3 Amount

Amount shall be captured as specified in 4.9.

4.11.4.1.4 Molecular fragment

For covalent modifications of proteins, nucleic acids, polymers or even structurally diverse material, fragments will be created and used to capture the extent of the modification.

4.11.4.1.4.1 Fragment role

For modified proteins covalent modifications will be captured as a fragment modification. For many protein modifications the fragment will typically be a modified amino acid.

Table 98 — Fragment role

User Guidance	The role of the fragment.
Example(s)	N-Terminal Pyroglu Formation
Conformance	CONDITIONAL
Data Type	CD

4.11.4.1.4.2 Molecular fragment ID

Each fragment will be identified with a Substance ID if available. If not available a new ID will be assigned with the submission.

Table 99 — Molecular fragment ID

User Guidance	The unique identifier assigned to the substance representing the fragment based on the ISO 11238 substance controlled vocabulary.
Conformance	MANDATORY
Data Type	CD
Values Allowed	ISO 11238 Substance ID

4.11.4.1.4.3 Molecular fragment name

Each fragment will also have at least one name associated with the fragment.

Table100 — Molecular fragment name

User Guidance	The name of the moiety shall be provided.
Example(s)	propylene glycol
Conformance	MANDATORY
Data Type	CD
Values Allowed	ISO 11238 substance name.
Business Rule(s)	ISO 11238 substance preferred name is selected as default value. The substance name is implicit and derived from the Moiety ID described in 5.1.5.2.2.

4.11.4.1.4.4 Amount type

Typically there will be an amount type relationship that indicates the degree of fragment substitution. Typically a single moiety will typically be assigned the value 1 and ranges for moieties that are stoichiometric will be relative to. The amount type should be captured according to 4.7.12.3.1.

4.11.4.1.4.4.1 Amount

The amount shall be captured as specified in 4.9.

4.11.4.2 Agent modification

This type describes modifications that do not result in the addition of a single well-defined chemical entity (i.e. formaldehyde, glutaraldehyde treatment, peroxide treatment). In the inactivation of vaccines or treatment of tissue matrices, agent modification is essential for the description of these materials. Agent modification is also used in describing the culturing of cells that result in phenotypic differentiation. The agent modification could also be used to capture the actual agent that reacts with a protein that results in the attachment of a well-defined chemical fragment or entity. The following descriptors would be used to define agent modifications.

4.11.4.2.1 Agent modification type

Table 101 — Agent modification type

User Guidance	Refers to the type of the agent caused the modification.
Example(s)	chemical, enzymatic, immunological, organism
Conformance	MANDATORY
Data Type	CD

4.11.4.2.2 Role

Table 102 — Role

User Guidance	For proteins, agent (a chemical that results in non-specific modifications of a protein) or moiety (a specific moiety added to a protein molecule).
Example(s)	antigen, linker, conjugate, inactivation, antigen priming
Conformance	MANDATORY
Data Type	CD

4.11.4.2.3 Modification agent

Table 103 — Modification agent

User Guidance	Established or primary name of the modifying agent.
Example(s)	Formaldehyde
Conformance	MANDATORY
Data Type	CD
Values Allowed	ISO 11238 substance name
Business Rule(s)	Implicitly derived from the value of the Modification Agent ID.

4.11.4.2.4 Modification agent ID

Table104 — Modification agent ID

User Guidance	Unique identifier for modifying agent.
Example(s)	UNII code
Conformance	CONDITIONAL
Data Type	II
Values Allowed	ISO 11238 Substance ID

4.11.4.2.5 Modification process

Table105 — Modification process

User Guidance	Refers to the description of the modification process.
Example(s)	cell culture, UV irradiation, etc.
Conformance	CONDITIONAL
Data Type	CD

4.11.4.2.6 Amount type

Amount type should be captured according to 4.7.12.3.1.

4.11.4.2.7 Amount

Amount shall be captured as specified in Section **Physical Modification.**

The physical modification is optional; when applicable the information shall be provided by means of the following descriptors:

4.11.4.2.8 Role

Table 106 — Role

User Guidance	To be used to describe the function of the modification.
Example(s)	inactivation, activation
Conformance	MANDATORY
Data Type	CD

4.11.4.2.9 Number of parameters

Table 107 — Number of parameters

User Guidance	The number of parameters that is going to be described as a result of the modification shall be provided.
Example(s)	For thermal denaturation two parameters, temperature and time, are typically captured.
Conformance	CONDITIONAL
Data Type	INT
Business Rule(s)	This number can be counted and is never explicitly represented. This value is calculated implicitly.

4.11.4.2.9.1 Parameters

Primarily used for nonspecific modification, specifies how a modification is quantified or extent of physical treatment.

4.11.4.2.9.1.1 Parameter

Table 108 — Parameter

User Guidance	Refers to the conditions under with the modification has been produced.
Example(s)	time, temperature, etc.
Conformance	CONDITIONAL
Data Type	ST
Business Rule(s)	Required parameters depend on process.

4.11.4.2.9.1.2 Amount type

Should be indicated per 4.7.12.3.1

4.11.4.2.9.1.3 Amount

The information related to the amount shall be provided as per specification described in 4.9.

4.12 Property (Conditional)

The section serves to provide information related to biological, physical or chemical characteristics associated with a substance. This information can be essential for the definition of each type of substance when structural elements are not sufficient to distinguish between similar substances.

EXAMPLE The pH of magnesium aluminometasilicate is necessary to distinguish low and high pH substances; viscosity is used to distinguish many polymers and should always be reported if known. At the specified substance Group 1 level, qualitative properties such as sterility are essential to distinguish materials from each other. Melting points and solubility that differ significantly for what is believed to be the same substance could either be indicative of purity or of polymorphism. Different polymorphs would be considered different specified substances at the specified substance Group 1 level.

Properties can have values that are quantitative, semi-quantitative or qualitative. For semi qualitative values a consistent scale should be used.

Property information is typically used for distinguishing substances or specified substances Group 1 from one another. It is possible however that a request for specific general property information that may be independent of defining information may be requested. Some of this information may be used to determine if two substances or specified substances are the same.

EXAMPLE For chemical substances; elemental analysis, mass and NMR spectra can be very helpful in determining whether two materials are different chemical substances and also in verifying that the molecular structure is correct. Solubility information, melting, boiling and critical points, partition coefficients (Log Kow), and pKa values should also be supplied if known.

The property information shall be provided by means of the following data elements:

4.12.1 Property type

Each property will be associated with a single property type and the list of property types should be maintained.

Table 109 — Property type

User Guidance	Type of the property for which the information is provided.
Example(s)	physical, chemical, enzymatic, immunological
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	Each property must be provided with a fully terminologically defined concept for measurements. Such definition would include any necessary "types" such as this.

4.12.2 Property name

Table 110 — Property name

User Guidance	The name of the property shall be specified based on a controlled vocabulary.
Example(s)	VISCOSITY, PH, CELL SURFACE ANTIGEN, SPECTRA
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	A fully defined coded concept must always be provided.

4.12.3 Property parameters (new class to be included in the second edition of ISO 11238)

Parameters are often associated with properties and should be captured. In order to compare or distinguish substances it is important that measurements are taken using the same parameters. Parameters are not elements that will typically be searchable, but standard formats for given property measurements may be defined.

Table 111 — Property parameters

User Guidance	A field that should be used to capture parameters that were used in the measurement of a property.
Example(s)	For Viscosity, solvent concentration and temperature should be captured as a single entity.
Conformance	CONDITIONAL
Data Type	ST

4.12.4 Substance name

Table 112 — Substance name

User Guidance	To be used to identify a substance related to a defining property, for example: cell surface antigens (the Substance ID for CD4 would be captured to defined CD4 positive cells).
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	ISO 11238 substance name
Business Rule(s)	The preferred term will be the default value implicitly derived from the Substance ID (4.4).

4.12.5 Substance ID

Table 113 — Substance ID

User Guidance	Substance ID of a substance upon which a defining property depends; the identifier of the substance related to a defining property shall be described where applicable and based on controlled vocabulary.
Example(s)	Solubility in Water, the Substance ID for Water shall be specified.
Conformance	CONDITIONAL
Data Type	II
Values Allowed	ISO 11238 Substance ID

4.12.6 Amount type

Should be indicated per 4.7.12.3.1.

4.12.6.1 Amount

The information related to the amount shall be provided as per specification described in 4.9. The non-numeric value in the case of spectra would be the actual spectra or a reference to the document containing the actual spectra.

4.13 Version (repeat as necessary) (Mandatory)

This section shall provide information on the version of the substance. Where no "Version Number" and "Effective Date" have been assigned by an authority source, the version number shall be set as 0 with the date set as the date of initial submission of the substance. Any changes or updates to a given substance will result in a new version. This could include changes in the definition of the substance or the addition of names, codes. The version of substances should be tracked. Submitters should also indicate if it is a new version of a previous submission.

4.13.1 Version number

Table 114 — Version number

User Guidance	The number of the version of the substance shall be provided.
Example(s)	1, 2, 3, 99
Conformance	MANDATORY
Data Type	INT
Business Rule(s)	If same document set exists, version must be greater than the last submitted version

4.13.2 Effective date

Table 115 — Effective date

User Guidance	The date when the substance was effective shall be provided in line with the ISO 8601 date format. This shall be defined when the substance is generated or modified.
Example(s)	20110219
Conformance	OPTIONAL
Data Type	TS
Business Rule(s)	The value may be implicitly derived.

4.13.3 Change made

Table 116 — Change made

User Guidance	The description of the updates or changes of the substance shall be specified. The field will be left empty when first insert of substance.
Example(s)	Update to previous version to include additional translation.
Conformance	CONDITIONAL
Data Type	CD

5 Substance definitions

All material will be traced back to one of the five types of substances or a mixture substance as described below. All substances will be described as a single substance type or as a mixture. The Annexes give additional examples into how to choose the correct type of substance and what specific elements to submit.

If the substance type as defined in 4.3 is equal to Chemical substance, the following data elements and conformances apply.

5.1 Chemical substance

Chemical substances are a single substance type whose primary defining element is the molecular structure. Chemical substances shall be defined on the basis of their complete covalent molecular structure; the presence of a salt (counter-ion) and/or solvates (water, alcohols) is also captured. Purity, grade, physical form or particle size are not taken into account in the definition of a chemical substance or in the assignment of a Substance ID.

EXAMPLE Purified Water, Water for Injection, Sterile Water for Injection USP, ice and steam all map to the same substance Water, but would be separate specified substances Group 1.

In order to assign a Substance ID for a chemical substance, a complete covalent structure with all stereochemistry (R or S and E or Z) defined shall be submitted. The stoichiometry (mole ratio) of counterions or solvates present in the material shall also be provided in the structural representation.

The molecular representation (molfile, SMILES, InChI, CDX) of the substance shall be provided with all stereochemistry assigned or sites of unknown stereochemistry identified. An actual image of the structure should be provided in an accompanying document.

The information typically provided to INN, USAN, BAN or JAN to assign a non-proprietary name is typically sufficient to define chemical substances. Active substances and excipients are typically defined in accordance with INN, JAN, or USAN definitions or European Pharmacopoeia (EP), Japanese Pharmacopoeia (JP), or United States Pharmacopeia (USP/NF) monographs. The labelling requirements provided by monographs should be taken into

account when defining and distinguishing substances. Information beyond monograph requirements may occasionally be necessary to distinguish substances.

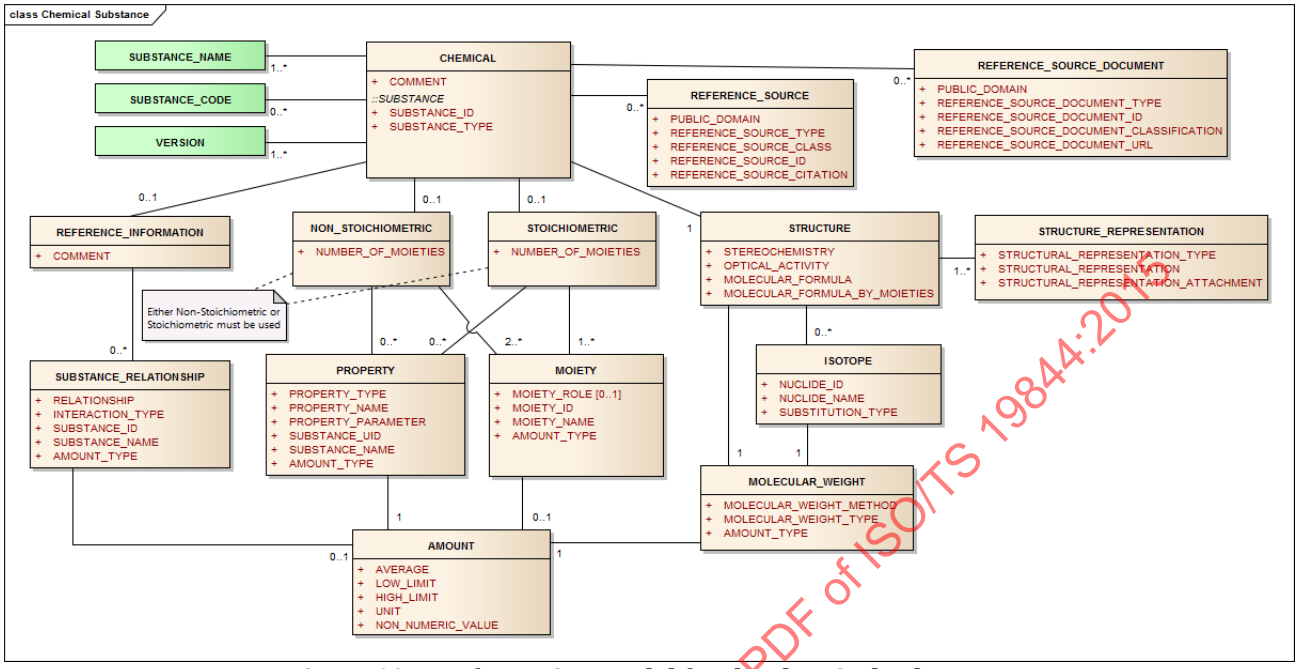


Figure 19 — Information model for the chemical substance

5.1.1 Structure

The information related to the structure shall be provided as per specification described in 4.9

Chemical substances can either be stoichiometric or non-stoichiometric (i.e. substances with variable or unknown stoichiometry). Variable or unknown stoichiometry should be indicated by brackets without a number associated with the brackets around moieties where the stoichiometry is either variable or unknown.

For ingredients in drug products, a preferred format may be specified (mol, SMILES, etc.) for the transmittal of information. Although most chemicals exist primarily with a single structure representation, there are a number of substances that have multiple structural representations in equilibrium. This is illustrated in the glucose (Dextrose) example above.

5.1.2 Stoichiometric

Table 117 — Stoichiometric

User Guidance	This field describes if a substance can be represented as single covalent structural moiety or a salt or solvate with a defined and constant ratio ions or molecules and solvents.
Example(s)	Yes, No
Conformance	MANDATORY
Data Type	BL
Values Allowed	Yes/No

5.1.3 Stoichiometric chemicals

Stoichiometric chemical substances are substances that contain complete chemical structures and definite stoichiometric ratios among moieties. Stoichiometric ratios between moieties should be whole integers and not

fractions. They can be defined by a single representation of the complete molecular structure and a stereochemistry descriptor. The representation can be of one of the four types listed above. The following examples of defining elements for various types of chemical substances help illustrate the information needed to define substances. There are a variety of drawing tools that can generate any of the formats described.

EXAMPLE Ethanol is an example of an achiral substance and can be completely defined by either one of the three following representations:

InChI Representation STRUCTURAL_REPRESENTATION_TYPE: "InChI"

STRUCTURAL_REPRESENTATION: "INCHI=1/C2H6O/c1-2-3/h3H,2H2,1H3"

STEREOCHEMISTRY: "ACHIRAL"

SMILES Representation STRUCTURAL_REPRESENTATION_TYPE: "SMILES"

STRUCTURAL_REPRESENTATION: "CCO"

STEREOCHEMISTRY: "ACHIRAL" (same as above)

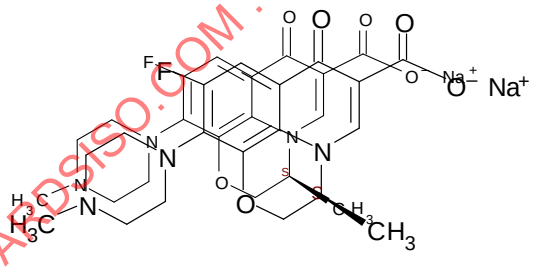
Mol representation STRUCTURAL_REPRESENTATION_TYPE: "MOL"

STRUCTURAL_REPRESENTATION:

```
  3  2  0  0  0  0  0  0  0  0999 V2000
    3.9063   -7.3125    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    4.6223   -6.8991    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    5.3383   -7.3125    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
  1  2  1  0  0  0  0
  2  3  1  0  0  0  0
M  END
```

STEREOCHEMISTRY "ACHIRAL" (same as above)

Levofloxacin Sodium. This is an example of a chiral substance and salt. An INCHI, SMILES or Molfile is sufficient to define this substance.



InChI Representation STRUCTURAL_REPRESENTATION_TYPE: "INCHI"

STRUCTURAL_REPRESENTATION:
"InChI=1S/C18H20FN3O4.Na/c1-10-9-26-17-14-11(16(23)12(18(24)25)8-22(10)14)7-13(19)15(17)21-5-3-20(2)4-6-21;/h7-8,10H,3-6,9H2,1-2H3,(H,24,25);/q;+1/p-1/t10-;/m0./s1"

STEREOCHEMISTRY "CHIRAL"

SMILES Representation

STRUCTURAL_REPRESENTATION_TYPE: “SMILES”

STRUCTURAL_REPRESENTATION:
“C[C@H]1COc2c3n1cc(c(=O)c3cc(c2N4CCN(CC4)C)F)C(=O)[O-].[Na+]”

STEREOCHEMISTRY: “CHIRAL”

Mol Representation STRUCTURAL_REPRESENTATION_TYPE: “MOL”

STRUCTURAL_REPRESENTATION:

```
27 29  0  0  1  0  0  0  0  0999 V2000
   14.5167   -9.0507    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
   15.1250   -7.9548    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   13.8500   -8.7215    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   13.2625   -9.0840    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   13.8500   -8.0048    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   15.1250   -8.7048    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   14.4667   -7.6340    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   12.6334   -8.7673    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   13.2625   -9.8506    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   12.6334   -8.0048    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   13.2292   -7.6548    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   12.0167   -9.1298    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
   15.8250   -7.6007    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   14.5167   -9.7840    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   10.7417   -9.8965    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
   14.4667   -6.8965    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   13.9125  -10.1632    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   12.0167   -9.8798    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   11.3584   -8.7757    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   15.8250   -6.9548    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   11.9542   -7.7132    0.0000 F   0  0  0  0  0  0  0  0  0  0  0  0
   10.7417   -9.1340    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   11.4000  -10.2465    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   16.6458   -7.9631    0.0000 O   0  5  0  0  0  0  0  0  0  0  0  0
   10.0792  -10.1715    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   15.3500  -10.3215    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   17.3000   -7.9630    0.0000 Na  0  3  0  0  0  0  0  0  0  0  0  0
 21 10  1  0  0  0  0
 22 19  1  0  0  0  0
 23 18  1  0  0  0  0
 24 13  1  0  0  0  0
 25 15  1  0  0  0  0
 14 26  1  1  0  0  0
  4  9  1  0  0  0  0
  5  7  1  0  0  0  0
  8 10  2  0  0  0  0
 15 23  1  0  0  0  0
  2  6  2  0  0  0  0
  3  1  1  0  0  0  0
  4  3  2  0  0  0  0
  5  3  1  0  0  0  0
  6  1  1  0  0  0  0
  7  2  1  0  0  0  0
  8  4  1  0  0  0  0
  9 17  1  0  0  0  0
 10 11  1  0  0  0  0
 11  5  2  0  0  0  0
 12  8  1  0  0  0  0
 13  2  1  0  0  0  0
 14  1  1  0  0  0  0
 15 22  1  0  0  0  0
 16  7  2  0  0  0  0
 17 14  1  0  0  0  0
 18 12  1  0  0  0  0
 19 12  1  0  0  0  0
 20 13  2  0  0  0  0
M  CHG  2  24  -1  27   1
M  END
```

STEREOCHEMISTRY “CHIRAL”

5.1.4 Comment

Comments should be used sparingly to capture information that does not fit into other fields. They may be used to indicate that alternate structures exist.

Table 118 — Comment

User Guidance	Any comment can be provided in this field, if necessary.
Conformance	OPTIONAL
Data Type	ST

Chemical substances, whose primary defining element is the molecular structure, shall be defined on the basis of their complete covalent molecular structure; the presence of a salt (counter-ion) and/ or solvates is also captured. The molecular structure, the molecular formula, the **molecular weight** and optical activity, together with the representation of the stereochemistry are mandatory elements to be provided. Although the molecular weight can be derived from the structure, this element should be presented at the substance level for chemical substances in order to substantiate the provided structural information.

The ISO 11238 Substance standard should be updated for the element molecular weight at the next version since this element is missing in the Class Chemical substances.

In addition the molecular formula should be described according to the moieties even for stoichiometric substances. See above Molecular formula by moieties and Molecular formula, which is ment to be equal to the sum of molecular formula by moieties.

In addition to earlier versions the element group moiety Is connected with the stoichiometric chemical substance.

5.1.5 Non-stoichiometric chemicals

The section applies to chemical substances that do not have defined stoichiometry.

This section is used only when the Stoichiometric Element (Section 0) is equal to No.

Every non-stoichiometric substance will have more than one moiety. Each moiety will be enclosed in brackets indicating stoichiometry where known and transmitted in a single structural representation. Every non-stoichiometric moiety will have a complete structural representation and also be described as a substance in its own right. If ranges of amounts for the non-stoichiometric moieties are known they will be preferably captured as mole ratios or weight percent if mole ratios cannot be determined.

5.1.5.1 Number of moieties

Table 119 — Number of moieties

User Guidance	The number of moieties specified shall be provided. Non-stoichiometric chemical substances must have at least two moieties. Each moiety shall be represented by a chemical structure.
Example(s)	2
Conformance	CONDITIONAL
Data Type	INT
Business Rule(s)	The numeric value shall be always ≥ 2; The value is automatically calculated (Implicit) based on the number of Moiety classes as described in 5.1.5.2)

5.1.5.2 Moiety (repeat as necessary)

The moiety section serves for the description of both the moiety and the other fragment constituting the substance. Each moiety within the chemical substance is to be identified and the composition range of the moieties when known is to be provided by means of the following data elements:

The element group Moiety should be attached also to the element class Chemical, Stoichiometric substances in order to describe a salt and solvates relationship to the active moiety in an unambiguous way with respect to the description of the molecular formula and sometimes the official names.

The element molecular weight must be included in the next version of the standard and be included in the Structure Class Information Model together with the element group amount. The element group moiety and amount should also be tied to the element group stoichiometric substances.

This will help to describe salts and solvates and hydrated substances in a more unambiguous way with respect to the description of the molecular formula and sometimes official names, as in Figure 10.

EXAMPLE Aluminium Sesquichlorohydrex Propylene Glycol is defined by relationships between five moieties, the aluminium cation, chloride and hydroxide ions, propylene glycol and water; three of the five moieties have mole ratios to each other. The ratio of aluminium cation to chloride to hydroxide is 2:3:3. The amount of propylene glycol and water is variable and not stoichiometric.

5.1.5.2.1 Moiety role

For many peptides the moiety role of counterions such as acetate should be indicated.

Table 120 — Moiety role

User Guidance	The role of the moiety should be specified if there is a specific role the moiety is playing.
Example(s)	Counter-ion
Conformance	OPTIONAL
Data Type	CD

5.1.5.2.2 Moiety ID

Each moiety in a non-stoichiometric substance will be identified with a Substance ID if available. If not available a new ID will be assigned with the submission.

Table 121 — Moiety ID

User Guidance	The unique identifier assigned to the substance representing the moiety based on the ISO 11238 substance controlled vocabulary.
Conformance	MANDATORY
Data Type	II
Values Allowed	ISO 11238 Substance ID

5.1.5.2.3 Moiety name

Each moiety will also have at least one name associated with the moiety.

Table 122 — Moiety name

User Guidance	The name of the moiety shall be provided.
Example(s)	propylene glycol
Conformance	MANDATORY
Data Type	CD
Values Allowed	ISO 11238 substance name.
Business Rule(s)	ISO 11238 substance preferred name is selected as default value. The substance name is implicit and derived from the Moiety ID described in 5.1.5.2.2.

5.1.5.2.4 Amount type

Typically there will be an amount type relationship that will be consistent for all of the moieties. A single moiety will typically be assigned the value 1. Ranges of moieties that are stoichiometric will be assigned successive values. The amount type should be captured according to 4.7.12.3.1.

5.1.5.2.5 Amount

The amount shall be captured as specified in 4.9.

5.1.5.3 Property (repeat as necessary)

Property shall be captured as specified in 4.11.

5.2 Proteins/peptides

A protein is defined as a single unit of a linear amino acid sequence, or a combination of subunits that are either covalently linked or have a defined invariant stoichiometric relationship. This includes all synthetic, recombinant and purified proteins of defined sequence, whether the use is therapeutic or prophylactic. This set of elements will be used to describe albumins, coagulation factors, cytokines, growth factors, peptide/protein hormones, enzymes, toxins, toxoids, recombinant vaccines, and immunomodulators.

Proteins and peptides are defined by their molecular structure based on the amino acid sequence, disulfide linkages, sites and a general type of glycosylation, based on the cell or organism type from which the protein was isolated from or produced (e.g. yeast, plant, mammalian, human). The method of production is generally not a defining element for proteins and peptides. For a given non-glycosylated peptide, whether naturally isolated, produced by recombinant technology, or chemically synthesised, it will be defined as the same substance when there are no resultant differences in the amino acid sequence and disulfide linkages.

EXAMPLE: Recombinant and chemically synthesised salmon calcitonin are chemically identical and therefore considered to be the same substance with the same Substance ID.

Proteins that have consistently different types of glycosylation are assigned separate Substance IDs. Human and primate glycosylation differ from other mammalian glycosylation in that terminal sialic acid residues are only acetylated and not glycolated. Glycosylation in human and old world monkeys also differ from other mammals in lacking the alpha, 1,3-galactose epitopes. Avian, insect, plant, fish and yeast also have consistently different glycosylation. **Yeast, plant, insect and avian glycosylation are also significantly different from mammalian and human glycosylation and glycoproteins produced in such systems would each have a separate Substance ID even if the amino acid sequences and disulfide linkages are the same**

EXAMPLE Coagulation Factor VIII isolated from human blood versus that produced by recombinant technology in Chinese hamster ovary (CHO) cells are considered to be different substances even if they have the same amino acid sequence and disulfide linkages. **The glycosylation type would distinguish these types.**

In defining proteins, microheterogeneity is typically not taken into account in the assignment of a Substance ID. Microheterogeneity generally refers to slight differences in the structure of proteins. Most of these differences are due to post-translational modification. There can also be slight differences in the ends of molecules due to differential protease activity. Many of these differences are in the extent of occupancy and type of glycan present on a given site. Therefore a given glycoprotein produced in different mammalian cell lines would have the same Substance ID. Post–translational modifications will be captured when they are both complete and non-variable or when they are essential for bioactivity.

EXAMPLE In many blood clotting factors, glutamate residues are converted to gamma-carboxyglutamate residues. These modifications are essential for activity and will be captured as a structural modification.

5.2.1 Microheterogeneity

The ISO 11238 Substance Standard does not explicitly capture microheterogeneity. Microheterogeneity is typically caused by differential post-translational modifications that lead to differences in the type and extent of glycosylation, phosphorylation, sulphation and/or other modifications throughout the ensemble of molecules of a given protein. Recently regulatory agencies are receiving more detailed information on the microheterogeneity of proteins in medicinal products. This information may be used to further distinguish proteins at the specified substance Group 1 level. Detailed guidelines on capturing this information may be given in subsequent editions of this Technical Specification and Protein Annex.

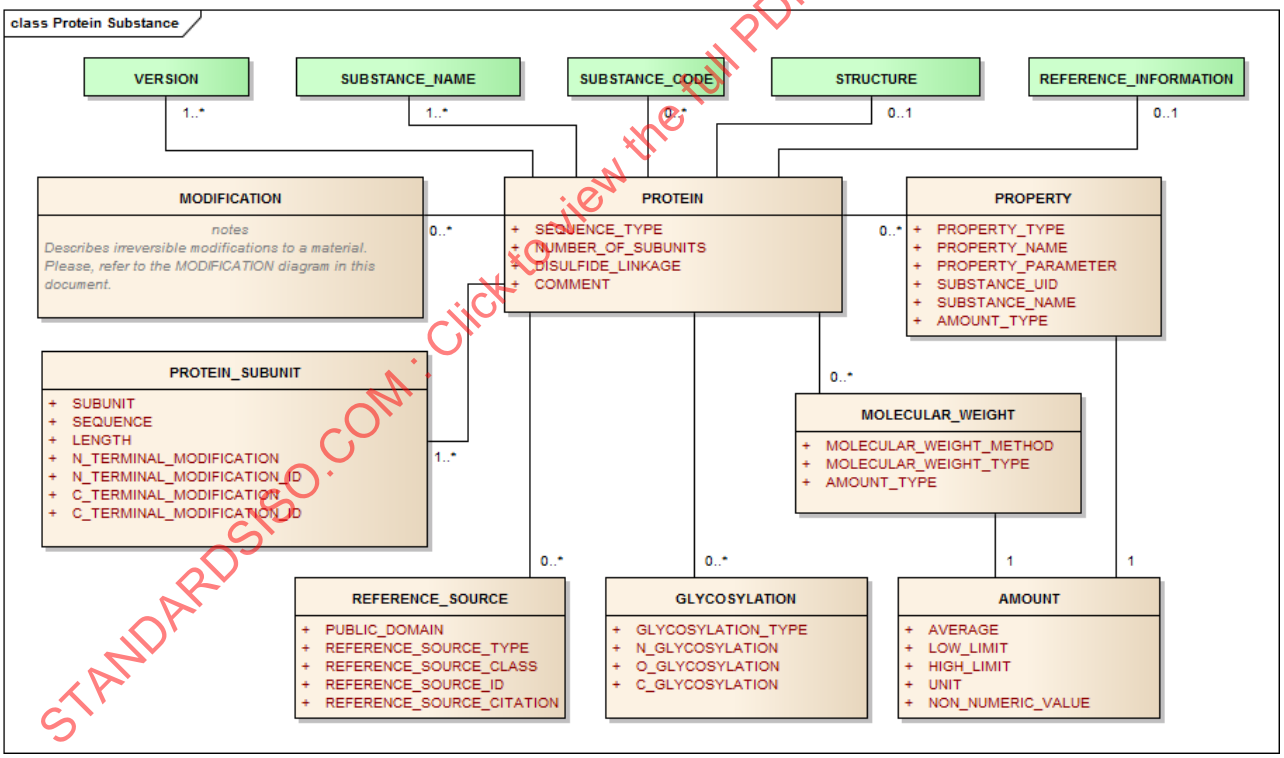


Figure 11 — Information model for the protein substance

If the substance type as defined in 4.3 is equal to *Protein substance*, the following data elements and conformances apply in order to assign Substance IDs.

5.2.2 Sequence type

Table 123 — Sequence type

User Guidance	The protein descriptive elements will only be used when a complete or partial amino acid sequence is available or derivable from a nucleic acid sequence.
Example(s)	COMPLETE; PARTIAL
Conformance	MANDATORY
Data Type	CD

5.2.3 Number of subunits

Although many proteins form multimeric complexes, particularly at high concentrations, subunits are typically either connected through disulfide linkages or other covalent linkages or have strong non-covalent interactions with defined stoichiometry largely independent of concentration.

Table 124 — Number of subunits

User Guidance	Number of linear sequences of amino acids linked through peptide bonds. The number of subunits constituting the protein shall be described. It is possible that the number of subunits can be variable.
Example(s)	1
Conformance	CONDITIONAL
Data Type	INT
Business Rule(s)	This number is implicit and derived from number of protein subunits.

5.2.4 Disulfide linkage

Disulfide linkages imply that two cysteines connect forming a sulfur to sulfur linkage. There are also instances where disulfide linkages can form through other modified or substituted acids or amino acids containing a thiol (See Desmopressin example in Annex C: Protein/Peptide Substances).

Table 125 — Disulfide linkage

User Guidance	The disulphide bond between two cysteine residues either on the same subunit or on two different subunits shall be described. The position of the disulfide bonds in the protein shall be listed in increasing order of subunit number and position within subunit followed by the abbreviation of the amino acids involved. The disulfide linkage positions must actually contain the amino acid Cysteine at the respective positions.
Example(s)	"Subunit 1 position 10 — Subunit 2 position 16" refers to a disulfide linkage between the residue of cysteine in the position 10 and 16 respectively of the first and second subunit. A convenient shorthand such as 1_10-2_16 could also be used.
Conformance	MANDATORY
Data Type	ST
Values Allowed	The values have to be separated with semi-colons as defined in the examples in the Annex C.

5.2.5 Comment

Table 126 — Comment

User Guidance	Any comment can be provided in this field, if necessary.
Conformance	OPTIONAL
Data Type	ST

5.2.6 Protein subunit (repeat as necessary)

This section refers to the description of each subunit constituting the protein. A subunit is a linear sequence of amino acids linked through peptide bonds. The Subunit information shall be provided when the finished protein is a complex of multiple sequences; subunits are not used to delineate domains within a single sequence. Subunits are listed in order of decreasing length; sequences of the same length will be ordered by decreasing molecular weight; subunits that have identical sequences will be repeated multiple times.

5.2.6.1 Subunit

Table 127 — Subunit

User Guidance	Index of primary sequences of amino acids linked through peptide bonds in order of decreasing length. Sequences of the same length will be ordered by molecular weight. Subunits that have identical sequences will be repeated and have sequential subscripts.
Example(s)	1, 2, 3 (..)
Conformance	MANDATORY
Data Type	INT

5.2.6.2 Sequence

Table 128 — Sequence

User Guidance	The sequence information shall be provided enumerating the amino acids from N- to C-terminal end using standard single-letter amino acid codes. Uppercase shall be used for L-amino acids and lowercase for D-amino acids. Transcribed proteins will always be described using the translated sequence; for synthetic peptide containing amino acids that are not represented with a single letter code an X should be used within the sequence. The modified amino acids will be distinguished by their position in the sequence..
Conformance	MANDATORY
Data Type	ST

5.2.6.3 Length

Table 129 — Length

User Guidance	Length of linear sequences of amino acids contained in the subunit
Example(s)	25
Conformance	MANDATORY
Data Type	INT
Business Rule(s)	The sequence length is implicitly derived from the linear sequences of amino acids contained in the subunit.

5.2.6.4 N_Terminal modification

Table 130 — N_Terminal modification

User Guidance	The name of the fragment modified at the N-terminal of the protein shall be specified.
Example(s)	N-Terminal Pyroglu Formation; Pyroglutamic acid (pE),
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	ISO 11238 substance names
Business Rule(s)	If an N-terminal modification exists, this element becomes mandatory. The preferred name as specified in ISO 11238 is the default value. The value is implicit and derived based on the N_Terminal Modification ID.

5.2.6.5 N_Terminal modification ID

Table 131 — N_Terminal modification ID

User Guidance	Unique identifier for molecular fragment modification based on the ISO 11238 Substance ID.
Example(s)	UNII code; SZB8301W42 (UNII)
Conformance	CONDITIONAL
Data Type	II
Values Allowed	ISO 11238 Substance ID

5.2.6.6 C_Terminal modification

Table 132 — C_Terminal modification

User Guidance	The modification at the C-terminal shall be specified.
Example(s)	amide, ethyl ester etc., C_Terminal Lysine removal
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	ISO 11238 substance names
Business Rule(s)	If a C-terminal modification exists, this element becomes mandatory. The preferred name as specified in ISO 11238 is the default value. The value is implicit and derived based on the C_Terminal Modification ID.

5.2.6.7 C_Terminal modification ID

Table 133 — C_Terminal modification ID

User Guidance	Unique identifier for molecular fragment modification based on the ISO 11238 Substance ID.
Example(s)	UNII code, K3Z4F99H6 (UNII), Lysine (K)
Conformance	CONDITIONAL
Data Type	II
Values Allowed	ISO 11238 Substance ID

5.2.7 Molecular weight (repeat as necessary)

The molecular weight or a molecular weight range for proteins, polymers or nucleic acids should be provided if known. Multiple molecular weights that either depend on the method or type should also be provided.

If the molecular weight is available, the following specifications apply:

5.2.7.1 Molecular weight method

There are a variety of methods used to determine molecular weight. The method is often associated with the type of molecular weight. Each method should be captured.

Table 134 — Molecular weight method

User Guidance	The method by which the molecular weight was determined.
Example(s)	SDS-PAGE, calculated, light scattering, viscosity, gel permeation centrifugation; end group analysis.
Conformance	MANDATORY
Data Type	CD

5.2.7.2 Molecular weight type

There are basically four types of molecular weights associated with polymers, proteins and nucleic acids. The four types are number average, M_n; viscosity average, M_v; weight average, M_w; and Z-average, M_z. Polydispersity is an important property that will distinguish different polymers from one another. Polydispersity is typically defined as the ratio of weight average molecular weight to the number average molecular weight (M_w/M_n).

Table 135 — Molecular weight type

User Guidance	Type of molecular weight such as exact, average (also known as. number average), weight average
Example(s)	Number Average, Weight Average
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	M_n; viscosity average, M_v; weight average, M_w; and Z-average M_z. Polydispersity
Business Rule(s)	If known it shall be provided

5.2.7.2.1 Amount

The information on amount shall be captured as specified in 4.9.

5.2.8 Glycosylation

Glycosylation is both variable and heterogeneous. Glycosylation sites typically have multiple glycans attached to them and even differ on the extent of glycosylation on a given site. For Glycosylated proteins, the type and sites of glycosylation should be provided. Glycosylation is applicable to both proteins and structurally diverse substances. Although sites of N-glycosylation can generally be predicted from the amino acid sequence, sites of O-glycosylation and C-glycosylation are usually determined experimentally and may not be completely known at the time of submission and can be added to a substance when known. Generally all sites that have occupancy greater than 5% should be submitted to define the glycoprotein. Although analytical techniques are evolving to better characterise the glycans attached to proteins this data will not be required to define a substance but could be captured at the specified substance Group 1 level. There can be substantial differences in glycosylation that are

often dependent on the particular clone of a cell line that is used to produce a protein or even vary on a batch to batch basis.

5.2.8.1 Glycosylation type

The different glycosylation types are determined by the cell that either the protein or specified substance was synthesised in. The different types are based on consistent differences in the glycosylation. Human glycosylation differs from mammalian glycosylation in several ways, human glycans do not contain glycolic acid residues on terminal sialic acid and they do not contain terminal .alpha.1-> 3-galactose. Old world monkeys also lack the ability to form .alpha. 1-> 3-galactose, but do have glycolic acid esters on terminal sialic acid residues. Other glycosylation types typically have substantial differences in both glycans used and the extent of site occupancy. Glycosylation type may also be captured for structurally diverse substances, particularly vaccines which may be produced in either human, mammalian cell or old world monkey cells. The glycosylation type will often have substantial effects on the immunogenicity of the protein or vaccine.

Table 136 — Glycosylation type

User Guidance	The type of the glycan shall be specified based on a controlled vocabulary.
Example(s)	HUMAN, OLD WORLD MONKEY, MAMMALIAN, AVIAN, REPTILIAN; FUNGAL, BACTERIAL, PLANT, INSECT, MAMMALIAN AFUCOSYLATED, YEAST HUMANISED
Conformance	CONDITIONAL
Data Type	CD

5.2.8.2 N-Glycosylation

N-linked glycosylation: The sites of N-glycosylation can typically be predicted from the amino acid sequence. The sequence NX(S/T) which indicates an asparagine followed any amino acid other than proline and then followed by either a serine or threonine will typically be a site of N-glycosylation.

Table 137 — N-Glycosylation

User Guidance	Information on the site of n-glycosylation (asparagine); n-glycosylation is to be listed according to the protein sequence.
Example(s)	Position 34 of this Subunit, an Asparagine; for a single subunit protein 1_34
Conformance	CONDITIONAL
Data Type	ST
Values Allowed	The position of that subunit must actually contain an amino acid that can have N-linked glycosylation.

5.2.8.3 O-Glycosylation

O-linked glycosylation: The sites of O-glycosylation cannot typically be predicted from the sequence alone and are discovered through analytical methods. Sites may become identified at later stages in clinical development. The Substance ID typically should not change as new sites are identified but the version of the substance would.

Table 138 — O-Glycosylation

User Guidance	Information on the site of o-glycosylation (serine, threonine, tyrosine, hydroxylysine, hydroxyproline) shall be provided.
Example(s)	Position 34 of this Subunit, a Serine.
Conformance	CONDITIONAL
Data Type	ST
Values Allowed	The position of that subunit must actually contain an amino acid with an OH group (i.e. serine or threonine or tyrosine).

5.2.8.4 C-Glycosylation

C-linked glycosylation: The consensus sequence for C-glycosylation is WxxW/F where the first tryptophan undergoes c-mannosylation.

Table 139 — C-Glycosylation

User Guidance	Information on the C-Glycosylation: site of C-glycosylation (tryptophan) shall be provided.
Example(s)	Position 45 of this Subunit, a Tryptophan
Conformance	CONDITIONAL
Data Type	ST

5.2.9 Structure

The information related to the structure shall be provided as per specification described in Section A structural representation should be provided for any structural modifications of the proteins. The representation should include the first atom of the protein which is modified. For amino acid replacements and fragment additions a substance will be created in its own right. In addition or in lieu of the data structure listed above, HELM notation can be a suitable format for transmitting definitive information on selected proteins and peptides (see http://www.pistoiaalliance.org/projects/active-projects/hierarchical-editing-language-for-macromolecules-helm/).

5.2.10 Modification

If applicable, the modification should be described based on specification provided in 4.10.

5.2.11 Property

If applicable, the Property should be described based on specification provided in 4.11.

5.2.11.1 Amount type

See 4.7.12.3.1.

5.2.11.1.1 Amount

Amount shall be captured as specified in 4.9.

5.2.12 Molecular weight

Information on molecular weight should be described based on specifications in 5.2.7.

5.3 Nucleic acids

Nucleic acids are defined by three distinct elements: the base, sugar and linkage. Individual substance/moiety IDs will be created for each of these elements. The nucleotide sequence will be always entered in the 5'-3' direction.

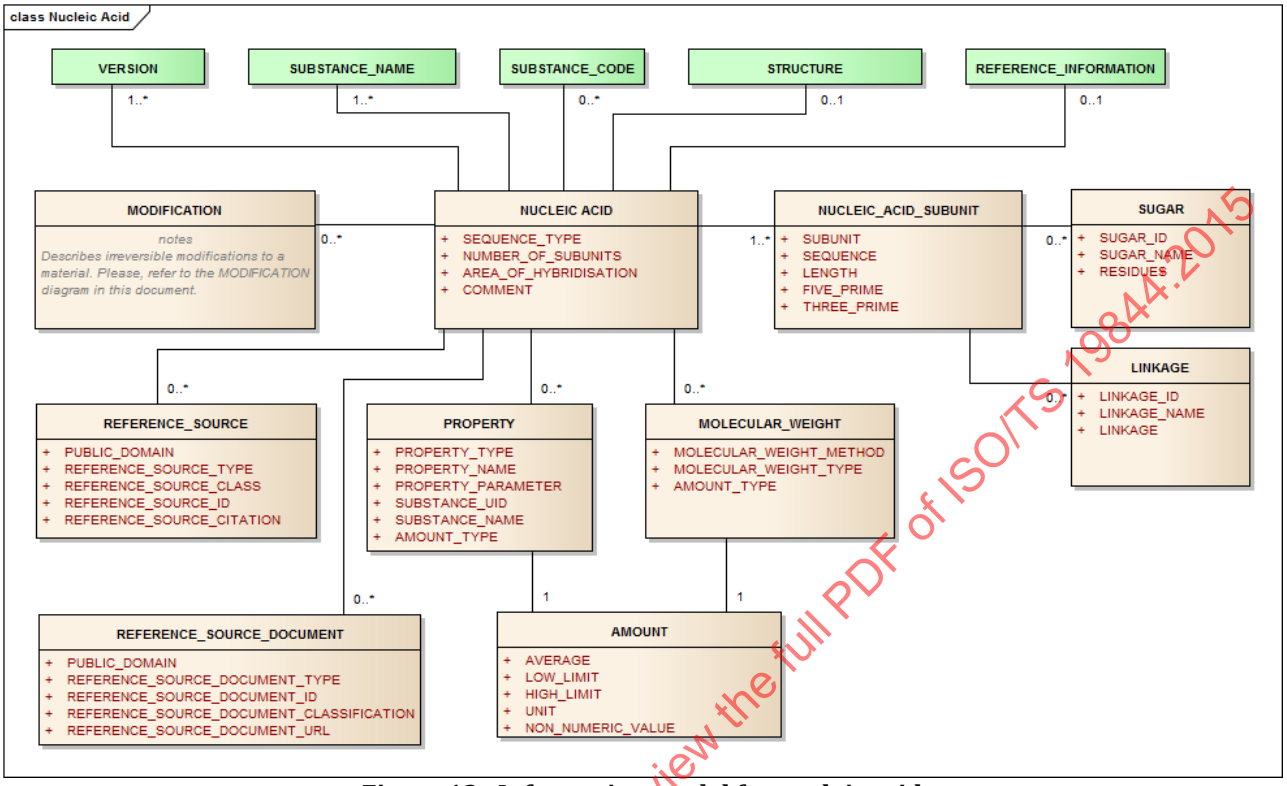


Figure 12– Information model for nucleic acid

The nucleic acid set of elements are only to be used to describe nucleic acids that have:

- A length greater than three bases or base pairs;
- Elements to be used for oligonucleotides
- Genes used in gene therapy
- Any nucleic acid aptomers

For the description of nucleic acids the following information can be used to assign Substance IDs:

- Nucleotide Sequence
- Sugar or sugar-like entities
- Linkage (typically phosphate)
- Nucleic Acid Type (RNA, DNA, plasmid, single or double stranded)
- Any Modifications to the substance either sequence specific (i.e. modified or unnatural base, sugar, phosphate linkage) or non-sequence specific modifications.

5.3.1 Structure

There are multiple alternative ways of representing the structure of nucleic acids. Figure 13 provides an example of one of the possible structural representations. While the sequence of nitrogenous bases is given as a single letter notation, the backbone structure is represented as a polymeric single repeat unit (5.4 "Polymers"). A structural representation can be provided for relatively short oligonucleotides but the textual fields illustrated below should be sufficient to describe nucleic acid based substances.

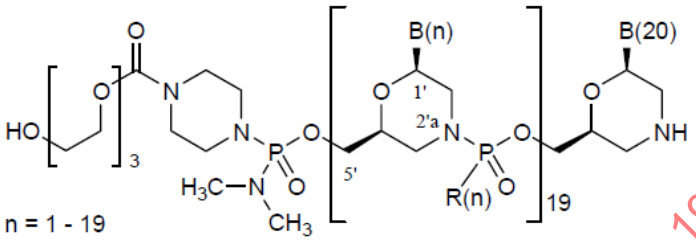


Figure 13 — Representative Structural Representation of nucleic acids

5.3.2 Sequence type

Table 140 — Sequence type

User Guidance	The type of the sequence shall be specified based on a controlled vocabulary.
Example(s)	COMPLETE; PARTIAL
Conformance	MANDATORY
Data Type	CD

5.3.3 Number of subunits

Table 141 — Number of subunits

User Guidance	The number of linear sequences of nucleotides linked through phosphodiester bonds shall be described. Subunits would be strands of nucleic acids that are tightly associated typically through Watson-Crick base pairing. NOTE: If not specified in the reference source, the assumption is that there is 1 subunit.
Example(s)	1, 2, or 3
Conformance	MANDATORY
Data Type	INT
Business Rule(s)	This number is implicit and derived from number of nucleic acid subunits.

5.3.4 Area of hybridisation

Table 142 — Area of hybridisation

User Guidance	The area of hybridisation shall be described if applicable for double stranded RNA or DNA. The number associated with the subunit followed by the number associated to the residue shall be specified in increasing order. The underscore "_" shall be used as separator as follows: "Subunit_number Residue".
Example(s)	Each residue shall be specified followed by a comma and space.
Conformance	CONDITIONAL
Data Type	ST

5.3.5 Comment

Table 143 — Comment

User Guidance	Any comment can be provided in this field, if necessary.
Conformance	OPTIONAL
Data Type	ST

5.3.6 Nucleic acid subunit (repeat as necessary)

Subunits are listed in order of decreasing length; sequences of the same length will be ordered by molecular weight; subunits that have identical sequences will be repeated multiple times.

5.3.6.1 Subunit

Table 144 — Subunit

User Guidance	Index of linear sequences of nucleic acids in order of decreasing length. Sequences of the same length will be ordered by molecular weight. Subunits that have identical sequences will be repeated and have sequential subscripts.
Example(s)	1, 2, 3 (..)
Conformance	MANDATORY
Data Type	INT

5.3.6.2 Sequence

Table 145 — Sequence

User Guidance	Actual nucleotide sequence notation from 5' to 3' end using standard single letter codes. In addition to the base sequence, sugar and type of phosphate or non-phosphate linkage should also be captured.
Example(s)	GATTCA
Conformance	MANDATORY
Data Type	ST
Values Allowed	Sequence of the letters G, C, T, A, and U

5.3.6.3 Length

Table 146 — Length

User Guidance	The length of the sequence
Example(s)	5
Conformance	MANDATORY
Data Type	INT
Business Rule(s)	The sequence length is implicitly derived by the sequence description.

5.3.6.4 Sugar (Repeat as necessary) (new classes to be included in the second edition of ISO 11238)

5.3.6.4.1 Residues

The residues that contain a given sugar entity will be captured. Each sugar present in the nucleic acid will be represented.

Table 147 — Residues

User Guidance	The residues that contain a given sugar will be captured. The order of given residues will be captured in the 5'-3'direction consistent with the base sequences listed above.
Example(s)	1_1-1_20 would indicate that residues 1-20 of a given strand contain a particular sugar.
Conformance	MANDATORY
Data Type	ST
Values Allowed	ISO 11238 Substance ID

5.3.6.4.2 Sugar ID

The Substance ID type of sugar to which the base is connected. For naturally occurring nucleic acids the sugar is typically either .beta.-D-ribose or .beta.-D-deoxyribose. The actual configuration of the sugar shall be captured.

Table 148 — Sugar ID

User Guidance	The Substance ID of the sugar or sugar-like component that make up the nucleotide
Example(s)	UNII code for ribose:
Conformance	MANDATORY
Data Type	II
Values Allowed	ISO 11238 Substance ID

5.3.6.4.3 Sugar name

The type of sugar to which the base is connected; for naturally occurring nucleic acids, the sugar is typically either .beta.-d-ribose or .beta.-d-deoxyribose.

Table 149 — Sugar name

User Guidance	The name of the sugar or sugar-like component that make up the nucleotide
Example(s)	Ribose, deoxyribose, 2,5-dihydroxy-morphilino
Conformance	MANDATORY
Data Type	CD
Values Allowed	ISO 11238 substance name
Business Rule(s)	This is implicit and derived from the Sugar ID. The preferred name is the default value.

5.3.6.5 Linkage group (new classes to be included in the second edition of ISO 11238)

The linkages between sugar residues will also be captured.

5.3.6.5.1 Linkage

This element will indicate the orientation of the linkage. The 3‘-5‘ linkage is the most prevalent form.

Table 150 — Linkage

User Guidance	The entity that links the sugar residues together should also be captured for nearly all naturally occurring nucleic acid the linkage is a phosphate group. For many synthetic oligonucleotides phosphorothioate linkages are often seen. Linkage orientation is assumed to be 3’-5’. If the linkage is either 3’-3’ or 5’-5’ this should be specified.
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	3’-5’, 3’-3’ or 5’-5’ this should be specified.

5.3.6.5.2 Linkage name

Table 151 — Linkage name

User Guidance	Each linkage will be registered as a fragment and have at least one name. A single name shall be assigned to each linkage.
Example(s)	Phosphothioate
Conformance	CONDITIONAL
Data Type	CD

5.3.6.5.3 Linkage ID

Table 152 — Linkage ID

User Guidance	Each linkage will be registered as a fragment and have an ID.
Conformance	CONDITIONAL
Data Type	II

5.3.6.6 Five prime

Table 153 — Five prime

User Guidance	The nucleotide present at the 5’ terminal shall be specified based on a controlled vocabulary. Since the sequence is represented from the 5' to the 3' end, the 5’ prime nucleotide is the letter at the first position in the sequence. A separate representation would be redundant.
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	Implicit NOTE: If not specified in the available reference source this value shall be derived from the nucleic acid sequence.

5.3.6.7 Three prime

Table 154 — Three prime

User Guidance	The nucleotide present at the 3’ terminal shall be specified based on a controlled vocabulary. Since the sequence is represented from the 5' to the 3' end, the 5’ prime nucleotide is the letter at the last position in the sequence. A separate representation would be redundant.
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	NOTE: If not specified in the available reference source this value shall be derived from the nucleic acid sequence.

5.3.7 Modification

If applicable, the modification should be described based on specification provided in 4.10.

5.3.8 Property

If applicable, the Property should be described based on specification provided in 4.11.

5.3.8.1 Amount type

Refer to 4.7.12.3.1

5.3.8.1.1 Amount

Amount shall be captured as specified in 4.9.

5.3.9 Molecular weight

Information on molecular weight should be described based on specifications in 5.2.7.

5.3.9.1 Reference source

Information on Reference source should be described based on specifications in 4.6.

5.4 Polymers (to be addressed in more detail in the second edition of this Technical Specification)

Polymers are polydisperse molecular ensembles defined using combination structural and descriptive elements. The structural elements are the structural repeating units along with end groups and salt forms. The number and/or weight average molecular weight, degree of polymerisation, degree or extent of substitution of fragments and physical properties related to molecular weight are descriptive elements that may be required to assign a Substance ID. For synthetic polymers, the monomers used to prepare the polymer should also be provided. For copolymers, the type of copolymer (random, block, branched, etc.), along with the block/branch size and number of blocks or branches, should also be provided. A technical specification sheet and/or information sheet from a manufacturer is often sufficient to generate a Substance ID. For polymers derived from biological matrixes, the source of the polymer should also be provided.

EXAMPLE Starch is a mixture of two polymers amylose and amylopectin. It is normally isolated from wheat, corn, rice, potato, or tapioca and source material elements should be used to capture the source.

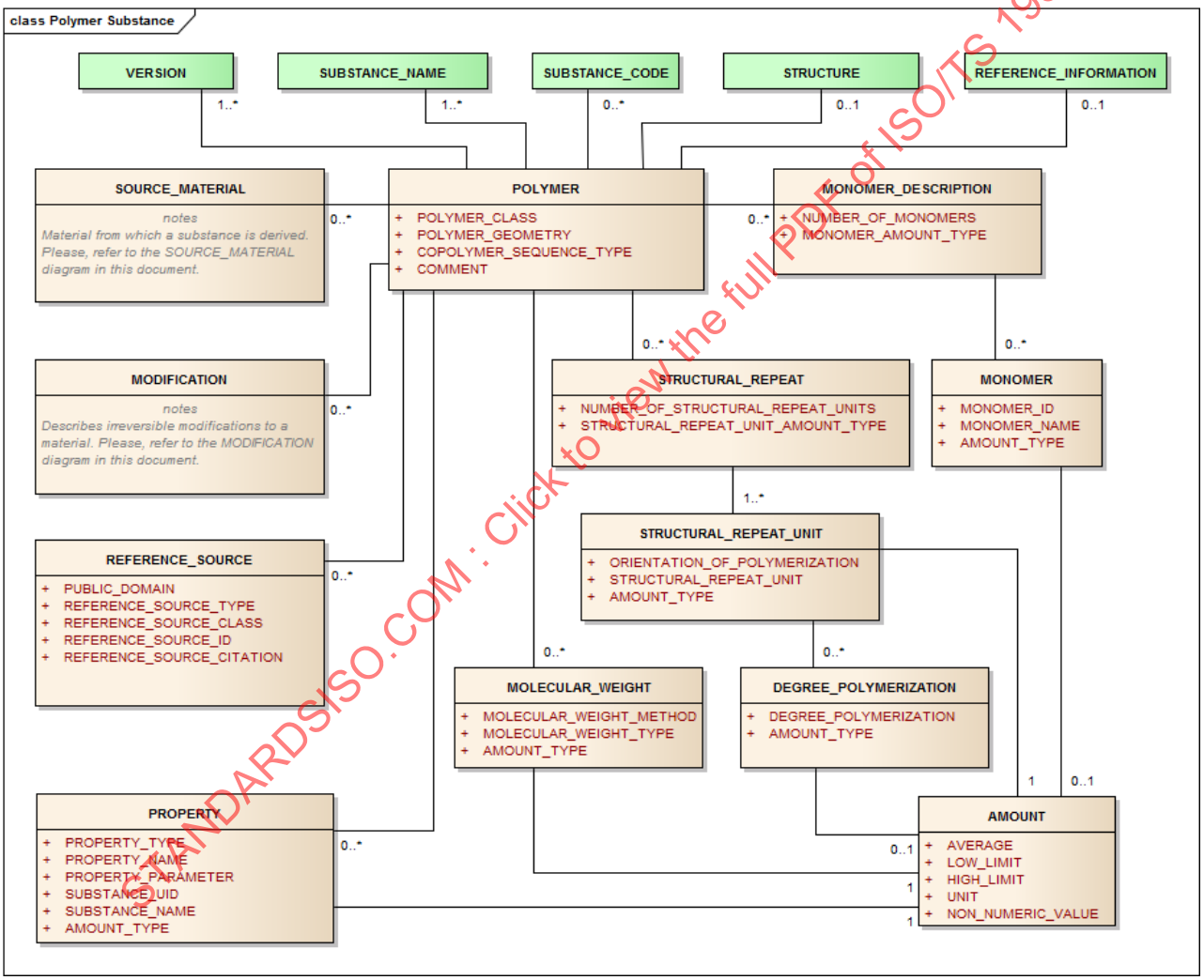


Figure 14 — Information model for polymer substance

For the description of polymer substances a combination of the following elements should be submitted to assign Substance IDs:

- Structural Representations of the Structural Repeating Units and/or Monomers
- Explicit Connectivity between Structural Repeating Units and End-groups

- Number Average Molecular Weight and/or Limits
- Weight Average Molecular Weight and/or Limits
- Degree of Polymerisation
- Physical Properties Related to Molecular Structure (i.e. Viscosity, Density, etc.)
- For Copolymers: Relative Amounts of Monomers
- For Block Copolymers: Block Size and Block Number
- For Substituted Polymers: Substituents and the Degree of Substitution
- Any Process that Modifies the Chemical Structure

The general set of information shall be provided by means of the followings data elements.

5.4.1 Substance name

The information related to the name of the substance shall be provided as per specification described in 4.5. The naming of polymers can be different from other types of substances. Official Names are typically given to classes of substances and not individual polymers. A naming convention to distinguish polymers within a class may be developed.

EXAMPLE POLYQUATERNIUM-47 is an INCI for an acrylate copolymer but it is necessary to indicate the relative amounts of the different acrylate Structural Repeat Units and the molecular weight. A specific name for a distinct polymer would be POLYQUATERNIUM-47 (METHACRYLAMIDOPROPYLTRIMETHYLAMMONIUM CHLORIDE-CO-METHYL ACRYLATE-CO-ACRYLIC ACID 45:10:45; 1200000 MW).

5.4.2 Structure

The information related to the structure shall be provided as per specification described in 4.9. The structure wherever possible should be representative of the structural repeating units that make up the polymer. The structure should be contained within brackets. The end groups or an asterisk indicating connectivity to another structural repeating unit should be drawn outside the brackets. For random copolymers, each structural repeating unit should be drawn in the same structural representation, but not directly connected to each other. For block copolymers, the structural repeating units will be directly connected to each other. End groups, when known, should also be attached to the structural repeating units. MolFile are the preferred format for the molecular structure of the repeating unit.

5.4.3 Polymer class

Table 155 — Polymer class

User Guidance	Type of polymer.
Example(s)	homopolymer, copolymer, crosspolymer
Conformance	MANDATORY
Data Type	CD

5.4.4 Polymer geometry

Table 156 — Polymer geometry

User Guidance	The geometry of the polymer shall be described.
Example(s)	linear, branched, cross linked, network, dendritic, star, comb, brush
Conformance	MANDATORY
Data Type	CD

5.4.5 Copolymer sequence type

Table 157 — Copolymer sequence type

User Guidance	The type of copolymer of the polymer sequence shall be described
Example(s)	random, alternating, block, graft, statistical
Conformance	MANDATORY
Data Type	CD

5.4.6 Comment

Table 158 — Comment

User Guidance	Any comment can be provided in this field, if necessary.
Conformance	OPTIONAL
Data Type	ST

5.4.7 Monomer description (repeat as necessary)

For synthetic polymers, this set of descriptors intends to specify and quantify the monomers used for the synthesis of the polymer or copolymer. The information on the monomer shall be provided by means of the following data elements:

5.4.7.1 Number of monomers

Table 159 — Number of monomers

User Guidance	The number of diverse monomers used to synthesise the polymer shall be specified.
Example(s)	2
Conformance	MANDATORY
Data Type	INT
Business Rule(s)	This number can be counted and is never explicitly represented.

5.4.7.2 Amount type

Should be captured according to 4.7.12.3.1. Indicate how the quantitative amount of Structural Repeating Units is captured (e.g. mole ratio, weight ratio, mole percent, weight percent).

5.4.7.3 Monomer (repeat as necessary)

This section identifies and quantifies the monomer(s) used in the synthesis of the polymer.

5.4.7.3.1 Monomer ID

Table 160 — Monomer ID

User Guidance	The unique identifier assigned to the monomer.
Conformance	MANDATORY
Data Type	II

5.4.7.3.2 Monomer name

Table 161 — Monomer name

User Guidance	The name of the monomer shall be provided; established or primary name of monomer.
Conformance	MANDATORY
Data Type	CD

5.4.7.3.3 Amount

The information related to the amount shall be provided as per specification described in 4.9.

5.4.8 Structural repeat (repeat as necessary)

This section specifies and quantifies the repeated units and their configuration. A structural representation should also be submitted. Information on the structural repeat unit shall be provided by means of the following data elements:

5.4.8.1 Number of structural repeat units

Table 162 — Number of structural repeat units

User Guidance	The number of diverse structural repeated units represented in the structure of the polymer shall be specified.
Example(s)	Homopolymers will only have 1 Structural Repeat Unit; copolymers will typically have more than 1.
Conformance	MANDATORY
Data Type	INT
Business Rule(s)	Implicitly derived

5.4.8.2 Amount type

Should be captured according to 4.7.12.3.1 .Indicate how the quantitative amount of Structural Repeating Units is captured (e.g. mole ratio, weight ratio, mole percent, weight percent).

5.4.8.3 Structural repeat unit (repeat as necessary)

For synthetic polymers the structural repeat units are typically generated from the polymerisation of monomers. The Structural Repeat Unit for polyethylene is ethylene and not methylene.

5.4.8.3.1 Orientation of polymerisation

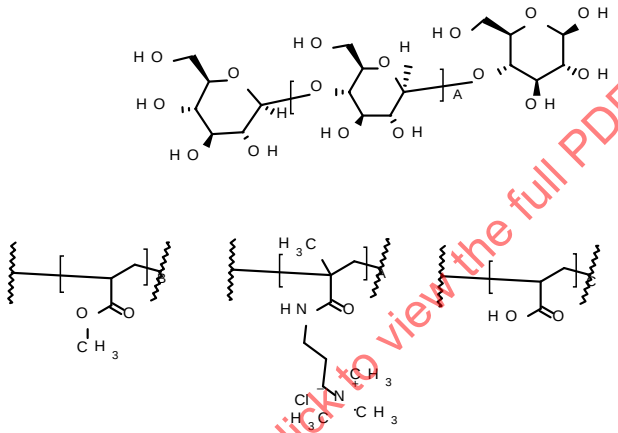
This element is used to indicate how Structural Repeat Units connect to each other. In addition to the Orientation of Polymerisation explicit connections between Structural Repeating Units can be indicated by connection points.

Table 163 — Orientation of polymerisation

User Guidance	The orientation of the polymerisation shall be described.
Example(s)	head-tail, head-head, random
Conformance	MANDATORY
Data Type	CD

5.4.8.3.2 Structural repeat unit

Table 164 — Structural repeat unit

User Guidance	Structural repeat units are essential elements for defining polymers. Specific guidance will be given in the polymer annex
Example(s)	
Conformance	MANDATORY
Data Type	Molfile; SMILES; InChI
Business Rule(s)	End groups if defined can be attached to the Structural Repeat Unit. The Structural Repeat Unit should be the largest repeating structural fragment

5.4.8.3.3 Amount

The information related to the amount shall be provided as per specification described in 4.9.

5.4.8.3.4 Degree of polymerisation (repeat as necessary)

This section applies to homopolymer and block co-polymers where the degree of polymerisation within a block can be described. The following information shall be provided where available:

a) Polymerisation degree

Table 165 — Polymerisation degree

User Guidance	The type of the degree of polymerisation shall be described.
Conformance	MANDATORY
Data Type	CD

b) Amount

The information related to the amount shall be provided as per specification described in 4.9.

5.4.9 Molecular weight (repeat as necessary)

The information related to the molecular weight shall be provided as per specification described in 5.2.6.

5.4.10 Property (repeat as necessary)

The information related to the property of the polymer shall be provided as per specification described in 4.11.

5.4.11 Reference source (repeat as necessary)

The information related to the source of the property of the polymer shall be provided as per specification described in 4.6.

5.5 Structurally diverse substances

There are a wide variety of substances described as structurally diverse substances. Structurally diverse substances are typically substances that have a variety of structurally diverse constituents that cannot be captured as a mixture of a limited number of related single substances. This category would be used to describe the following class of substances:

- Herbals -- The herbal preparation OIL is discussed in the Chemical Annex B, B.5.1 to B.5.3. The Herbal Preparation, Extract, exudates as well as the (Herbal) Substance (fresh), Herbal Drug are discussed in the Herbal Annex E. and will be addressed in the next version of this Technical Specification.
- Homeopathic Substances -- To be addressed in the next edition of this Technical Specification.
- Vaccines — To be addressed in the next edition of this Technical Specification.
- Complex blood products — To be addressed in the next edition of this Technical Specification.
- Polyclonal Immunoglobulins — To be addressed in the next edition of this Technical Specification.
- Cells, tissues, complex animal derived products — To be addressed in the next edition of this Technical Specification.
- Minerals — To be addressed in the next edition of this Technical Specification.

For organism-based substances, the parent organism is essential defining information.

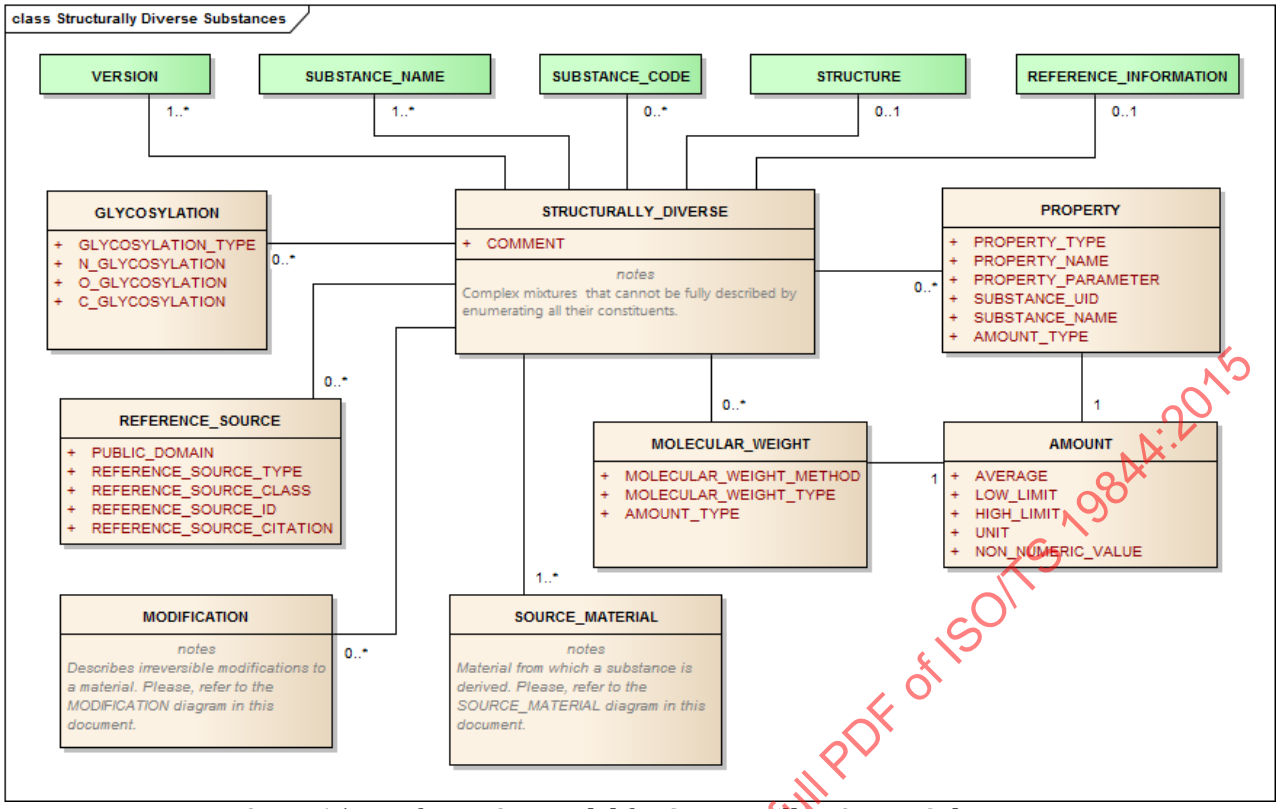


Figure 15 — Information Model for Structurally-Diverse Substance

The majority of these substances will be derived from a biological organism, but they could be complex natural materials such as coal tar or mineral oil.

5.5.1 Herbals and substances used in the preparation of plant-based allergenic extracts

Substances in herbal preparations and most substances in allergenic preparations shall be defined simply by taxonomic information and the plant part. Taxonomic information, particularly the scientific name of a medicinal plant, is essential for defining herbal substances in medicinal products. The Kew Gardens Medicinal Plant Names Services (MPNS) should be the authoritative source for all scientific names and should also maintain a list of plant parts and fractions. A controlled vocabulary for medicinal plant taxonomy has been developed and will be maintained by the Kew Gardens Medicinal Plant Names Services. A submitter should choose the current accepted scientific name provided by MPNS. The MPNS Portal should be accessed via www.kew.org/mpns. The service should be contacted if the herbal substance material cannot be mapped to records in the MPNS. It should be noted that many common and even pharmacopoeia names of herbs often map to more than one plant species, but it is important that the actual herbal substance (defined as a particular species) shall be used in defining a substance.

Although herbal products in some jurisdictions are not regulated as medicinal products, they are a major class of products covered in the IDMP standards. Many jurisdictions have a clear distinction between herbal substances and herbal preparations and that is what is followed in this implementation guide. herbal substances are defined by the taxonomic definition of the plant represented by the plant's scientific name which includes the genus, species and author. Additional taxonomy such as kingdom, phylum, class order, and family could also be captured to assist in the classification of substances. The part or parts or the plant that are used in the preparation are also included in the substance *definition*. Identification of a substance may also require an indication of the plant life cycle captured in the name of the plant part such as inflorescence or sprout.

EXAMPLE The common term chamomile typically refers to the flowers of two distinct plants identified by the accepted scientific names *Matricaria chamomilla* L. and *Chamaemelum nobile* (L.) All. The term chamomile is ambiguous and a submitter should refer to the specific species that is contained in the product.

Time and place of harvest, type of soil, the use of fertilisers, pesticides and herbicides, the amount of daylight and water, although important they are not captured as part of the substance definition. A cultivar or variety of a plant

may be defined as a different substance if known differences in constituents exist (e.g. broccoli and cauliflower are defined as different substances even though they share the same genus and species). herbal preparations are considered fractions of the herbal substance will be captured at the substance level for oils, and to some extent for extracts. Additional information of the extract will be defined at the specified substances Group 1 information level. Fractions of plants parts such as commodity oils, juices, and exudates such as acacia gum and herbal extracts such as teas and tinctures and will always map to the parent organism and part or parts. A variety of examples of herbal substances and herbal preparations are present in Annex E: herbal substances and herbal specified substances.

The herbal Annex E should be referred to for user guidance and business rules for the various elements of the source material for structurally diverse substances as they apply to herbals. Examples are given for herbals derived from plant material, as well as material from algae and fungi, which are also classified as herbals. Other types of natural materials, such as animals and minerals are also treated as herbals in Kampo extracts and other systems of medicine. Although Annex E doesn't include any examples, algae, fungi, animals and minerals are also classified as herbals.

5.5.1.1 Naming vegetable oils

The business rules for naming of vegetable oils are described in ISO/TS 19844 Section Structurally Diverse Substance, herbals, 5.5.1 and in the section Source material (4.10).

NOTE The information model for Structurally Diverse substances has to be updated in the Main body document of the ISO/TS 19844 as by the latest version and in the ISO 11238 standard. Since in the next version of the Implementation Guide a proposal will be made to include both the information for the herbal drug (a term used here in preference to herbal substance) and the herbal preparation at the substance information level. Additional herbal drug and herbal preparation information can be included in the Substance_type Structurally Diverse specified substance Group 1 information level. This is needed to capture the marker and signature substances at this level in order to comply with the Regulatory regulation.

The terms Substance and specified substance are used as defined by ISO 11238. The terms herbal drug (herbal substance) and herbal preparation are used as defined by pharmacopoeias, and will both be captured at the Substance level. In addition, a more basic unit, comprising unprocessed, fresh material of a single species or infraspecies (occasionally a genus) + part, with no affiliation to any particular pharmacopoeia, will also be captured at the Substance level; for clarity, these will be referred to as Substance (fresh) in this Annex, The information model also accommodates fraction information at the substance level. The element <Fraction> will be considered as two herbal preparation_Types: <Oil, Exudate, Juice> or <Extract>, each requiring different information.

Specified substance group 1 includes additional information for substances and multi-substance material. Additional information for herbal drugs and for herbal preparations e.g. additional information for an extract such as extraction solvent composition, herbal Drug/ herbal extract ratio will be captured at this level along with the constituents and marker or signature substances.

5.5.1.1.1 Example: Olive Oil, Virgin

Olive Oil, Virgin EP is defined as a herbal preparation (because it has been obtained by expression). There is no defined herbal drug for this herbal preparation, but instead its source material is the ripe drupes (fruit) of Olea europaea L.

For the purposes of the ISO 11238, the herbal preparation is defined as a Substance as [<Scientific genus/binomial/trinomial **without Author**>,<Part(s)>,<Fraction>], i.e. Olea europaea, Fruit, Oil.; Common name: Olive Oil, Virgin. The source material for the herbal preparation, with the **relationship** of **Parent substance**, giving the Substance_Role the value <Parent>, is a separate Substance (fresh) defined as [<Scientific genus/binomial/trinomial with Author>,<Part(s)>], i.e. Olea europaea L., Fruit.

The name [<Scientific genus/binomial/trinomial **with Author**>,<Part(s)>] refers to a **Substance Name** in terms of the ISO 11238 that is not an official name in any pharmacopoeia or other legislation. The inclusion of the author is to ensure that it is not confused with the official names for herbal drugs and herbal preparations which are also defined as Substances by the ISO 11238. The scientific element of the name is usually a binomial (species), but

may be a trinomial (infraspecies such as a variety or subspecies) or a genus, as appropriate, together with the author of the name. The Part element of the name should be chosen from a CV consisting of a restricted number of terms.

The **name** [<Scientific genus/binomial/trinomial **without Author**>,<Part(s)>] refers to a **Substance Name** in terms of the ISO 11238 and is the equivalent to a herbal drug Name. There is no such Substance Name for Olive Oil in this example from the EP, although other pharmacopoeias might define a herbal drug for Olea europaea, Fruit.

The **name** [<Scientific genus/binomial/trinomial without Author>,<Part(s)>], [Fraction] refers to a **Substance Name** in terms of the ISO 11238 and is **equivalent** to an **herbal preparation Name**.

The herbal preparation Name has two herbal preparation Name_Types; **one** for the Substance fraction: Oil, Exudate, Juice and one for the Substance fraction: Extract. This is done to differentiate between the cardinalities of the herbal preparation_Types.

The herbal preparation-Substance (fresh) is a two way relationship. The herbal preparation_Type <Oil, Exudate, Juice> is prepared from one Substance (fresh).

NOTE Several herbal preparations, e.g. Juices, can be a mixture of separately obtained juices obtained from the herbal substances and mixed after expression. The final preparation is then a multi-substance material which is captured at the Specified substance Group 1 level.

The herbal preparation_Type <Extract> can be prepared from **one or more** herbal drugs and by **one or more** extraction steps and by **one or different** extraction solvent compositions. Only the first extraction solvent composition is used in the element solvent. Further details of the herbal preparation and herbal drug are described in the ANNEX E, Structurally Diverse substances, herbals and for the Oils the information is provided the Chemical ANNEX B.

For the herbal preparation Olive Oil, Virgin the result of the information model will read:

Substance_ID: HJFTE78543 (Artificial ID); (UNII: 6UYK2W1W1E)
Substance Name (herbal preparation name): Olea europaea, Fruit, Oil
Substance Name_Type: Other
herbal preparation_Type: Oil 1)
Substance Name: Olive Oil, Virgin
Substance Name_Type: Official (for further elements see tabular format in the Chemical Annex B.)
Parent Substance_ID: OLIJF5643S
Parent Substance_Name: Olea europaea L., Fruit

5.5.1.2 Substance type

Table 166 — Substance type

User Guidance	Information about the substance type as described and in accordance with ISO 11238 definitions
Example(s)	Structurally Diverse
Conformance	MANDATORY
Data Type	CD
Values Allowed	Structurally Diverse

5.5.1.3 Substance ID

Table 167 — Substance ID

User Guidance	ID to be used in all electronic submissions to identify a substance. Generated when sufficient information is available to unambiguously define a substance. ID will be permanently associated with a given substance and each substance at the substance level shall have one and only one ID. NOTE If a unique "Substance ID" has been assigned, this "Substance ID" shall be specified based on the Substance Name controlled vocabulary. In the absence of a unique "Substance ID", e. g. for the initial submission of the substance, this data element is not mandatory.
Example(s)	DEVTYS543H
Conformance	MANDATORY
Data Type	II
Values Allowed	Value could be a code associated with a preferred term, or a specific type of data. If a code is transmitted the preferred term associated with that code should also be transmitted.
Business Rule(s)	All substances will be identified by a single ID. NOTE: the Substance ID is NOT SPECIFIED for Substance ID Request NOTE: The ID will only be released to the public if the defining information is in the public domain or if a company that provides the defining information requests public release or releases the code in public marketing materials. Defining information found in patents will usually not be sufficient to release the ID to the public. Even if an ID can be released there may be elements that will not be released to the general public. An ID will always be released to an organisation that requests an ID and supplies information to define a substance. A Flag to control the release of the ID to the general public is part of the Reference Source information (4.6).

5.5.1.4 Substance name

Table 168 — Substance name

User Guidance	Name or company code associated with the Substance. A Substance is always a single species, or infraspecies, or is occasionally a single genus. It will have a Name which is composed of: [Scientific genus/binomial/trinomial with Author] [Part(s)] The scientific element of the name is usually a binomial, but may be a trinomial or a genus, as appropriate, together with the author of the name. The Name is not an official name in any pharmacopoeia or other legislation, and the inclusion of the author is to ensure that it is not confused with the official names for specified substances that are derived from a single species substance. It can be written in any language and its jurisdiction can be restricted using Name_Status.
Example(s)	*Harpagophytum procumbens* (Burch.) DC. ex Meisn., Root
Conformance	MANDATORY
Data Type	ST
Values Allowed	Free text, multiple names allowed, each in its own <asNamedEntity> element.
Business Rule(s)	Mandatory all substances shall have at least one name or company code associated with the substance.

5.5.1.5 Substance name type

Table 169 — Substance name type

User Guidance	Each name shall be associated with a type.
Example(s)	Official name, systematic name, generic name, brand name, company code
Conformance	MANDATORY
Data Type	CD
Values Allowed	OFFICIAL, SYSTEMATIC, COMPANY CODE, OTHER
Business Rule(s)	A name shall have one and only one name type.

5.5.1.6 Language

Table 170 — Language

User Guidance	If the name is language dependent, that language shall be specified. Company codes are not language dependent and no language should be specified.
Example(s)	en — English, de — German, fr –French
Conformance	MANDATORY
Data Type	CD
Values Allowed	ISO 639-1, alpha-2 codes
Business Rule(s)	Multiple codes assigned to the same language are to be considered synonyms and only one of them needs to be specified. While xml:lang and JSON allows country-specification of language, e.g. en_US vs. en_UK or pt_PT vs. pt_BR, such specificity is to be avoided except if the country variant of the language exceptionally does give rise to an alternative spelling. In that case, only the exceptional spelling needs to be tagged with the country-specific language code, not the regular spelling for that language in all other countries.

5.5.1.7 Official name type

Table 171 — Official name type

User Guidance	Designation of which authority assigned the Official Name. All official names need to have at least one such designation. The name type is the name of the authority or authorities that have assigned or have adopted the name.
Example(s)	BAN, COSING, EP, FCC, INCI, INN, JAN, JECFA, MARTINDALE, USAN, USP.
Conformance	MANDATORY — all official names require a designation of naming authority
Data Type	CD
Values Allowed	All values in the name assigning authority identifier system.
Business Rule(s)	All official names must have a naming authority designated. If multiple authorities have assigned the same name, a separate <asNamedEntity> is specified for each authority that needs to be listed.

5.5.1.8 Reference sources

5.5.1.8.1 Public domain

Table 172 — Public domain

User Guidance	Indication of whether the source is in the public domain. Company codes will only be released if the defining information is associated with the code in a public source such as Chemical Abstracts (CAS), a journal article, or a poster at scientific meeting attributable to the company.
Example(s)	Yes, No
Conformance	OPTIONAL
Data Type	BL

5.5.1.8.2 Reference source type

Table 173 — Reference source type

User Guidance	The reference source type in which the data elements were actually found.
Example(s)	BP, CHEMID (NLM), EP, IND, INN, ITIS, JAN, JOURNAL, JP, KEGG, MARTINDALE, NDA, PERSONAL CARE PRODUCTS COUNCIL (PCPC), PUBCHEM, USAN, USP, WEB PAGE
Conformance	MANDATORY
Data Type	CD
Business Rule(s)	Mandatory all names should have at least one source.

5.5.1.9 Source material

5.5.1.9.1 Source material class

Table 174 — Source material class

User Guidance	General classification of the source material;
Example(s)	Organic
Conformance	MANDATORY
Data Type	CD

5.5.1.9.2 Source Material Type

Table 175 — Source material type

User Guidance	The type of the source material shall be specified based on controlled vocabulary
Example(s)	fungus, plant, animal
Conformance	MANDATORY
Data Type	CD

5.5.1.9.3 Organism ID

Table 176 — Organism ID

User Guidance	The unique identifier associated with the plant material shall be specified.
Conformance	MANDATORY
Data Type	II

5.5.1.9.4 Organism name

Table 177 — Organism name

User Guidance	The organism accepted Scientific name shall be provided based on the genus, binomial or trinomial.
Example(s)	*Harpagophytum procumbens* (Burch.) DC. ex Meisn.
Conformance	MANDATORY
Data Type	CD
Business Rule(s)	The accepted scientific name for plants and fungi is provided by Kew's Medicinal Plant Names Services (MPNS). The names for other organisms will be provided by other authoritative sources.

5.5.1.9.5 Development stage

Table 178 — Development stage

User Guidance	Stage of life for plants. This information shall be provided only when the substance is significantly different in these stages.
Example(s)	leafing, pre-flowering, flowering, fruiting, etc.
Conformance	OPTIONAL
Data Type	CD
Business Rule(s)	If it is a distinguishing factor, even if it is implicitly understood to experts.

5.5.1.10 Organism (repeat as necessary)

This section describes the plant which the substance derived from.

5.5.1.10.1 Kingdom

Table 179 — Kingdom

User Guidance	The kingdom of a plant shall be specified.
Example(s)	Plantae
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	This can be derived from the Organism name as referenced in the appropriate terminology / taxonomy (e.g. the Kew's Medicinal Plant Names Services (MPNS), NCBI Entrez Taxonomy or the Catalog of Life.). Catalog of Life typically capture complete taxonomy in a manner consistent with this standard.

5.5.1.10.2 Phylum

Table 180 — Phylum

User Guidance	The phylum of a plant shall be specified.
Example(s)	Tracheophyta
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	This is implied by the organism ID and name as referenced in the appropriate terminology / taxonomy (e.g. the NCBI Entrez Taxonomy ID implies all higher levels.)

5.5.1.10.3 Class

Table 181 — Class

User Guidance	The class of the plant shall be specified.
Example(s)	Magnoliopsida
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	This is implied by the Organism ID and name as referenced in the appropriate terminology / taxonomy (e.g. the NCBI Entrez Taxonomy ID implies all higher levels.)

5.5.1.10.4 Order

Table 182 — Order

User Guidance	The order of the plant shall be specified,
Example(s)	Lamiales
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	This is implied by the Organism ID and name as referenced in the appropriate terminology / taxonomy (e.g. the NCBI Entrez Taxonomy ID implies all higher levels.)

5.5.1.10.5 Family

Table 183 — Family

User Guidance	The family of the plant shall be specified.
Example(s)	Pedaliaceae
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	This is implied by the Organism ID and name as referenced in the appropriate terminology / taxonomy (e.g. the NCBI Entrez Taxonomy ID implies all higher levels.)

5.5.1.10.6 Genus

Table 184 — Genus

User Guidance	The genus of the organism shall be specified; refers to the genus element of the plant/animal scientific name; it is present in names for genera, species and infraspecies.
Example(s)	Harpagophytum
Conformance	MANDATORY
Data Type	CD
Business Rule(s)	This is implied by the Organism name as referenced in the appropriate terminology / taxonomy. For plants, it will be the current accepted genus name provided by Medicinal Plant Names Services.

5.5.1.10.7 Species

Table 185 — Species

User Guidance	The species of the organism shall be specified; refers to the scientific epithet of the species of the plant/animal; it is present in names for species and infraspecies.
Example(s)	Procumbens
Conformance	MANDATORY
Data Type	CD
Business Rule(s)	This is implied by the Organism name as referenced in the appropriate terminology / taxonomy. For plants, it will be the current accepted species name provided by Medicinal Plant Names Services.

5.5.1.11 Part description (repeat as necessary)

5.5.1.11.1 Part

Table 186 — Part

User Guidance	The portion of a plant with a definable anatomical location.
Example(s)	Tuburous Root
Conformance	MANDATORY
Data Type	CD
Business Rule(s)	A controlled vocabulary for plant parts is maintained by Kew's Medicinal Plant Names Services.

.

5.5.1.11.2 Part location

Table 187 — Part location

User Guidance	The location of the part in the plant with a definable anatomical location.
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	If recognised by the expert to be relevant for the definition of the substance, it shall be provided.

All the other elements related to the Organism class as referenced in ISO 11238 are optional for the herbal substance description.

5.5.1.12 Version (repeat as necessary)

Version associated with the substance will be captured as described in 4.13.

5.5.2 Vaccines - Annex addressing this will be included in the next edition of this Technical Specification

Most vaccines will be described as structurally diverse substances. For viral and many bacteria vaccines the taxonomic description will nearly always include infraspecific information. The strain and/or serovar should be captured for all of the vaccines. The state of the virus or bacteria shall also be captured along with the strain.

EXAMPLE Many vaccines are live attenuated or inactivated and this should be captured.

Some organisms such as Influenza A virus require identification of subspecies, variety, strain, serovar, type, form, or cultivar group to accurately describe them. Relevant taxonomic identification numbers are helpful in parent organism identification.

5.5.3 Purified blood products and polyclonal antibodies — an Annex addressing this will be included in the next edition of this Technical Specification

Purified blood products (distinct clotting factors, human serum albumin) and monoclonal immunoglobulins are described as proteins. Polyclonal immunoglobulins are described as structurally diverse materials and require identification of the immunoglobulin type and targeted antigen if applicable.

5.5.4 Cells and tissues — an Annex addressing this will be included in the next edition of this Technical Specification

Cells and tissues are also described as structurally diverse substances. Information on individual donors or extent of pooling is not captured at the substance level. Information on a particular cell line shall be captured as infraspecific information.

Many natural substances are modified chemically, physically, or biologically. A new substance will be generated if such a modification changes the chemical structure of one or more chemical entities in the substance. A polysaccharide conjugate vaccine is described as a component of an organism part chemically conjugated to another substance or substances. The type of conjugation chemistry and the identity of the conjugated chemical entity or entities are needed to describe the resulting structurally diverse conjugate. When a genetically-modified organism or cell is used, the inserted gene biological genetic modifications must be described.

5.5.5 Minerals — an Annex addressing this will be included in the next edition of this Technical Specification

Structurally-diverse substances not derived from organisms often require property descriptions as defining characteristics. In this way, light mineral oil is distinguished from mineral based on density or viscosity while petroleum distillates require a boiling range. These natural mixtures may be chemically, physically, or biologically modified prior to use so information about modifications may also be necessary to define them.

In order to assign a Substance ID for other structurally-diverse substances (non-organism derived), the following information shall be provided:

- Source material (e.g. coal, petroleum)
- Physical properties related to molecular structure of the substance or ensemble of substances (i.e. viscosity, density, etc.)
- Any process that fractionates or modifies the source chemically

5.6 Mixture substance (repeat as necessary)

Material that contains multiple substances can be mixture substance if the substances are either isolated or synthesised together. Racemic mixtures or substances containing unknown or mixed stereochemistry will not be defined as mixtures, but will be represented as substances that contain impurities or degradants. Mixtures of mixtures will not be allowed. Each component of a mixture should be listed. Substances present in trace amounts will not be listed in a mixture unless they are known to have a specific effect. Mixtures are also used when substance ambiguity exists in authoritative sources (e.g. aloe).

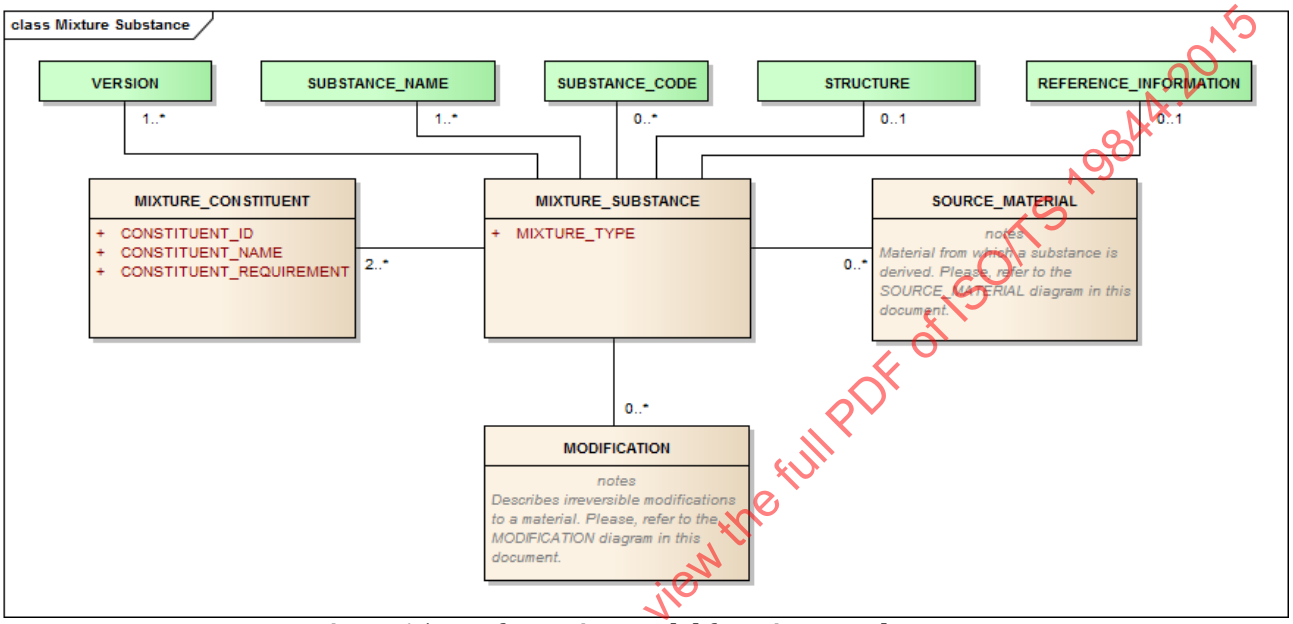


Figure 15 — Information Model for mixture substance

The following information on mixture shall be provided:

5.6.1 Mixture type

There are basically three potential mixture types. Most mixtures will be “all of” mixtures meaning that all of the constituents are present in the mixture. The other two types of mixture are “one of” or “any of”. “Any of” mixtures will mean that at least one of the constituents must be present but other constituents may or may not be present.

Table 188 — Mixture type

Example(s)	ALL OF; ONE OF; ANY OF
Conformance	MANDATORY
Data Type	CD

5.6.2 Mixture constituent (repeat as necessary)

The mixture constituent set of elements aims to describe the constituents of the mixture substance. Such information shall be provided by means of the following data elements:

5.6.2.1 Constituent ID

Table 189 — Constituent ID

User Guidance	The unique identifier assigned to the constituent substance. NOTE: If a unique "Substance ID" has been assigned, this "Substance ID" shall be specified based on the Substance Name controlled vocabulary. In the absence of a unique "Substance ID" e. g. for the initial submission of the substance this data element is not required.
Conformance	CONDITIONAL
Data Type	II

5.6.2.2 Constituent name

Table 190 — Constituent name

User Guidance	The name of the constituent of the mixture shall be described: established or primary substance name. Teicoplanin A_{2-1} Is one of the five constituents of the mixture of glycopeptides Teicoplanin produced by certain strains of Actinoplanes teichomyceticus sp.
Example(s)	Teicoplanin A_{2-1}
Conformance	CONDITIONAL
Data Type	ST

5.6.2.3 Constituent requirement

Table 191 — Constituent requirement

User Guidance	A flag indicating if this component is required.
Example(s)	Always present, may be present.
Conformance	OPTIONAL
Data Type	CD

6 Specified substance (Optional)

There are a number of regulatory needs when defining materials in medicinal products. The globalisation of supply chains, widespread contract manufacturing and the high value of many medicinal products places a much greater burden on both regulatory agencies and companies to ensure that the material used in medicinal products is correctly identified with known pedigree.

Although the substance model captures information essential to the description of materials in medicinal products there is often a strong regulatory need for additional information that is not captured at the substance level. The four groups of specified substance elements allow for the explicit capture of information essential for the evaluation and tracking of material used in medicinal products. Each of the four groups of elements provides information essential for these regulatory needs in a manner that should facilitate compliance. The implementation of the specified substance Groups is optional. Should a region implement any of the specified substance Groups, the following conformances as described in this section are applicable.

6.1 Specified substance Group 1 (repeat as necessary)

Multiple substance materials (e.g. simethicone, aluminium lakes, and flavours), herbal extracts, juices, oils and tinctures and polymorphic forms of materials will be described and differentiated using specified substance Group 1 elements.

Specified substances Group 1 will always have at least one constituent; all constituents will either be substance or specified substance Group 1. The modification group elements and fraction elements are also necessary to define many specified substances Group 1. For herbal preparations, the drying process could be considered as a modification process. An extraction is a preparation. There can be a liquid extract or dry extract being made from the (herbal) Substance (fresh) or from the (herbal) Substance (dry) which is the herbal Drug in terms of the Pharmacopoeia.

Single component substances that differ in physical form will be defined as different specified substances Group 1. For example crystalline insulin and amorphous insulin are two different specified substances Group 1. To avoid confusion, if substance information is sufficient for regulatory needs, the substance information and ID should be used to describe the medicinal product.

The information model for specified substances Group 1 is listed in Figure 16.

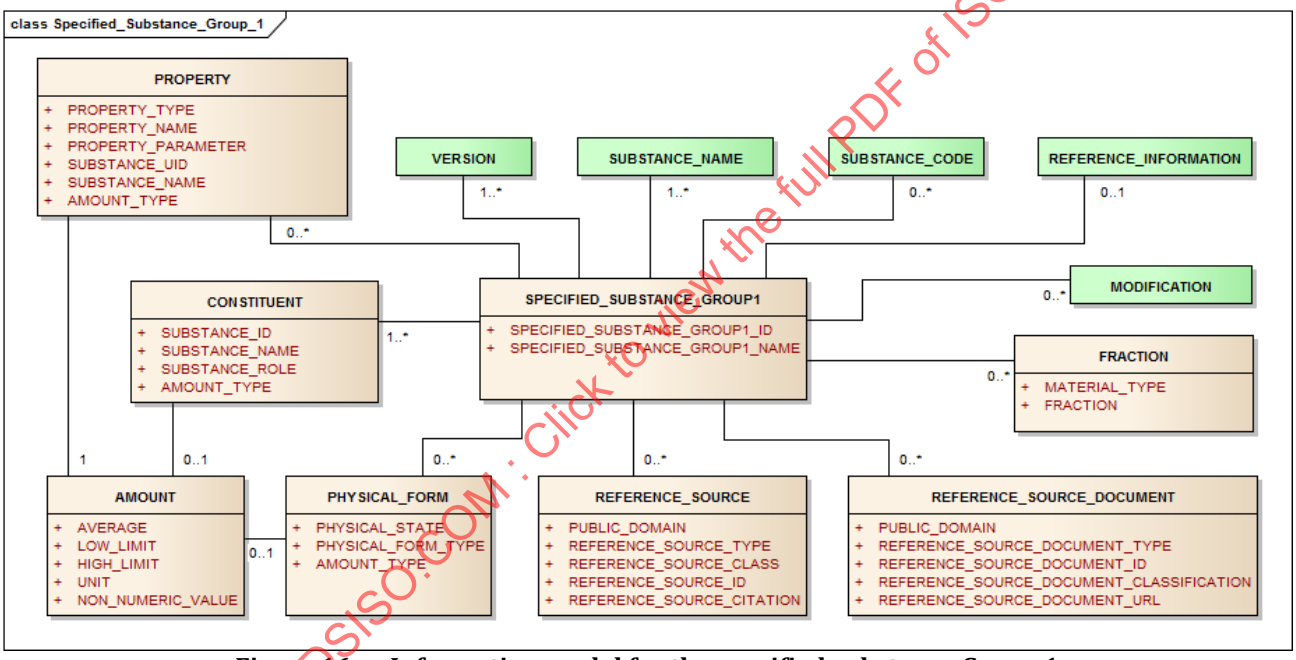


Figure 16 — Information model for the specified substance Group 1

6.1.1 Specified substance Group 1 ID

Table 192 — Specified substance Group 1 ID

User Guidance	The unique identifier assigned to the specified substance Group 1 shall be specified. NOTE If a unique "specified substance ID" has been assigned, this "specified substance ID" shall be specified based on the Substance Name controlled vocabulary. In the absence of a unique "specified substance ID" e. g. for the initial submission of the substance this data element is not required.
Conformance	MANDATORY
Data Type	II
Business Rule(s)	The ID of the substance will be automatically assigned by the system once the message will be processed. NOTE: the specified substance ID is NOT SPECIFIED for specified substance ID Request

6.1.2 Specified substance Group1 Name

Table 193 — Specified substance Group 1 name

User Guidance	The name of the specified substance Group 1 shall be provided in this field. [Scientific genus/binomial/trinomial **without** Author] [Part(s)] [Fraction][Extraction solvent composition][herbal Drug/Native herbal Preparation(DER)]
Example(s)	*Ginkgo biloba,* Leaf, Dry Extract acetone-Water (60 — 40)(65-37 = 1 w,w)
Conformance	MANDATORY
Data Type	ST
Business Rule(s)	Name conventions can be adopted regionally as the latest version of the Naming Conventions document.

6.1.3 Substance Name (repeat as necessary)

Names associated with a specified substance Group 1 will be captured in manner similar to Substance Name (4.5). The name type should be specified and, for Official Names, the same format should be followed (4.5.4).

6.1.4 Substance Code

Codes associated with a specified substance Group 1 will be captured in manner similar to Substance Codes (4.7).

6.1.5 Version (repeat as necessary)

Version associated with a specified substance Group 1 will be captured in manner similar to Substance Version (4.13).

6.1.6 Reference sources

Reference information will also be captured in a manner similar to the substance reference information (4.6).

6.1.7 Property (repeat as necessary)

Property values will be captured in a manner similar to those described for substances (Clause 4). It is possible to capture both definitional properties and properties that provide additional information about a substance. For

example the melting point of a polymorphic substance or solubility or rate of dissolution of a crystalline material could be a defining property of a given specified substance Group 1.

6.1.8 Fraction (new class to be included in the second edition of ISO 11238)

For herbal substances the fraction element should be used to capture oils and juices derived from specific parts of plants.

Table 194 — Fraction

User Guidance	The fraction of the plant where the herbal substance is derived.
Example(s)	Oils, juices, essential oil
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	When the herbal substance is derived from a fraction of the plant, the description of the fraction is required.

6.1.9 Modification

Modifications, particularly those involving herbal preparations, should be captured at the specified substance Group 1 level when they are not captured at the substance level (4.11).

EXAMPLE The refinement of olive oil to remove or reduce free acids would be a modification captured at the specified substance Group 1 level. Partial or complete hydrogenation of given oil would also be captured at the specified substance Group 1 level.

6.1.10 Reference Information (repeat as necessary)

The Reference Information is not essential for the definition of the substance. If provided, the conformance as described in 4.7.7.1.

6.1.11 Constituent (repeat as necessary)

The constituents group of elements serves several roles, each substance in multiple substance materials will be captured as a component. Signature, active markers, and limit substances and extraction solvents will also be captured for herbal extracts. The amount of components, constituents and extraction solvents will also be captured. A different specified substance Group 1 will be created if these amounts consistently vary.

At least one constituent is necessary for every specified substance Group 1 (i.e. the parent Substance ID).

6.1.11.1 Substance ID

Each constituent should be identified by either a Substance ID or a specified substance ID.

Table 195 — Substance ID

User Guidance	Every specified substance Group 1should have at least one constituent and the Substance ID for each constituent should be captured.
Example(s)	ISO 11238 Substance or specified substance ID
Conformance	MANDATORY
Data Type	II
Values Allowed	ISO 11238 Substance or specified substance ID
Business Rule(s)	A constituent can be a substance or a specified substance Group 1.

6.1.11.2 Substance name

Table 196 — Substance name

User Guidance	The name of the substance which is the constituent of the specified substance Group 1 shall be described; Preferred or Official Name of the constituent.
Conformance	MANDATORY
Data Type	CD
Values Allowed	ISO 11238 Substance or specified substance name
Business Rule(s)	This value is implicit and derived from the Substance ID. The Preferred name is the default value.

6.1.11.3 Substance role

Table 197 — Substance role

User Guidance	The role of the substance in the specified substance shall be described based on a controlled vocabulary. An impurity component is not described in the specified substance Group 1. The parent ID of a specified substance Group 1shall always be specified.
Example(s)	active marker, component, degradant, metabolite, active moiety, parent substance
Conformance	MANDATORY
Data Type	CD
Values Allowed	COMPONENT, EXTRACTION SOLVENT, SIGNATURE SUBSTANCE; MARKER SUBSTANCE

6.1.11.4 Amount type

As in 4.7.12.3.1.

6.1.11.4.1 Amount

The information related to the amount shall be provided as per specification described in 4.9.

6.1.12 Physical form (repeat as necessary)

This section is to capture information on the physical form of the specified substance (e.g. crystalline amorphous, tincture, dry extract), the state of matter, and slightly more detailed form of the final specified substance (e.g. solid, liquid, gas, or emulsion). Biphasic insulin is one example of a crystal in which part is dissolved and part is in solution. There are also instances of substances being partially crystalline and partially amorphous which would be two different form types and a mixture of specified substances, different polymorphic crystalline types. If two polymorphs are distinguished by crystalline type, the symmetry group that distinguishes the crystalline types should be captured as a property.

The Physical form is not always applicable (e.g. for multiple substances) and therefore is conditional; however when it is applicable (e.g. herbal preparation or when the form of the substance is available), the physical form is required to be provided in accordance with the following specifications:

6.1.12.1 Physical state

Table 198 — Physical state

User Guidance	The actual state of the substance in packaged product or administered product
Example(s)	solid, liquid
Conformance	MANDATORY
Data Type	CD
Values Allowed	SOLID, LIQUID, GAS, EMULSION, GEL
Business Rule(s)	This information is implicit and derived from the Physical Form Type

6.1.12.2 Physical form type

Table 199 — Physical form type

User Guidance	Type of the physical form
Example(s)	crystalline, amorphous, tincture, dry, extract
Conformance	MANDATORY
Data Type	CD
Business Rule(s)	When the physical state is available, it shall be provided as it defines a new Substance ID

6.1.12.3 Amount type

In 4.7.12.3.1.

6.1.12.4 Amount

The information related to the amount shall be provided as per specification described in 4.9.

When there is an amount, it must be specified as an observation with a measurable property code.

6.2 Specified substance Group 1 intended for herbal substance and herbal preparation

This iteration of the document covers specifically herbal specified substance Group 1.

6.2.1 Specified substance Group 1 ID

Table 200 — specified substance Group 1 ID

User Guidance	The unique identifier assigned to the specified substance Group 1 shall be specified. NOTE If a unique "specified substance ID" has been assigned, this "specified substance ID" shall be specified based on the Substance Name controlled vocabulary. In the absence of a unique "specified substance ID" e. g. for the initial submission of the substance this data element is not required.
Example(s)	UFR78E23Y
Conformance	MANDATORY
Data Type	II
Business Rule(s)	The ID of the substance will be automatically assigned by the system once the message will be processed. NOTE: the specified substance ID is NOT SPECIFIED for specified substance ID Request

6.2.2 Specified substance Group1 Name

Table 201 — specified substance Group 1 name

User Guidance	The name of the specified substance Group 1 shall be provided in this field as available in official name sources.
Example(s)	*Harpagophytum zeyheri*, Root, Dry Extract Ethanol-Water (60-40)(1.5-3.0 = 1 w/w)
Conformance	MANDATORY
Data Type	ST

6.2.2.1 Substance name (repeat as necessary)

Names associated with a specified substance Group 1 will be captured in manner similar to Substance Name (4.5). The name type should be specified and, for Official Names, the same format should be followed (4.5.4). Names used in pharmacopoeias, legislation or regulations should always be used when available.

6.2.2.2 Version (repeat as necessary)

Version associated with a specified substance Group 1 will be captured in manner similar to Substance Version (4.13).

6.2.3 Reference sources

Reference information will also be captured in a manner similar to that describe in a manner similar to the substance reference information (4.6).

6.2.4 Fraction (new class to be included in the second edition of ISO 11238)

Note that the fraction Oil, will be captured at the substance information level. The fraction Class at the specified substance Group 1 information level is meant to capture additional information regarding the Extraction Types Dry, Liquid extract captured at the Substance level.

Information at the specified substance Group 1 level is used to capture the extraction solvent composition and the (herbal) Drug to Native herbal Preparation (Extract) ratio, and to describe the constituents present in the extract.

Table 202 — Fraction

User Guidance	The fraction of the plant where the herbal substance is derived.
Example(s)	Dry Extract DER (Drug extract Ratio 20 g plant material = 1 g Dry extract) (20 = 1)
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	When the herbal substance is derived from a fraction of the plant, the fraction is required.

6.2.5 Modification (new classes to be included in the second edition of ISO 11238)

Modifications, particularly those involving herbal preparations, should be captured at the specified substance Group 1 level when they are not captured at the substance level (4.11).

6.2.6 Constituent (repeat as necessary)

6.2.6.1 Substance ID

Table 203 — Substance ID

User Guidance	Every specified substance Group 1 should have at least one constituent and the Substance ID each constituent should be captured.
Example(s)	ISO 11238 Substance or specified substance ID
Conformance	MANDATORY
Data Type	II
Values Allowed	ISO 11238 Substance or specified substance ID
Business Rule(s)	Mandatory, at least one constituent is necessary for every specified substance Group 1. A constituent can be a substance or a specified substance.

6.2.6.2 Substance name

Table 204 — Substance name

User Guidance	The name of the substance which is the constituent of the specified substance Group 1 shall be described, Preferred or Official Name of the constituent.
Example(s)	**Devil's claw** root characteristic constituents are: Iridoid glucosides, harpagide, 8-pcoumaroylharpagide, procumbide and its 6'-p-coumaroyl ester.
Conformance	MANDATORY
Data Type	CD
Values Allowed	ISO 11238 Substance or specified substance name
Business Rule(s)	This value is implicit and derived from the Substance ID. The Preferred name is the default value.

6.2.6.3 Substance role

Table 205 — Substance role

User Guidance	The role of the substance in the specified substance shall be described based on a controlled vocabulary. An impurity component is not described in the specified substance Group 1. The parent ID of a specified substance Group 1 shall always be specified.
Example(s)	active marker, parent substance
Conformance	MANDATORY
Data Type	CD
Values Allowed	COMPONENT, EXTRACTION SOLVENT, SIGNATURE SUBSTANCE; ACTIVE MARKER
Business Rule(s)	Markers for herbal specified substance are conditional, when available information on markers shall be provided.

6.2.6.4 Amount type

As in 4.7.12.3.1.

6.2.6.4.1 Amount

The information related to the amount, where applicable, shall be provided as per specification described in 4.9.

6.2.7 Physical form (repeat as necessary)

6.2.7.1 Physical state

Table 206 — Physical state

User Guidance	The actual state of the substance in packaged product or administered product
Example(s)	solid, liquid
Conformance	MANDATORY
Data Type	CD
Values Allowed	SOLID, LIQUID, GAS, EMULSION, GEL
Business Rule(s)	This information is implicit and derived from the Physical Form Type

6.2.7.2 Physical form type

Table 207 — Physical form type

User Guidance	Type of the physical form
Example(s)	Dry Extract
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	When the physical state is available, it shall be provided as it defines a new Substance ID

6.2.7.3 Amount type

Table 208 — Amount type

User Guidance	Most elements that have a quantitative value will also have a field called amount type. Amount type shall always be specified because the actual value of the amount is often dependent on it. Example: In capturing the actual relative amounts of substances or molecular fragments it is essential to indicate whether the amount refers to a mole ratio or weight ratio. For any given element an effort should be made to use the same amount type for all related definitional elements.
Example(s)	Mole percent; Weight percent; Mole ratio; Weight ratio, Weight, Moles
Conformance	CONDITIONAL
Data Type	CD

6.2.7.3.1 Amount

The information related to the amount, when applicable, shall be provided as per specification described in 4.9.

6.3 Specified substance Group 2 (repeat as necessary)

Specified substances Group 2 will be used to capture the manufacturer of a given substance as well as limited manufacturing information. The defining elements are the parent substance which will be either a substance or a specified substance Group 1. Many biosimilar substances will be distinguished at this level. Any manufacturing information specific to the finished product will be described based on the ISO 11615:2012 model and according to the specifications outlined in ISO TS 20443 and 20451.

Each manufacturer should be identified with a code, such as D.U.N.S. number (as ID), and a name. The manufacturing type can be manufacturer. The description of the repackager or distributor will be specified at the medicinal product level and be based on the ISO 11615:2012 model, and the specifications outlined in the ISO TS 20443 and 20451. It is also important to capture the production method type (i.e. synthetic, extractive, biosynthetic) production type and the production system that describes the cell line or animal from which a given substance is isolated from.

The manufacturer is not necessarily the entity that has market authorisation, but the entity that manufactures or repackages the material. **If a substantial change in a critical manufacturing process occurs, a new specified substance Group 2 ID should be generated.** This would result in a new version of the specified substance Group 2. The grade of a material as described for specified substance Group 3 can also be provided as a non-definitional field when relevant.

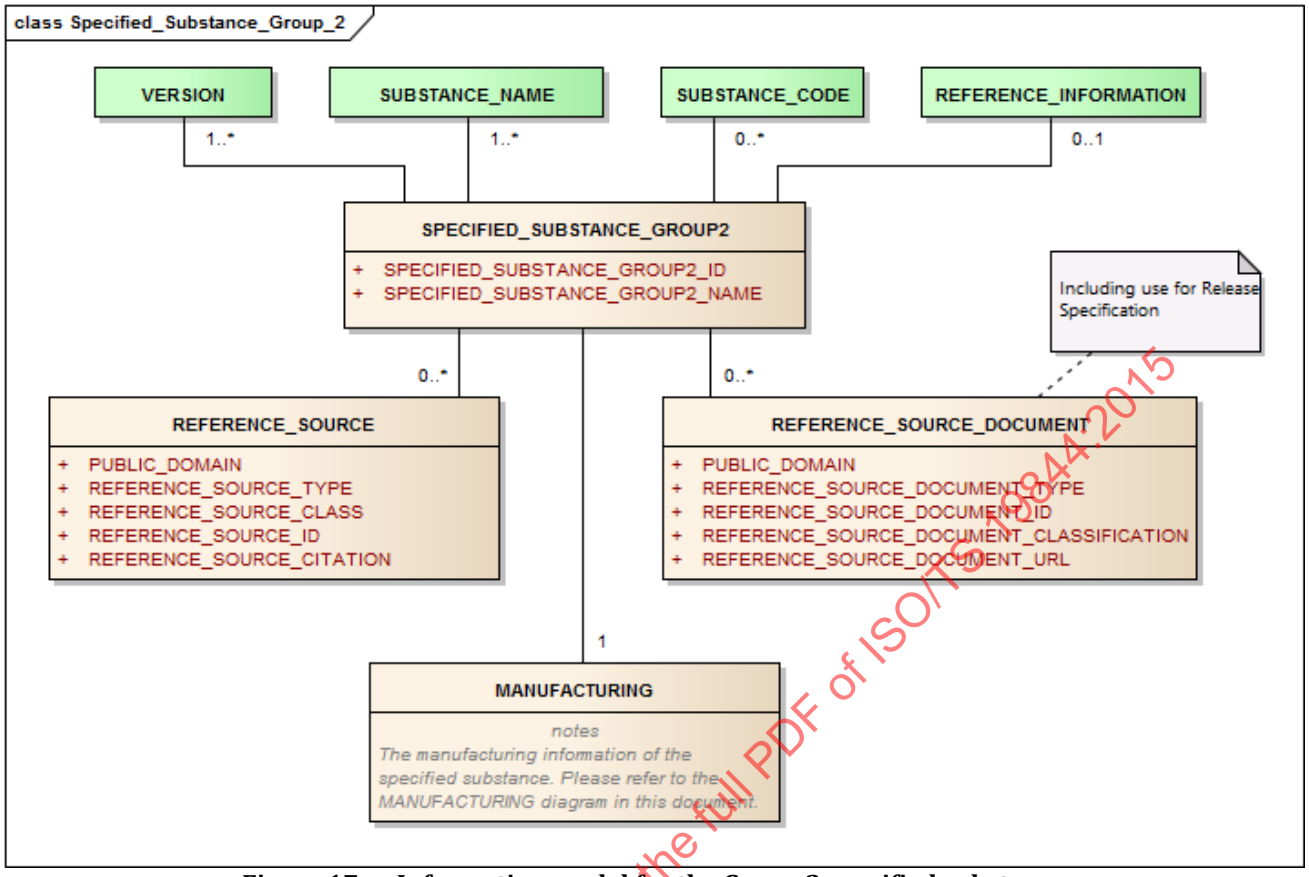


Figure 17 — Information model for the Group 2 specified substance

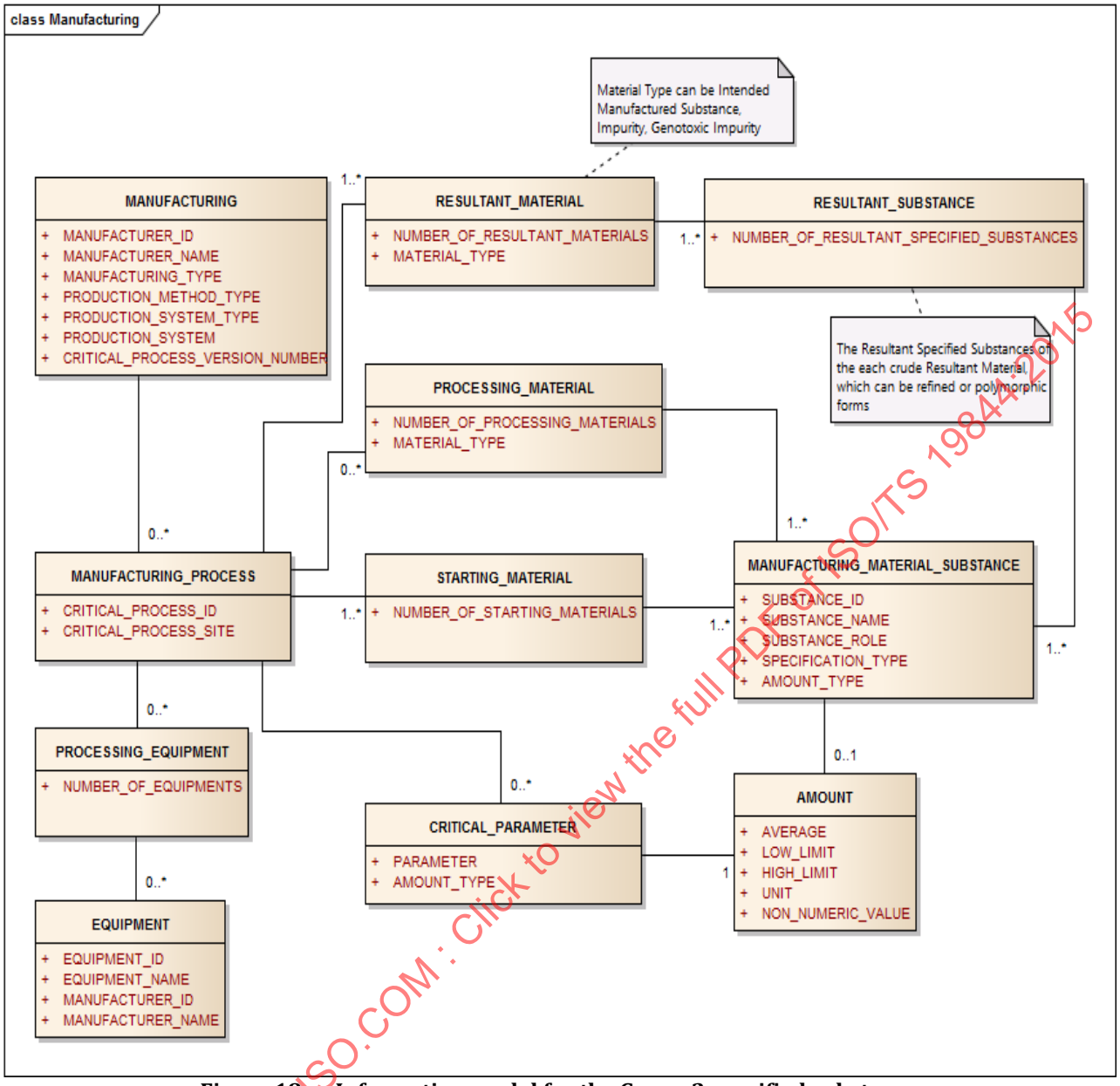


Figure 18 — Information model for the Group 2 specified substance

In this Substance Class specified substance Group 2 the Element group Reference Source Document is included. The Reference Source Document Type provides the manufacturer to submit the Release specification of the manufactured preparation, according to Figure 17.

Only the Element Group Manufacturing in Figure 18 is meant to be implemented yet up to the minimum information captured for the attributes in this element group. However in future versions of the Implementation Guide this element group can be extended up to a Manufacturing Class with the information model as is shown in Figure 18. The element group Manufacturing information is extended as a class and is yet allocated at the specified substance Group 4 information level in the ISO 11238 Substance Standard. This (extended) manufacturing information is shown only for informational purposes and is meant to capture the starting materials, impurities and the intended manufactured substance. The model follows up to now not any further information than the element group Manufacturing.

6.3.1 Specified substance Group2 ID

Table 209 — Amount type

User Guidance	The unique identifier assigned to the specified substance Group 2 shall be specified. NOTE: If a unique "specified substance ID" has been assigned, this "specified substance ID" shall be specified based on the Substance Name controlled vocabulary. In the absence of a unique "specified substance ID" e. g. for the initial submission of the substance this data element is not required.
Conformance	MANDATORY
Data Type	II
Business Rule(s)	NOTE: the specified substance ID is NOT SPECIFIED for specified substance ID Request

6.3.2 Specified substance Group2 Name

Table 210 — specified substance Group 2 name

User Guidance	The name of the specified substance Group 2 shall be provided in this field. The name should be coined from the preferred term of the substance and the manufacturer.
Example(s)	Human Insulin Drugco
Conformance	MANDATORY
Data Type	ST
Business Rule(s)	The name should be coined from a concatenation of the preferred term of the parent substance and the manufacturer (i.e. Human Insulin Drugco). Regions may provide additional guidance as the latest version of the Naming Conventions document (Ref. Doc.: EMA/720247/2011).

6.3.3 Parent Substance ID

Table 211 — Parent Substance ID

User Guidance	The Substance or specified substance Group 1 ID that identifies the manufactured substance. The unique identifier assigned to the substance that is the parent of the specified substance shall be provided based on controlled vocabulary. The ID of the parent for the specified substance Group 2 may refer to either a Substance ID or a specified substance Group 1 ID. NOTE: If a unique "ID" has been assigned, this "ID" shall be specified based on the Substance Name controlled vocabulary. In the absence of a unique "ID" e. g. for the initial submission of specified substance and parent substance, the name of the parent substance shall be specified.
Conformance	MANDATORY
Data Type	II
Values Allowed	ISO 11238 substance or specified Substance ID

6.3.4 Manufacturing

The manufacturing information shall be provided according to the following specifications:

6.3.4.1 Manufacturer ID

Table 212 — Manufacturer ID

User Guidance	The unique code used to track manufacturers.
Conformance	CONDITIONAL
Data Type	II
Business Rule(s)	An ID associated with each manufacturer may or may not be linked to a manufacturer site; when available it shall be provided. A necessary condition for the specified substance Group 2 to exist is: either the Manufacturer ID, Name and type is available AND/OR the production method type, product system type and production system exist. The manufacturer and production description can coexist, at least one of those shall be provided.

6.3.4.2 Manufacturer name

Table 213 — Manufacturer name

User Guidance	The name of the manufacturer of the specified substance.
Conformance	CONDITIONAL
Data Type	ST
Business Rule(s)	The general name of the manufacturer not linked to a specific site. This is implicit and derived from the Manufacturer ID. A necessary condition for the specified substance Group 2 to exist is: either the Manufacturer ID, Name and type is available AND/OR the production method type, product system type and production system exist. The manufacturer and production description can coexist, at least one of those shall be provided.

6.3.4.3 Manufacturing type

Table 214 — Manufacturing type

User Guidance	The type of operation performed by the entity associated with the substance.
Example(s)	Manufacturer (of Substance or specified substance)
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	If a repackager the original manufacturer should also be captured. A necessary condition for the specified substance Group 2 to exist is: either the Manufacturer ID, Name and type is available AND/OR the production method type, product system type and production system exist. The manufacturer and production description can coexist, at least one of those shall be provided.

6.3.4.4 Production method type

Table 215 — Production method type

User Guidance	The overall type of production system, synthetic, extractive, biosynthetic.
Example(s)	synthetic, extractive, biosynthetic, semi-synthetic
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	When applicable and available it shall be provided. A necessary condition for the specified substance Group 2 to exist is: either the Manufacturer ID, Name and type is available AND/OR the production method type, product system type and production system exist. The manufacturer and production description can coexist, at least one of those shall be provided.

6.3.4.5 Production system type

Table 216 — Production system type

User Guidance	Should be captured for material derived from both extractive and biosynthetic production method, for synthesised peptides and nucleic acids.
Example(s)	Plant, animal, bacterial, fungal, insect cell line, yeast, mammalian cell line, human cell line, animal tissue, human tissue, solid phase chemistry, solution chemistry
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	Should be captured for all material derived from extractive or biosynthetic production methods and some synthetic peptides and nucleic acids. When applicable and available it shall be provided. A necessary condition for the specified substance Group 2 to exist is: either the Manufacturer ID, Name and type is available AND/OR the production method type, product system type and production system exist. The manufacturer and production description can coexist, at least one of those shall be provided.

6.3.4.6 Production system

Table 217 — Production system

User Guidance	The production system description shall be provided when available.
Example(s)	CHO cell, goat, bovine lungs
Conformance	CONDITIONAL
Data Type	CD
Values Allowed	Should be captured for all material derived from extractive or biosynthetic production methods and some synthetic peptides and nucleic acids. When applicable and available it shall be provided. A necessary condition for the specified substance Group 2 to exist is: either the Manufacturer ID, Name and type is available AND/OR the production method type, product system type and production system exist. The manufacturer and production description can coexist, at least one of those shall be provided.
Business Rule(s)	Production System, System Type and Method are all to be represented in one terminology. When applicable and available it shall be provided.

	A necessary condition for the specified substance Group 2 to exist is: either the Manufacturer ID, Name and type is available AND/OR the production method type, product system type and production system exist. The manufacturer and production description can coexist, at least one of those shall be provided.

6.3.4.7 Critical process version number

Table 218 — Critical process version number

User Guidance	Should be captured for material derived from both extractive and biosynthetic production method, for synthesised peptides and nucleic acids.
Example(s)	2
Conformance	OPTIONAL
Data Type	INT
Values Allowed	Start at one and increases if a major critical process changes occur (i.e. changes in master cell bank; elimination or addition of a chromatographic purification process).
Business Rule(s)	The critical process version number shall be tied to the production system.

NOTE The version should be tied to the Production System Type and will change if a critical process changes.

6.3.4.8 Substance name (repeat as necessary)

Names associated with a specified substance Group 2 will be captured in manner similar to Substance Name (4.5). Trade names associated with the manufacturer can be captured at this level.

6.3.4.9 Substance Code

Codes associated with a specified substance Group 2 will be captured in manner similar to Substance Codes (4.7)

6.3.4.10 Reference sources

Reference Information will also be captured in a manner similar to that describe in a manner similar to the substance reference information (4.6).

6.3.4.11 Version (repeat as necessary)

Version associated with a specified substance Group 2 will be captured in manner similar to Substance Version (4.13).

6.4 Specified substance Group 2 for herbal preparations

6.4.1 Specified substance Group2 ID

Table 219 — Specified substance Group 2 ID

User Guidance	The unique identifier assigned to the specified substance Group 2 shall be specified. NOTE: If a unique "specified substance ID" has been assigned, this "specified substance ID" shall be specified based on the Substance Name controlled vocabulary. In the absence of a unique "specified substance ID" e. g. for the initial submission of the substance this data element is not required.
Example(s)	AGHFT4352E
Conformance	MANDATORY
Data Type	II
Business Rule(s)	NOTE: the specified substance ID is NOT SPECIFIED for specified substance ID Request

6.4.2 Specified substance Group2 Name

Table 220 — Specified substance Group 2 ID

User Guidance	The name of the specified substance Group 2 shall be provided in this field. The name should be coined from the preferred term of the substance and the manufacturer. The Name Extract and Parent relationship with the herbal preparation. Coupling to the manufacturer.
Example(s)	*Harpagophytum procumbens*, Root, Dry Extract Ethanol-Water (60-40) (1.5–3.0 = 1 w/w) — Manufacturer Ex.
Conformance	MANDATORY
Data Type	ST
Business Rule(s)	The name should be coined from a concatenation of the preferred term of the parent substance and the manufacturer (i.e. Human Insulin Lilly).

6.4.3 Parent Substance ID

Table 221 — Parent Substance ID

User Guidance	The Substance or specified substance Group 1 ID that identifies the manufactured substance. The unique identifier assigned to the substance that is the parent of the specified substance shall be provided based on controlled vocabulary. The ID of the parent for the specified substance Groups 1, 2 or 3may refer to either a Substance ID or a specified substance Group 1 ID. NOTE: If a unique "ID" has been assigned, this "ID" shall be specified based on the Substance Name controlled vocabulary. In the absence of a unique "ID" e. g. for the initial submission of specified substance and parent substance, the name of the parent substance shall be specified.
Example(s)	EXCTER654E (= specified substance Group 1 ID of the herbal preparation)
Conformance	MANDATORY
Data Type	II
Values Allowed	ISO 11238 Substance or specified substance ID

6.4.4 Manufacturing

6.4.4.1 Manufacturer ID

Table 222 — Manufacturer ID

User Guidance	The unique code used to track manufacturers.
Conformance	CONDITIONAL
Data Type	II
Business Rule(s)	An ID associated with each manufacturer may or may not be linked to a manufacturer site; when available it shall be provided. A necessary condition for the specified substance Group 2 to exist is: either the Manufacturer ID, Name and type is available AND/OR the production method type, product system type and production system exist. The manufacturer and production description can coexist, at least one of those shall be provided.

6.4.4.2 Manufacturer name

Table 223 — Manufacturer name

User Guidance	The name of the manufacturer of the specified substance.
Conformance	CONDITIONAL
Data Type	ST
Business Rule(s)	The general name of the manufacturer not linked to a specific site. This is implicit and derived from the Manufacturer ID. A necessary condition for the specified substance Group 2 to exist is: either the Manufacturer ID, Name and type is available AND/OR the production method type, product system type and production system exist. The manufacturer and production description can coexist, at least one of those shall be provided.

6.4.4.3 Manufacturing type

Table 224 — Manufacturing type

User Guidance	The type of operation performed by the entity associated with the substance.
Example(s)	Manufacturer (of the Substance or specified substance).
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	A necessary condition for the specified substance Group 2 to exist is: either the Manufacturer ID, Name and type is available AND/OR the production method type, product system type and production system exist. The manufacturer and production description can coexist, at least one of those shall be provided.

6.4.4.4 Production method type

Table 225 — Production method type

User Guidance	The overall type of production system, synthetic, extractive, biosynthetic.
Example(s)	Extraction
Conformance	CONDITIONAL
Data Type	CD
Business Rule(s)	The general name of the manufacturer not linked to a specific site. This is implicit and derived from the Manufacturer ID. A necessary condition for the specified substance Group 2 to exist is: either the Manufacturer ID, Name and type is available AND/OR the production method type, product system type and production system exist. The manufacturer and production description can coexist, at least one of those shall be provided.

6.4.5 Version (repeat as necessary)

Version associated with a specified substance Group 2 will be captured in manner similar to Substance Version (4.13)

6.5 Specified substance Group 3 (repeat as necessary)

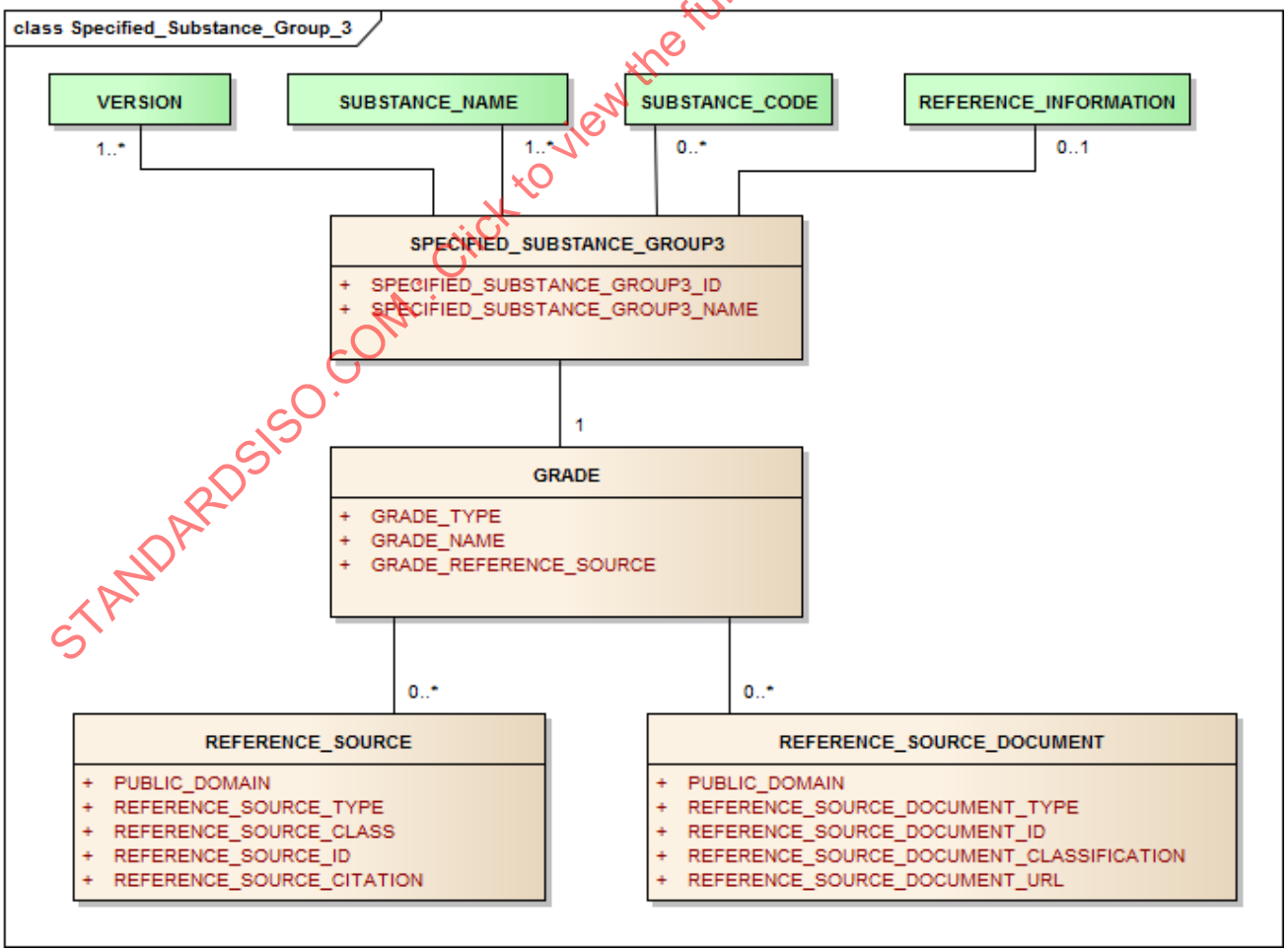


Figure 19 — Information model for the Group 3 specified substance

Figure 19 provides the information to capture the grade of a substance. This group is meant to be for Pharmacopoeial Grades (EP, USP). Each grade will lead for a separate specified substance Group 3 Identifier. However there are applications in which regulatory approval has been achieved for more than one grade and sometimes the sponsor will add an extra specification, such as particle size which is not captured in the Pharmacopoeial monograph. In this case the model provides in the Reference documentation type "In House" Grade in which a specification set can be laid down covering the EP, USP as well as the intended particle size.

The information on the grade shall be specified by means of the following data elements:

6.5.1 Specified substance Group 3 ID

Table 226 — Specified substance Group 3 ID

User Guidance	The unique identifier assigned to the specified substance Group 3 shall be specified. NOTE: If a unique "specified substance ID" has been assigned, this "specified substance ID" shall be specified based on the Substance Name controlled vocabulary. In the absence of a unique "specified substance ID" e. g. for the initial submission of the substance this data element is not required.
Conformance	MANDATORY
Data Type	II
Business Rule(s)	NOTE: the specified substance ID is NOT SPECIFIED for specified substance ID Request

6.5.2 Specified substance Group3 Name

Table 227 — Specified substance Group 3 name

User Guidance	The name of the specified substance Group 3 shall be provided in this field or it should be derived from the specified substance Group 1 Name (Preferred Term)
Example(s)	Name in monograph appended to Grade Type (i.e. sterile water for injection USP)
Conformance	MANDATORY
Data Type	ST

6.5.3 Parent Substance ID

Table 228 — Parent Substance ID

User Guidance	The unique identifier assigned to the substance that is the parent of the specified substance shall be provided based on controlled vocabulary. The ID of the parent for the specified substance Group 3 may refer to either a Substance ID or specified substance Group 1 ID. NOTE: If a unique "ID" has been assigned, this "ID" shall be specified based on the Substance Name controlled vocabulary. In the absence of a unique "ID" e. g. for the initial submission of specified substance and parent substance, the name of the parent substance shall be specified.
Conformance	MANDATORY
Data Type	II
Values Allowed	ISO 11238 Substance or specified substance ID

6.5.4 Grade

The characteristics of the grade of the substance shall be specified.

6.5.4.1 Grade type

Table 229 — Grade type

User Guidance	Pharmacopoeial specification type or other specification type.
Conformance	MANDATORY
Data Type	CD
Business Rule(s)	Each Pharmacopoeial Specification shall be given a separate record as described in the applicable Annexes.

6.5.4.2 Grade name

Table 230 — Grade name

User Guidance	Typically the Monograph Title that refers to a given substance or specified substance; for herbal substances standardised or non-standardised will be appended to the name along with standardisation.
Example(s)	quantified, standardised
Conformance	MANDATORY
Data Type	ST

6.5.4.3 (Grade) reference source

Table 231 — (Grade) reference source

User Guidance	Refers to the relevant reference source of the grade or specific source of monograph.
Example(s)	EP., Ed 7th 2011 (7.2), USP
Conformance	MANDATORY
Data Type	ST

6.5.5 Reference source (repeat as necessary)

Reference source will be captured in a manner similar to the one described in 4.6.

6.5.6 Version (repeat as necessary)

Version associated with a specified substance Group 1 will be captured in manner similar to Substance Version (4.13).

6.5.6.1 (Grade) reference source

Table 232 — (Grade) reference source

User Guidance	Refers to the relevant reference source of the grade or specific source of monograph.
Example(s)	EP., Ed 7th 2011 (7.2), USP
Conformance	MANDATORY
Data Type	ST

6.5.7 Reference source (repeat as necessary)

Reference source will be captured in a manner similar to the one described in 4.6.

6.5.8 Version (repeat as necessary)

Version associated with a Group 1 specified substance will be captured in manner similar to Substance Version (4.13).

Annex A
(normative)

Choosing a Substance ID

A.1 General

Public access to substance identifiers should be provided, Identifiers should be mapped to a variety of names and also provide defining information to assist in proper choice of a substance identifier. Before requesting a Substance ID, the Substance ID list on the public website should be checked to determine if a Substance ID has already been generated for the relevant substance. Most of the substances in medicinal products will already have a substance identifier associated with the material.

When selecting a Substance ID, always choose the Substance ID that is the most specific to the material being used. For example, for bacterial derived products, if a single strain of bacteria is used in a given product, the identifier should map to the particular strain. For other products the strain may not be known or important for the function of the product and in that instance an identifier for the bacterial species may be sufficient to describe the material. In general for structurally diverse material derived from biological matrices an attempt will always be made to map the material at least to the species level.

For chemical substances, material that differs either in salt or solvate stoichiometry will be assigned separate identifiers. It is important to choose the identifier that maps to the actual material used in the product. It should be noted that the same USAN and INN name will frequently map to different substances. INN names typically do not refer to solvated substances while USAN names do. For example the USAN term Amoxicillin is defined as a trihydrate while the INN term Amoxicillin refers to the anhydrous substance.

Care should always be taken to ensure that the correct substance identifier is used in describing the formulation of a product. In the Medicinal Product standard (ISO 11615), substance identifiers are also needed for both the active moiety and the basis of strength. Links between substance and active moieties for all active substances that can be used in a product should be provided. Pharmacopoeias or the maintenance organisation may also need to be consulted to determine the substance identifier to be used for the basis of strength for a given product. The concept of what constitutes an active moiety could vary between jurisdictions or even by type of product. The substance or concept related to basis of strength will often be indicated in regional pharmacopoeias and could vary between regions.

A.2 Requesting a Substance ID and providing information

If the appropriate substance ID is not available on this list, a Substance ID request should be sent to the maintenance organisation.

In submitting information for the assignment of a Substance ID, any information considered confidential should be so indicated. Any confidential information linked to the substance will only be released with the permission of the submitter or if it is unambiguously found in the public domain associated with the substance.

If a chemical structure or protein sequence is found in INN, USAN, or JAN, that information along with the Substance ID should be released to the public even if it was marked confidential when the Substance ID was generated. Substance identifiers will not be released to the public unless the defining information is in the public domain and unambiguously associated with the preferred term of a given substance. For substances identified with company codes or trade name, the association of the company code or trade name with Substance ID will not be released unless the defining information is found within a single reputable source associated with the company code or trade name. A company can always request to have it trade name or company code associated with a given substance. The company should ensure that any substance associated with a trade name or company is completely defined and the trade name unambiguously refers to that single substance or specified substance.

At least one name or company code should be associated with each request for a Substance ID. To facilitate mapping and to limit ambiguity, the submitter should include all codes and common names that are known to have been associated with a given substance. Once an ID is assigned it will be permanently associated with that substance. A submitter should explicitly indicate which information, whether defining or reference, is considered to be in the public domain and which is confidential.

The accompanying annex documents will provide information on the specific information that is necessary to assign a Substance ID that is compliant with the ISO standard.

Annex B
(normative)

Chemical substance

B.1 General

In Figure B.1 the high level information is provided regarding chemical substances.

A substance is described as any matter that has a discrete existence, irrespective of origin, which may be biological or chemical. Characteristics are described by two logical levels: Substance level and Specified substance level (blue blocks). The chemical substance class is one of the five single substance types (see 4.3 Substance types).

For both levels at least one substance name or specified substance name is a mandatory object. When in special cases no name has been provided, a (company)-code should be provided with a short description of the substance.

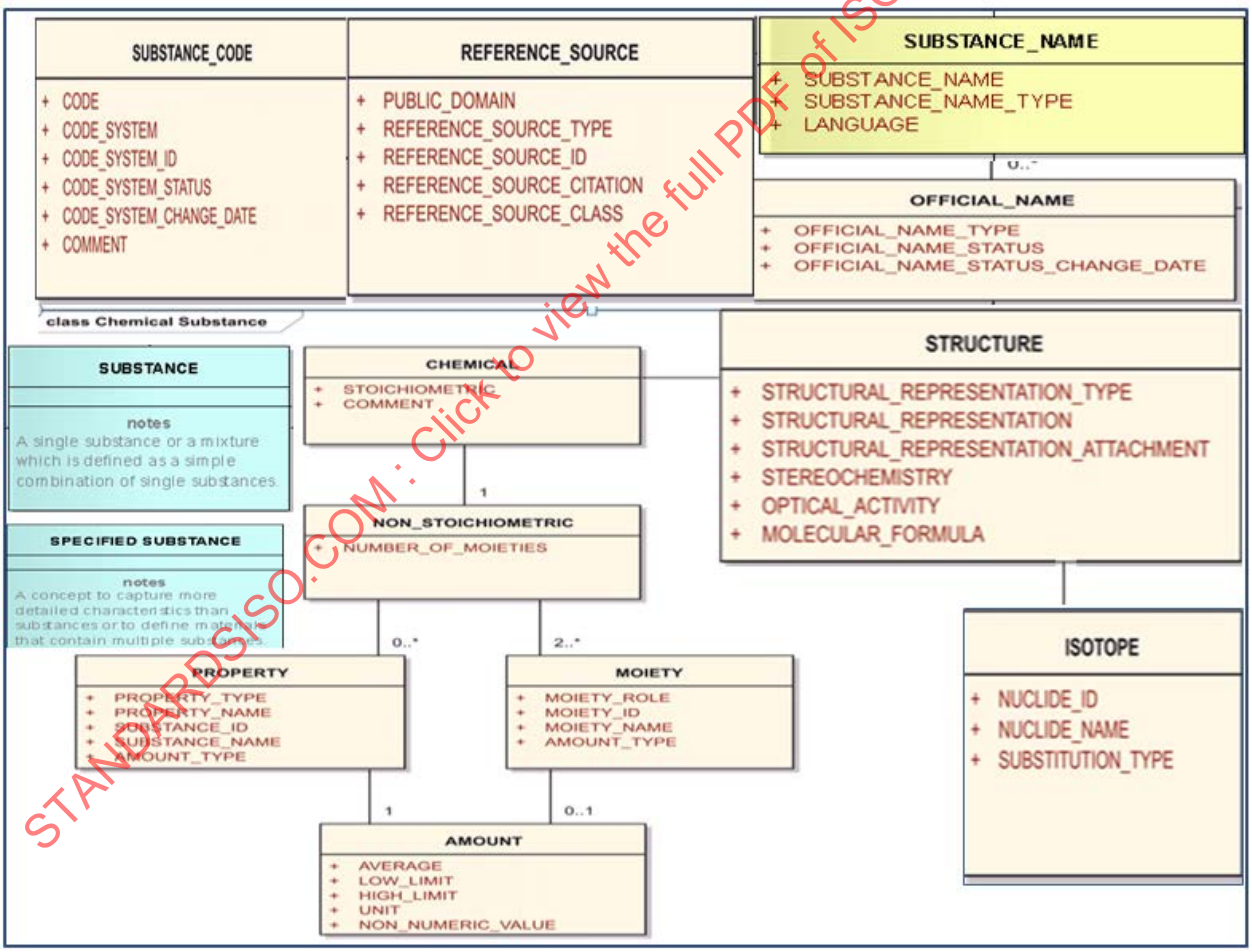


Figure B.1 — High level overview of substance information: Class Chemical Substance

NOTE 1 All element names are defined within the Glossary of this Annex.

NOTE 2 In the box Structure of Figure B.1 the element molecular weight is not mentioned.

Chemical substances, whose primary defining element is the molecular structure, shall be defined on the basis of their complete covalent molecular structure; the presence of a salt (counter-ion) and/ or solvates is also captured.

The molecular structure, the molecular formula, the **molecular weight** and optical activity, together with the representation of the stereochemistry are mandatory elements to be provided.

Although the molecular weight can be derived from the structure, this element should be presented at the substance level for chemical substances in order to substantiate the provided structural information. **The ISO 11238 Substance standard should be updated for the element molecular weight at the next version since this element is missing from the class Chemical substances.**

In addition the molecular formula should be described according to the moieties even for stoichiometric substances.

Example: Amlodipine besilate, Element Molecular formula**:**

Molecular formula: C20H25ClN2O5·C6H6O3S

NOTE 1 The molecular formula, 4.8.6 and 4.8.7, is specified in accordance with the Hill system. For a salt the base or the acid and the salt moiety are separated by a dot. In this case C20H25ClN2O5·C6H6O3S.

NOTE 2 The representation of the molecular formula will be presented per moiety next to (in addition to) the representation of the total elements: C26H31ClN2O8S.

NOTE 3 For stoichiometric chemicals the molecular formula should be directly derived from the molecular structure of the chemical.

Element Group: Molecular weight
Molecular Weight Method: Calculated
Molecular weight Type: Number Average

NOTE 4 The molecular weight type for small molecules is number average.There are basically four types of molecular weight available: number average M_n, viscosity average, M_v, weight average M_w, and Z-average, M_z. See for further information 5.2.6.2 of the ISO/TS 19844 Substance Implementation Guide.

Element group: Amount
Average [Numeric Value]: 567,06 = (408.88 + 158.18) or rounded at 567,1 in accordance with the EP monograph
Unit: g/mol.

NOTE 5 For stoichiometric chemicals the molecular weight should be directly derived from the molecular structure of the chemical.

The element group Moiety should be attached also to the element class Chemical, Stoichiometric substances in order to describe a salt and solvates relationship to the active moiety in an unambiguous way with respect to the description of the molecular formula and sometimes the official names.

B.1.1 Proposal for the update of the ISO 11238 Substance standard

The element molecular weight must be included in the next version of the standard and be included in the Structure Class Information Model together with the element group amount. The element group moiety and amount should also be tied to the element group stoichiometric substances.

This will help to describe salts and solvates and hydrated substances in a more unambiguous way with respect to the description of the molecular formula and sometimes official names. See Figure B.2.

Justification for the introduction of the molecular weight, moiety and amount element groups for stoichiometric chemical substances

According to the (EU)-legislation the qualitative and quantitative composition of a medicinal product should be described in terms of the active substances and constituents of the excipient.

This means that weight calculation from the active substance salt/ solvate/ salt hydrate form has to be calculated and expressed as the active substance base/acid. For this, the molecular formula is needed for the active substance salt/solvate/hydrate form and of the free base. In order to perform this calculation the molecular weight must be provided and not only calculated from the structure of the active substance. Therefore the name of the active substance, including the crystal form, molecular formula and molecular weight are mandatory fields next to the structure to comply to this (EU)-legislation.

NOTE Directive 2001/83/EC on the Community code relating to medicinal products for human use.

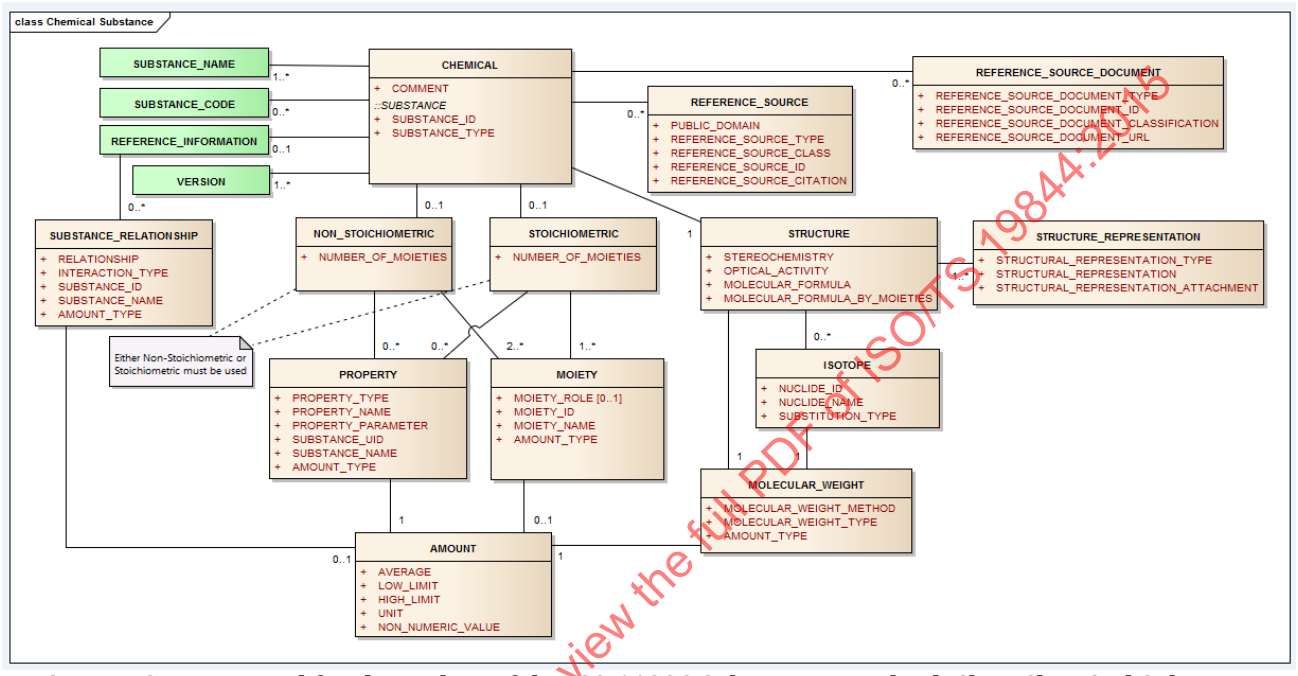


Figure B.2 — Proposal for the update of the ISO 11238 Substance standard, Class Chemical Substance

The moeity group also tied to the Stoichiometric Chemical Substance: *e.g.*:
Describing Amlodipine besylate hydrate (1:1:1) in equal mol ratios with the Molecular formula:
$C_{20}H_{25}ClN_2O_5 \cdot C_6H_6O_3S \cdot H_2O$ in stead of $C_{26}H_{33}ClN_2O_9S$; rINN: = Amlodipine besilate
Molecular weight: 585,06 = (408.88 + 158.18 + 18.00), rounded to 585,1 g/mol.
There will be two terms used to describe the Molecular formula:

- **Molecular formula by moieties:** $C_{20}H_{25}ClN_2O_5 \cdot C_6H_6O_3S \cdot H_2O$
- **Molecular formula:** $C_{26}H_{33}ClN_2O_9S$, which is equal to the sum of the molecular formula of the moieties.

Relations with other element groups

Element Group Structural Modification:
Structural Modification Type : Moiety

Element Group Moiety:
Moiety Role: Salt
Moiety_ ID: 685928Z18A (UNII)
Moiety_ Name: Benzenesulfonic acid
Amount Type: Mole Ratio to Amlodipine

Element Group Amount:
Average (Numeric value) : 1

Element Group Structural Modification:
Structural Modification Type: Moiety

Element Group Moiety:
Moiety Role: Solvate
Moiety_ID: 059QF0KO0R (UNII)
Moiety Name: Water
Amount Type: Mole Ratio to Amlodipine

Element Group Amount:
Average (Numeric value): 1

Element Group Substance Relationship:
- The Active Moiety relationship is a one way relationship;
- Parent molecule to a salt/ solvate is a two way relationship.
- Both one-and two-way relationships are possible.

Relationship: Active Moiety
Substance ID: PTYF75643O (Artificial ID); 1J444QC288 (UNII)
Substance Name : Amlodipine

NOTE The relationship from a parent molecule to a salt is an example of a two way relationship.

Relationship: Parent (Salt/Solvate)
Substance ID: PTYF75643O (Artificial ID); 1J444QC288 (UNII)
Substance Name: Amlodipine

Relationship: Salt
Substance ID: JUTYR9087S (Artificial ID); 685928Z18A (UNII)
Substance Name: Benzenesulphonic acid

NOTE For the further documentation in this Chemical Annex below it is assumed that the proposed changes of the ISO 11238 Substance Standard will be accepted for the addition of the element molecular weight and amount group and the change for the element moiety also connected to stoichiometric chemical substances when applicable.

When requesting a Substance ID for a substance in the development stage or substance under investigation, only a minimum set of information should be provided.

When the name is not available at least the Company code must be provided and a short description of the chemical substance with a representation of the structure, molecular formula and molecular weight. Subclause 4.5.1 provides further guidance.

The reference source should be provided with the indication whether the name is in the public domain or not (e.g. public databases such as CAS, name and number).

B.1.2 Outline of Annex B

This Annex B follows the main classification of the substance_type: Chemical Substance in view of ISO 11238 and ISO/TS 19844 by describing first:

B3: Chemical Substance subtypes and Mixture Substance

Captured at the Substance level:

- Stoichiometric chemical substances with respect to their solid state: Co-crystals, chemical salts, solvates and hydrates and (polymorphic form captured at the specified substance level);
- Non-Stoichiometric chemical substance
- Mixture substance

Captured at the specified substance level:

- Multi substance material

B4: Discussion of the key elements of a chemical substance:

- Substance name and Company-code
- Substance structure
- Substance molecular formula
- Substance molecular weight
- Decisions for assigning a new Substance ID

B5: Discussing other elements and properties of importance regarding the characteristics of the substance

B6: Proving examples of chemical substances and describing the elements captured at the Substance level and Group 1, 2 or 3 Specified substance level when appropriate.

B7: Discussing Radionuclide Substance

B.2 Scope

This Annex to ISO/TS 19844 Implementation Guide, class Chemical Substance, Mixture Substance and Radionuclide, provides guidance to define elements that describe mandatory and conditional requirements/ fields in order to generate ISO 11238 Substance Identifiers in view of EN ISO 11238 (hereafter ISO 11238) and ISO/TS 19844 Implementation Guide (hereafter ISO/TS 19844) by providing examples of substances belonging to the classes described above.

B.3 Terms and definitions

For the purposes of this Annex B the terms and definitions given in ISO 11238 and the following apply.

B2.1
ATC Code, Substance classification code
Anatomic Therapeutic Chemical Classification code

Note 1 to entry: The Anatomical Therapeutic Chemical (ATC) Classification System is used for the classification of drugs. It is controlled by the WHO Collaborating Centre for Drug Statistics Methodology (WHOCC).

Note 2 to entry: This pharmaceutical coding system divides drugs into different groups according to the organ or system on which they act and/or their therapeutic and chemical characteristics. Each bottom-level ATC code stands for a pharmaceutically used substance in a single indication (or use). This means that one drug can have more than one code: Acetylsalicylic acid (Aspirin), for example, has A01AD05 as a drug for local oral treatment, B01AC06 as a platelet inhibitor, and N02BA01 as an analgesic and antipyretic. On the other hand, several different brands share the same code if they have the same active substance and indications.

B2.2
boiling point
corrected temperature at which the vapour pressure of a liquid is equal to 101.3 kPa.

B2.3
co-crystals
multi-component crystals based on hydrogen bonding interactions without the transfer of hydrogen ions to form salts.

B2.4
CV
controlled vocabulary
finite set of values that represent the only allowed values for a data item.

Note 1 to entry: The allowed values can be codes, text, or numeric.

Note 2 to entry: Adapted from CDISC Clinical Research Glossary V8.0, 2009.

B2.5
CAS Index name
Chemical Abstracts Service Index Name, also referred to as CAS Registry name

B2.6
CAS Registry Number /CAS number
unique numerical identifier

Note to entry: The numbers are sequential and are assigned as a substance enters the registry system. The numbers do not have a common length and lengths vary from 5 digits to nine digits. Each CAS number contains a single check digit. The CAS registry system is maintained by the Chemical Abstracts Service of the American Chemical Society. The CAS number for formaldehyde is 50-00-0.

B2.7
Cahn-Ingold-Prelog priority rules, CIP system or CIP conventions
set of rules used in organic chemistry to name the stereoisomers of a molecule

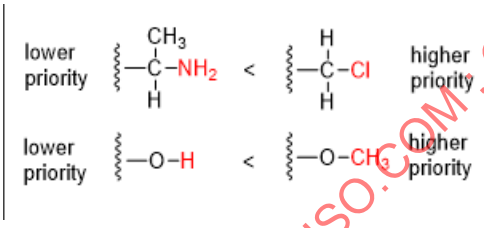
Note to entry: A molecule may contain any number of stereogenic carbon centres and any number of double bonds, and each gives rise to two possible configurations. The purpose of the CIP system is to assign an R or S descriptor to each stereogenic carbon centre and an E or Z descriptor to each double bond so that the configuration of the entire molecule can be specified uniquely by including the descriptors in its systematic name.

B2.8
configuration
method for indicating the three-dimensional arrangement of atoms at a stereogenic carbon centre

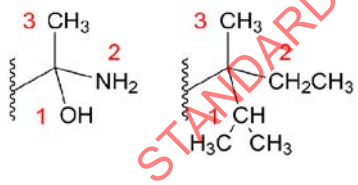
Note 1 to entry: Look at the four atoms directly attached to the stereogenic carbon centre and assign priorities in order of decreasing atomic number. The atom with the highest atomic number is ranked first; the atom with the lowest atomic number is ranked fourth

Note 2 to entry: If a decision about priority can't be reached by applying rule 1), compare atomic numbers of the second atoms in each substituent, continuing on as necessary through the third or fourth atoms until a point of difference is reached.

Note 3 to entry: Multiple-bonded atoms are considered equivalent to the same number of single-bonded atoms.

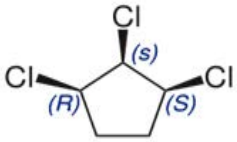


Stereogenic carbon centres: R/S



Two examples of stereogenic carbon centres. The lowest substituent (number 4) is shown only by a wavy line, and is assumed to be behind the rest of the molecule. Both centres shown are S isomers.

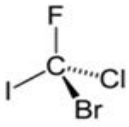
After the substituents of a stereogenic carbon centre have been assigned their priorities, the molecule is oriented in space so that the group with the lowest priority is pointed away from the observer. If the substituents are numbered from 1 (highest priority) to 4 (lowest priority), then the sense of rotation of a curve passing through 1, 2 and 3 distinguishes the stereoisomers. A centre with a **clockwise** sense of rotation is an ***R*** or *rectus* centre and a centre with a **counter clockwise** sense of rotation is an ***S*** or *sinister* centre. The names are derived from the Latin for right and left, respectively.



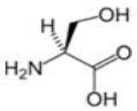
(1*R*,2*s*,3*S*)-1,2,3-trichlorocyclopentane

It is possible in rare cases that two substituents on an atom differ only in their absolute configuration (R or S). If the relative priorities of these substituents need to be established, R takes priority over S. When this happens, the descriptor of the stereogenic carbon centre is a lowercase letter (r or s) instead of the uppercase letter normally used.

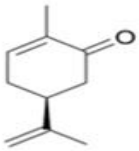
R/S assignments for several compounds



The hypothetical molecule bromochlorofluoroiodomethane shown in its R-configuration would be a very simple chiral compound. The priorities are assigned based on atomic number (Z): iodine (Z = 53) > bromine (Z = 35) > chlorine (Z = 17) > fluorine (Z = 9). Allowing fluorine (lowest priority) to point away from the viewer the rotation is clockwise hence the **R**-assignment.



In the assignment of L-serine highest priority is given to the nitrogen atom (Z = 7) in the amino group (NH_2). Both the methylalcohol group (CH_2OH) and the carboxylic acid group (COOH) have carbon atoms (Z = 6) but priority is given to the latter because the carbon atom in the COOH group is connected to a second oxygen (Z=8) whereas in the CH_2OH group carbon is connected to a hydrogen atom (Z=1). Lowest priority is given to the hydrogen atom and as this atom points away from the viewer the counterclockwise decrease in priority over the three remaining substituents completes the assignment as **S**.



The stereogenic carbon center in S-carvone is connected to one hydrogen atom (not shown, priority 4) and three carbon atoms. The isopropene group has priority 1 (carbon atoms only) and for the two remaining carbon atoms priority is decided with the carbon atoms two bonds removed from the stereogenic carbon center, one part of the keto group (O,O,C priority 2) and one part of an alkene (H,C,C priority 3). The resulting counterclockwise rotation results in a **S**.

EXAMPLES:

B2.9
critical point, critical temperature, critical pressure
behaviour of a heated liquid

Note to entry: The behaviour of a heated in an open vessel differs from that of a liquid in a sealed vessel.

In an open vessel, the liquid vaporises from its surface as it is heated. At the temperature at which its vapour pressure would be equal to the external pressure, vaporisation can occur throughout the bulk of the liquid and the vapour can expand freely into the surroundings. The condition of free vaporisation throughout the liquid is called **boiling.** The temperature at which the vapour pressure of a liquid is equal to the external pressure is called the **boiling temperature at that pressure.**

When a liquid is heated in a sealed vessel, boiling does not occur. Instead, the temperature, vapour pressure, and the density of the vapour rise continuously (Fig.2). At the same time, the density of the liquid decreases as a result of its expansion. (Fig. 2B) There comes a stage at which the density of the vapour is equal to that of the remaining liquid and the surface between the two phases disappears. (Fig. 2C) The temperature at which the surface disappears is the **critical temperature T_c** . The corresponding vapour pressure is the **critical pressure p_c.**

B2.10
distillation range
the temperature interval, corrected for a pressure of 101.3 kPa, within a liquid, or specified fraction of a liquid, distils under specified conditions as described in a Pharmacopoeia.

B2.11
elemental composition
pure chemical substance consisting of a single type of atom distinguished by its atomic number, which is the number of protons in its atomic nucleus. Elements are divided into metals, metalloids, and non-metals.

Note to entry: When two or more distinct elements are chemically combined, with the atoms held together by chemical bonds, the result is termed a chemical compound. Chemical compounds may be composed of elements combined in exact whole-number ratios of atoms to form a molecular formula. The **elemental composition** of the molecule Methane CH_4 is [C, H] since it consists of two chemical elements.

B2.12
EV-Code
EudraVigilance Code
code for a substance used in the Extended EudraVigilance Medicinal Product Dictionary (XEVMPD), developed by the European Medicines Agency (EMA). The XEVMPD is designed to support the collection, reporting, coding and evaluation of authorised and investigational medicinal product information in a standardised and structural way.

B2.13
hill system/hill notation
system of writing chemical formulas such that the number of carbon atoms in a molecule is indicated first, the number of hydrogen atoms next, and then the number of all other chemical elements subsequently, in alphabetical order. When the formula contains no carbon, all the elements, including hydrogen, are listed alphabetically.

B2.14
InChI
IUPAC International Chemical Identifier
textual identifier for chemical substances, designed to provide a standard and human-readable way to encode molecular information and to facilitate the search for such information in databases and on the web.

Note 1 to entry: InChI is a linear identifier that deals **with** chemical structure representation using a layered approach. The identifiers describe chemical substances in terms of layers of information There are currently six layers of information:

Note 2 to entry: Every InChI starts with the string "InChI=" followed by the version number. This is followed by the letter S for standard InChIs. The remaining information is structured as a sequence of layers and sub-layers, with each layer providing one specific **type** of information. The layers and sub-layers are separated by the delimiter "/" and start with a characteristic prefix letter (except for the chemical formula sub-layer of the main layer).

Note 3 to entry: The delimiter-prefix format has the advantage that a user can easily use a wildcard search to find identifiers that match only in certain layers. The InChI for benzene is:1/C6H6 /c1-2-4-6-5-3-1/h1-6H.

B2.15
InChI Key
fixed length (25 characters) condensed digital representation of the InChI designed to allow for easy web searches of chemical compounds.

Note 1 to entry: The full InChI is too lengthy for easy searching, and therefore the InChIKey was developed.

Note 2 to entry: This is sometimes referred to as hashed InChI.

B2.16
InChI resolvers
lookup service to make the links between the InChI key and InChI

Note to entry: The InChIKey always needs to be linked to the original InChI to get back to the original structure as the InChI cannot be reconstructed from the InChIKey.

B2.17
isomers

B2.17.1
stereoisomers
isomeric molecules that have the same molecular formula and sequence of bonded atoms (identical in atomic constitution) but differ in the three-dimensional arrangement of the atoms. Stereoisomers can either be enantiomers, diastereoisomers, geometric isomers or constitutional isomers

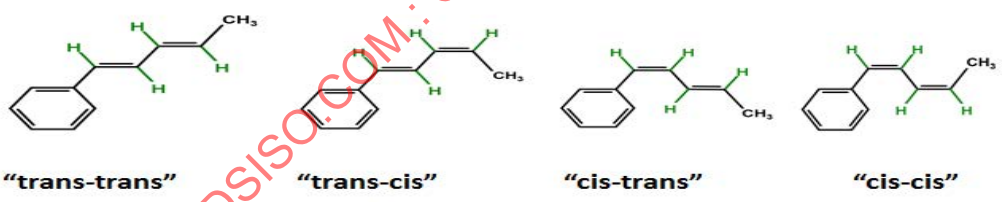
B2.17.2
enantiomers
two stereoisomers related to each other by a reflection. They are mirror images of each other which are non-superimposable. Every stereogenic centre in one molecule has the opposite configuration in the other. Two compounds that are enantiomers of each other have the same physical properties except for the direction in which they rotate polarised light and how they interact with different optical isomers of other compounds. As a result, different enantiomers of a compound may have substantially different biological effects. Pure enantiomers also exhibit the phenomenon of optical activity and can be separated only with the use of a chiral agent.

B2.17.3
diastereomers
stereoisomers not related through a reflection operation. Diastereomers are pairs of stereoisomers that are not mirror images of each other**.** Diastereomers have opposite configurations at one (or more) stereogenic centres, but have the same configuration at others. Diastereomers include meso compounds, *cis-trans* (E/Z) isomers, and non-enantiomeric optical isomers. Meso compounds are achiral, but they contain stereogenic centres of a (C)-atom. **Diastereomers will not have identical physical properties.**

B2.17.4
geometric Isomers
sub-class of diastereomers which exist due to restricted rotation in a molecule and do not require stereogenic centres to exist. Their relationship is commonly defined in terms of the relative stereochemical arrangement of groups rather than the absolute (R/S) configuration of stereogenic centres. The term *cis* and *trans* are also used to describe the relative position of two substituents on a ring: *Cis* on the same side, otherwise *trans.* Note: Z and E are not always interchangeable.

EXAMPLE 1: *Cis-Trans*; Z= Zusammen, together/ E=Entgegen, opposite.

EXAMPLE 2: Stereochemistry of 1-Phenyl-1,3-pentadiene



B2.17.5
conformers
form of isomerism that describes the phenomenon of molecules with the same structural formula having different shapes due to rotations around one or more bonds. Different conformations can have different energies, but usually interconvert, and are very rarely isolatable. For this reason most conformers will not be considered separate substances in this standard. Possible exceptions are substituted biphenyls where rotation is restricted and different conformers are isolatable.

B2.17.6
constitutional isomers
isomers that share the same molecular formula, but the bond connections (atomic constitution) and/ or their order differ(s) between different atoms/ groups. (e.g.*n*-Butane and *iso*-Butane; Ethyl alcohol and Dimethyl ether; Isopropylamine and *n*-Propylamine).

B2.18
IUPAC nomenclature
chemical nomenclature with a set of rules to generate systematic names for chemical compounds. The nomenclature is created and developed by the International Union of Pure and Applied Chemistry (IUPAC).

Note to entry: See http://www.chem.qmul.ac.uk/iupac/index.html#01; and http://www.chem.qmul.ac.uk/iupac/iupac.html

B2.19
melting point
melting point determined by the capillary method is the temperature at which the last solid particle of a compact column of a substance in a tube passes into the liquid phase.

B2.20
mixture substance
type of polydisperse substance that is a combination of single substances isolated or synthesised/obtained in the process together

EXAMPLE: Gentamicin would be defined as a mixture substance of Gentamicin C1, C1a, C2, C2a, and Gentamicin C2b. Glyceryl monoesters could be defined as a mixture substance of two single substances which differ in the position of esterification.

B2.21
moiety
moiety is an entity within a substance that has a complete and continuous molecular structure.

Note to entry The strength of a medicinal product is often based on what is referred to as the active moiety and should be defined in a consistent manner across all products. To avoid ambiguity the free acid and/or free base should be used as the moiety upon which strength is based.

B2.22
molecular formula
simplest type of formula, which uses only letters and numbers indicating atomic proportional ratios (the numerical proportions of atoms of one type to those of other types) of the chemical formula which is a way of expressing information about the proportions of atoms that constitute a particular chemical compound, using a single line of chemical element symbols, numbers, and sometimes also other symbols, such as parentheses, dashes, brackets, and plus (+) and minus (−) signs. These are limited to a single typographic line of symbols, which may include subscripts and superscripts. A chemical formula contains no words.(*e.g.* Glucose $C_6H_{12}O_6$).

B2.23
molecular structure
unambiguous representation of the arrangement of atoms. For the purposes of defining substances, three dimensional conformations shall not be captured.

B2.24
molecular weight
mass of one molecule of a homogenous substance or the average mass of molecules that comprise a heterogeneous substance. The unified atomic mass unit is the unit of molecular weight and includes the type of molecular weight (g/mol).

B2.25
multi-substance material
single substances of diverse origin that are brought together and do not undergo a chemical transformation will be defined as multi-substance material (Group 1 specified substance) and not a mixture substance.

B2.26
official name type
INN: International Nonproprietary Name, also known as rINN, recommended International Nonproprietary Name or pINN, proposed International Nonproprietary Name.

WHO: World Health Organization:
WHO-application International Nonproprietary Name rules:
(http://whoqlibdoc.who.int/hq/1997/WHO_PHARM_S_NOM_1570.pdf)
BAN: British Approved Name
EP: European Pharmacopoeia
JAN: Japanese approved name
JP: Japanese Pharmacopoeia
USAN: United States Adopted Names
USP: United States Pharmacopeia
DAB (EAB): Deutsches Arzneibuch. This has been followed up by Europäisches Arzneibuch.
HAB: Homöopatisches Arzneibuch 2010, Ambtliche Ausgabe. Deutschen Apotheken Verlag, Stuttgart.
HP: Pharmacopoea Homoepathica polyglotta. 8th edition. Leipzig. Dr. Willmar Schwabe, 1911.

B2.27
optical rotation
property displayed by chiral substances of rotating the plane of polarisation of polarised light. The amount of rotation observed in a polarimetry experiment depends on the number of optically actives molecules the light beam encounters. The more molecules the light encounters, the greater the observed rotation. Thus, the amount of rotation depends on both sample concentration and sample path length.

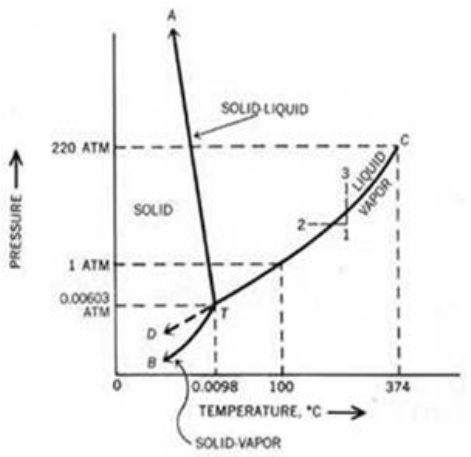
Note to entry Optical rotation is considered to be positive (+) for dextrorotatory substances and negative (-) for laevorotatory substances.

B2.27.1
specific optical rotation $[\alpha_m]^t_\lambda$
rotation, expressed in radians, measured at the temperature t (20°C) at a wavelength λ (usual D, light of 589,3 nanometer wavelength of the Sodium D-line, that is the yellow light emitted from common sodium lamps) and a path length of 1 decimeter (1 dm = 10 cm).

$$[\alpha]^{20}_D = \frac{\text{Observed rotation (degrees)}}{\text{Path length (dm) x Concentration C (g/ ml)}}.$$

B2.28
phase transition
occurs spontaneous of one phase to another phase at a characteristic temperature for a given pressure. Thus at 1 atm, ice is the stable phase of water below 0^0C., but above 0^0C the liquid is liquid water, $\mu(s)<\mu(l)$ (**Figure 1, 1A**). The Phase diagram of a substance shows the regions of pressure and temperature at which its various phases are thermodynamically stable. The boundaries between regions, the phase boundaries, show the values of p and T at which two phases coexist in equilibrium.

Note to entry: There is a set of conditions under which three different phases (typically solid, liquid and vapour) all simultaneously coexist in equilibrium. It is represented by the **triple point.**



(Fig.1 Phase diagram)

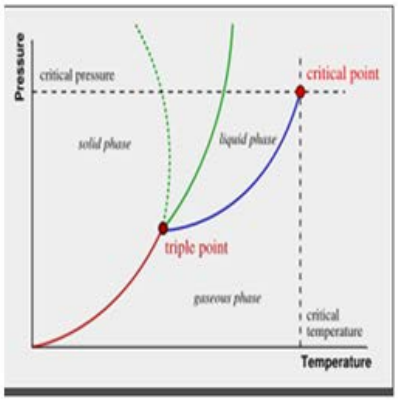


Fig. 1A Phase diagram, triple point critical point

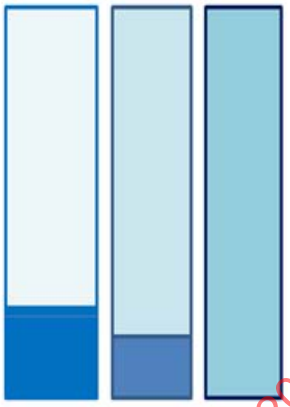


Fig. 2 Dark blue liquid.
Light blue vapour

Take the line TC which gives the vapour pressure of liquid water up to the critical point C. Along this line, liquid and vapour coexist in equilibrium. At temperatures higher than that of point C, condensation does not occur at any pressure. The line TB represents the vapour pressure of solid ice, which is a plot of the temperatures and pressures at which the solid and vapour are in equilibrium. Finally, line TA gives the melting point of ice and liquid water. The plot shows the temperatures and pressures at which ice and liquid water are in equilibrium.

B2.29
physical form
physical state, either gas, liquid, or solid, and the type of organisation for solid matter. Solids can be either crystalline or amorphous and can show polymorphism. Amorphous material is characterised by the absence of distinct reflections in the X-ray powder diffraction (XRPD) pattern.

B2.30
polymorphism
ability of a compound in the solid sate to exist in different crystalline forms having the same chemical composition

Note to entry: These different forms are formed by weak interactions between the components present in the solid state. These different forms may possess different physico-chemical properties. Different crystalline forms or solvates may be produced by varying the crystallization conditions (temperature, pressure, solvent, concentration rate of crystallisation, seeding of the crystallisation medium, presence and concentration of impurities etc.). Information about the techniques to study polymorphism, X-ray diffraction of powders or crystals, thermal analysis (differential scanning calorimetry), infrared absorption spectrophotometry and Near-infrared (NIR) spectrophotometry which is another technique used to measure the degree of crystallinity, and has been proven to be useful in studies of polymorphism and changes in the amorphous and crystalline state.

B2.31
refractive index
index, $n^{20}{}_D$, of a medium with reference to air equal to the ratio of the sine of the angle of incidence of a beam of light to the sine of the angle of refraction of the refracted beam in the given medium.

Note to entry: Unless otherwise prescribed, the refractive index is measured at 20°C ± 0,5°C, with reference to the wavelength of the D-line of sodium (λ = 589,3 nm).

B2.32
salt
ionic substances formed from the neutralisation reaction of an acid and base. Salts are ionic compounds composed of cations (positive ions) and anions (negative ions).

B2.33
solvate
substance formed through association of a solvent molecule (i.e. water, alcohol) with another moiety. Solvates can be either stoichiometric or non-stoichiometric and are predominately present in the solid form of substances.

B2.34
solubility
property of a solid, liquid, or gaseous chemical substance called solute to dissolve in a solid, liquid, or gaseous solvent to form a homogeneous solution of the solute in the solvent. The solubility of a substance fundamentally depends on the used solvent as well as on temperature and pressure. The extent of the solubility of a substance in a specific solvent is measured as the saturation concentration, where adding more solute does not increase the concentration of the solution.

B2.35
stoichiometric substances
substances that contains moieties in simple integral ratios.

Note 1 to entry: Defined stoichiometry shall be represented in the structural representation of a given substances. Moieties shall be represented using the lowest common factors such that a fractional representation is avoided. Substances will either be defined as stoichiometric or non-stoichiometric.

Note 2 to entry: Chemicals have defined stoichiometry when the ration of all moieties, (ion, counter ion and solvate) can be represented as simple integral ratios.

B2.36
UNII
10 character, randomly generated alpha-numeric string that is currently used to identify substances in medicinal products. The UNII is generated by the FDA/USP Substance Registry System, which is a robust system with detailed business rules for entry and the generation of UNIIs for both simple and complex substances. The first nine characters are randomly generated followed by a check character.

B2.37
viscosity
measure of a fluid of its resistance to gradual deformation by shear stress or tensile stress. For liquids, it corresponds to the informal notion of "thickness"

B2.37.1
dynamic viscosity or viscosity coefficient η
tangential force per unit surface, known as shearing stress T and expressed in pascals, necessary to move, parallel to the sliding plane, a layer of liquid of 1 m^2 at a rate (v) of 1 m/s relative to a parallel layer at a distance (x) of 1 meter.

Note to entry: The unit of dynamic viscosity is the pascal second (Pa.s). The most commonly used submultiple is the millipascal second (mPa.s).

B2.37.2
kinematic viscosity
ratio of the dynamic viscosity **η** to the density of the fluid *ρ* expressed in kilograms per m^2, of the liquid measured at the same temperature (ν) = η/ ρ. The kinematic viscosity is usually expressed in m^2/s.

B2.37.3
inherent viscosity
ratio of the natural logarithm of the relative viscosity to the mass concentration of the polymer. Inherent viscosity is defined as:

$$\eta_{inh} = \frac{\ln \eta_{rel}}{c}$$

where **c** is the mass concentration of the polymer (g/dL) and η_{rel} is the relative viscosity, which is defined as

$$\eta_{rel} = \frac{\eta}{\eta_0}$$

where η is the viscosity of the solution and η_0 is the viscosity of the neat solvent. The unit of inherent viscosity is dL/g.

B.4 Chemical Substance subtypes and Mixture Substance

B.4.1 Substance type, Chemical substance

B.4.1.1 Stoichiometric chemical substances

ISO/TS 19844 states in the main body that: stoichiometric chemical substances are substances that contain complete chemical structures and definite stoichiometric ratios among moieties. Stoichiometric ratios between moieties should be whole integers and not fractions. They can be defined by a single representation of the complete molecular structure and a stereochemistry descriptor.

Chemical substances are defined on the basis of their covalent molecular structure; the presence of a salt (counter-ion) and/or solvates (water, alcohols, acetone) is captured. (Purity, grade, physical form or particle size are not taken into account in the definition of a chemical substance or in the assignment of a Substance ID.)

EXAMPLE: Purified Water, Water for Injection, Sterile Water for Injection USP, ice and steam all map to the substance Water.

In order to assign a Substance ID for a chemical substance, a complete covalent structure with all stereochemistry (Axial R or S and geometric stereochemistry E or Z) defined is needed. The stoichiometry (mole ratio) of counterions or solvates present in the material should also be provided.

For all chemical substances a molecular representation of the substance should be provided with all stereochemistry assigned or sites of unknown stereochemistry identified. An actual image of the structure can be provided in an accompanied document or a structural representation type 'Molfile', 'SMILES', 'InChI'.

The structural information typically provided to INN, USAN, BAN or JAN to assign a non-proprietary name is usually sufficient to define chemical substances. Active substances and excipients are typically defined in accordance with INN, JAN, or USAN definitions or USP/NF and European Pharmacopoeia (EP) monographs. The labelling requirements provided by monographs should be taken into account when defining and distinguishing substances. Information beyond monograph requirements may occasionally be necessary to distinguish substances."

B.4.2 Solid state forms of the Substance

B.4.2.1 Co-crystals

Over the last decade, **co-crystals** have gained considerable attention as alternative solid-state forms in drug development. By making co-crystals of pharmaceutically interesting substances, their solid state properties such as solubility, hygroscopicity and stability may be improved as well as manufacturing behaviour (compaction, flowability, filterability etc.). The salt formation is already widely used for this purpose, but with co-crystal formation this can now be achieved also for substances that lack the possibility of salt formation.

A principal tool is the hydrogen bond, which is responsible for the majority of directed intermolecular interactions in molecular solids. Co-crystals are multi-component crystals based on hydrogen bonding interactions without the transfer of hydrogen ions to form salts; this is an important feature, since Bronsted acid-base chemistry is not a requirement for the formation of a co-crystal. Co-crystallisation is a manifestation of directed self-assembly of different components.

Salts are formed in an acid-base reaction between the active pharmaceutical ingredient (API) and an acidic or basic substance by proton (H^+) transfer from acid (A) to base (B). The components are arranged in the crystal lattice dominantly based on ion paring.

The components in co-crystals assemble via weaker interactions such as e.g. hydrogen bonding, n-n stacking or van der Waals interactions. From a scientific point, solvates and hydrates can be considered as a subgroup of co-crystals. The solvent or the water act as conformer in the same way as other conformers.

Salt formation: A-H + B ↔ AB^+ -H (salt);
Co-crystal formation: A-H + B ↔ A-H…B (Co-crystal).

Both Co-crystals and salts have defined stoichiometries, similar solution specification characteristics, as well as a solubility product K_{sp}.

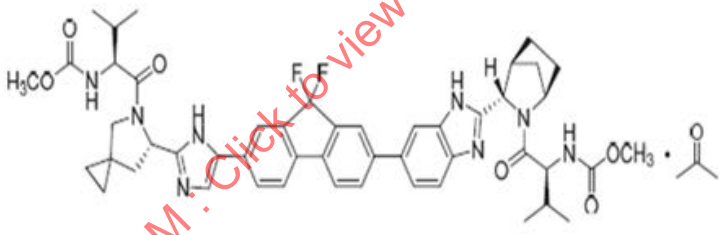
B.4.2.2 Solvates and hydrates

Water or solvent act as a conformer. They may be the result of a design with the aim to achieve crystals with certain properties, but may also rather be the result of a selection of the final solvent based on other criteria. To be able to distinguish hydrates and solvates from (other) co-crystals it has been stated that components of a co-crystal should exist as individual solids at ambient conditions. However, the terms "hydrates" and "solvates" are descriptive and widely used and should be therefore retained while keeping in mind that they are part of, rather than separate from, the general concept of co-crystals.

In ISO 11238 and ISO/TS 19844 these materials are considered as separate substances from the parent substance, describing the active moiety. Hydrates, solvates and co-crystals will have a separate Substance ID at the Substance level.

EXAMPLE 1 Disodium hydrogen phosphate (anhydrous) **[Na_2HPO_4]** will have a separate Substance ID from Disodium hydrogen phosphate dihydrate **[Na_2HPO_4. $2H_2O$]**.

EXAMPLE 2 Lepidasivir acetone solvate **[$C_{49}H_{54}F_2N_8O_6$. C_3H_6O]** will have a separate Substance ID from the parent substance Lepidasivir **[$C_{49}H_{54}F_2N_8O_6$]**.



Molecular formula: $C_{52}H_{60}F_2N_8O_7$ ($C_{49}H_{54}F_2N_8O_6$. C_3H_6O)
Molecular weight: 946,46 g/mol.

Systematic name (INN): methyl [(1*S*)-1-{(1*R*,3*S*,4*S*)-3-[5-(9,9-difluoro-7-{2-[(6*S*)-5-{(2*S*)-2-[(methoxycarbonyl)amino]-3-methylbutanoyl}-5-azaspiro[2.4]hept-6-yl]-1*H*-imidazol-4-yl}-9*H*-fluoren-2-yl)-1*H*-benzimidazol-2-yl]-2-azabicyclo[2.2.1]heptane-2-carbonyl}-2-methylpropyl]carbamate.

Figure B.3 — Illustrative representation of Lepidasvir acetone solvate

EXAMPLE 3 Edoxaban tosilate monohydrate **[$C_{24}H_{30}ClN_7O_4S$. $C_7H_8O_3S$. H_2O]** will have a separate Substance ID from the parent substance Edoxaban [$C_{24}H_{30}ClN_7O_4S$] and from Edoxaban tosilate anhydrous **[$C_{24}H_{30}ClN_7O_4S$. $C_7H_8O_3S$],** see Figure B.4.

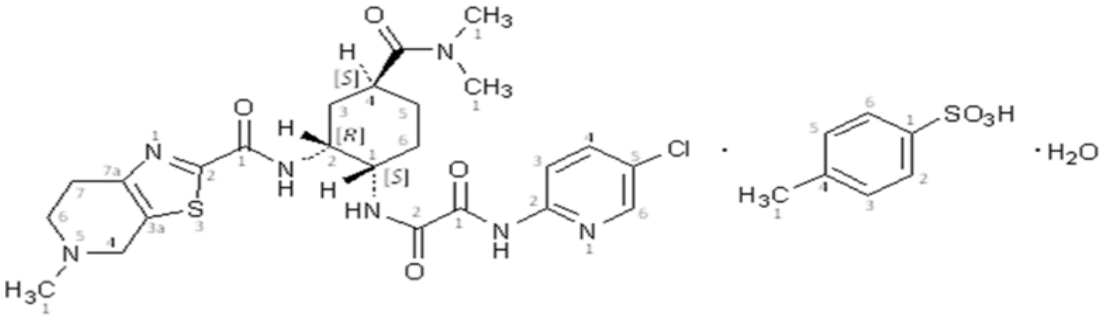


Figure B.4 — Full structure representation of Edoxaban tosilate monohydrate in accordance with the ISO/TS 19844, 4.9.1

NOTE **Systematic name (INN) of the base**: N-(5-chloropyridin-2-yl)-N'-[(1S,2R,4S)-4-(N,Ndimethylcarbamoyl)-2-(5-methyl-4,5,6,7-tetrahydro[1,3]thiazolo[5,4-c]pyridine-2-carboxamido)cyclohexyl]oxamide]

EXAMPLE 4 Trametinib **[$C_{26}H_{23}FIN_5O_4$]** is a novel active substance not described in any pharmacopoeia. Trametinib dimethyl sulfoxide **[$C_{26}H_{23}FIN_5O_4 \cdot C_2H_6OS$]** is a 1:1 stoichiometric DMSO solvate, where DMSO is fully incorporated into the crystal lattice. In solid state, one form has been identified (form 1). The active substance is a white to almost white solid, very slightly soluble in ethanol (non-solvated parent) and acetonitrile and slightly soluble in DMSO (solvated) and isopropyl acetate.

In Figure B.5 the structure of the parent substance Trametinib is provided as well as the structural relationship with its dimethyl sulfoxide solvate.

Structural representations of trametinib (parent), **figure A.1 and A.2 and its dimethyl sulfoxide solvate, figure B.1 and B.2.**

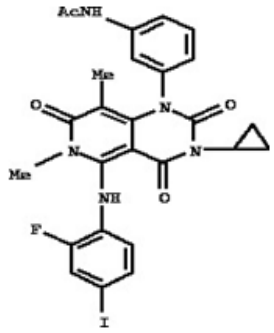


Fig: A .1 CAS Registry name and **CAS Number**: 871700-17-3 Acetamide, N-[3-[3-cyclopropyl-5-[(2-fluoro-4-iodophenyl)amino]-3,4,6,7-tetrahydro-6,8-dimethyl-2,4,7-trioxopyrido [4,3-d]pyrimidin-1(2H)-yl]phenyl]-

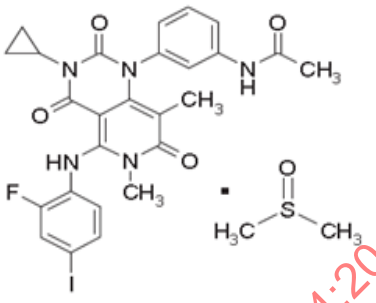


Fig: B.1: CAS Registry name and **CAS number** 2187431-43-1 Acetamide, N-[3-[3-cyclopropyl-5-[(2-fluoro-4-iodophenyl)amino] -3,4,6,7-tetrahydro-6,8-dimethyl-2,4,7-trioxopyrido[4,3-*d*]pyrimidin-1(2*H*)-yl]phenyl]-,compound with 1,1'-sulfinylbis[methane] (1:1)

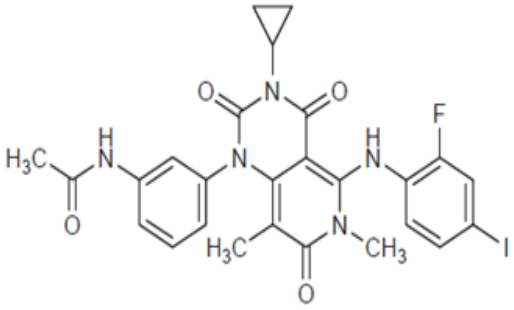


Fig: A.2 Recommended INN: List **67,** WHO Drug Information, Vol. 26, No. 1, 2012;
Trametinib; $C_{26}H_{23}FIN_5O_4$
N-{3-[3-cyclopropyl-5-[(2-fluoro-4-iodophenyl)amino]-6,8-dimethyl-2,4,7-trioxo-3,4,6,7-tetrahydropyrido[4,3-*d*]pyrimidin-1(2*H*)-yl]phenyl}acetamide; **Mw = 615.4**

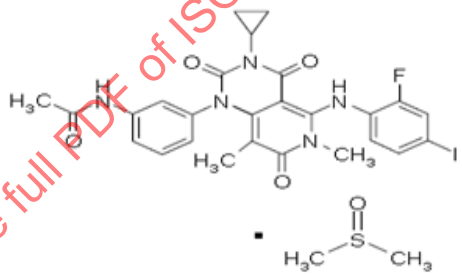


Fig: B.2: IUPAC name: Equimolecular combination of N-{3-[3-cyclopropyl-5-[(2-fluoro-4-iodophenyl)amino]-6,8-dimethyl-2,4,7-trioxo-3,4,6,7-tetrahydropyrido[4,3-d]pyrimidin-1(2H)-yl]phenyl}acetamide with (methylsulfinyl)methane;
Trametinib dimethyl sulfoxide; $C_{26}H_{23}FIN_5O_4 . C_2H_6OS$; Mw = 693.5 Da. (DMSO solvate of parent)

Figure B.5 — Structural representation of Trametinib and Trametinib dimethyl sulfoxide

NOTE Fig: A2, B1, B2 are correct structure representations in accordance with the ISO/TS 19844, as the complete chemical structures and definite stoichiometric ratios among moieties are described in 4.8.

B.4.3 Need to substantiate the chemical structure, molecular formula and molecular weight

When an INN-name and/or the systematic name of the substance is not described in a Pharmacopoeia of the EP, USP, or any other pharmacopoeia, the structural formula, molecular formula and molecular weight, as mandatory information, plus a Mass spectrum (and/or NMR-spectrum) should be supplemented as part of the Reference Source Document section (4.6.6-4.6.9) in order to substantiate the structure and molecular weight. The mass spectrum of trametinib dimethyl sulfoxide was obtained by positive ion electrospray ionisation with collisional induced dissociation (CID) of the protonated molecule of trametinib $(M+H)^+ = 616\ M/z$ to give the product ion spectrum. The fragment structure of the base peak $[C_{20}H_{17}FIN_4O_2]^+ = 491 M/z$ is shown in the Figure B.6.

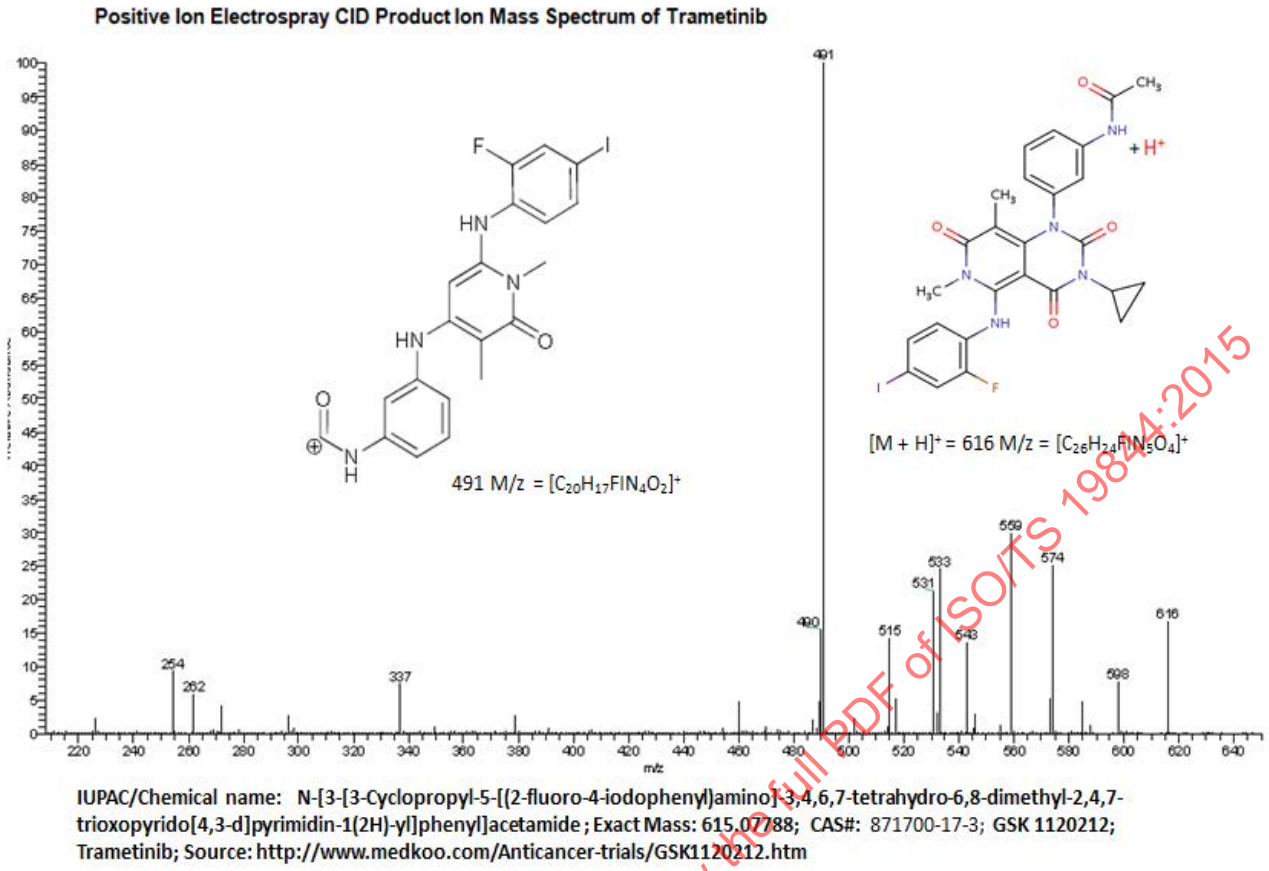


Figure B.6 — Mass Spectra and proposed structure of the molecular [M+1] ion and the [491]⁺ fragment Ion of Trametinib

B.4.4 Polymorphism

In the solid state, single as well as multiple entities, such as salt, hydrates, co-crystals etc., may exhibit polymorphism. Polymorphism is the ability of a compound in the solid state to exist in different crystalline forms having the same chemical composition. These different forms are formed by weak interactions between the components present in the solid state. These different forms may possess different physico-chemical properties.

In the ISO 11238 and ISO/TS 19844 these polymorphic forms are considered as a physical form type which is captured at the specified substance Group 1 level.

Separate polymorphic forms of the parent substance, describing the active moiety or the corresponding hydrates, solvates and co-crystals, will have a separate ID at the specified substance Group 1 level, in 6.1 and 6.1.12 Group 1 specified substance.

B.4.4.1 Example: Tibolone [$C_{21}H_{28}O_2$]

Tibolone is marketed containing a very small amount of the unwanted polymorphic form.

It is known that different polymorphic forms could have different influence on the solubility of Tibolone and consequently on the dissolution rate of Tibolone from the drug product. According to the Note for guidance on Specifications: Test Procedures and Acceptance Criteria for New Drug Substances and New Drug Products: Chemical Substance (ICH Topic Q 6A), the polymorphic forms of the active substance should be adequately specified.

Tibolone, form 1 has a monoclinic structure with four molecules in the unit cell, two molecules A and two molecules B linked one after another. Crystal form 1 has two types of hydrogen bond, between $O_{(17)}H$ of one type of molecule and the $O_{(3)}$ carbonyl of the other type of the molecule.

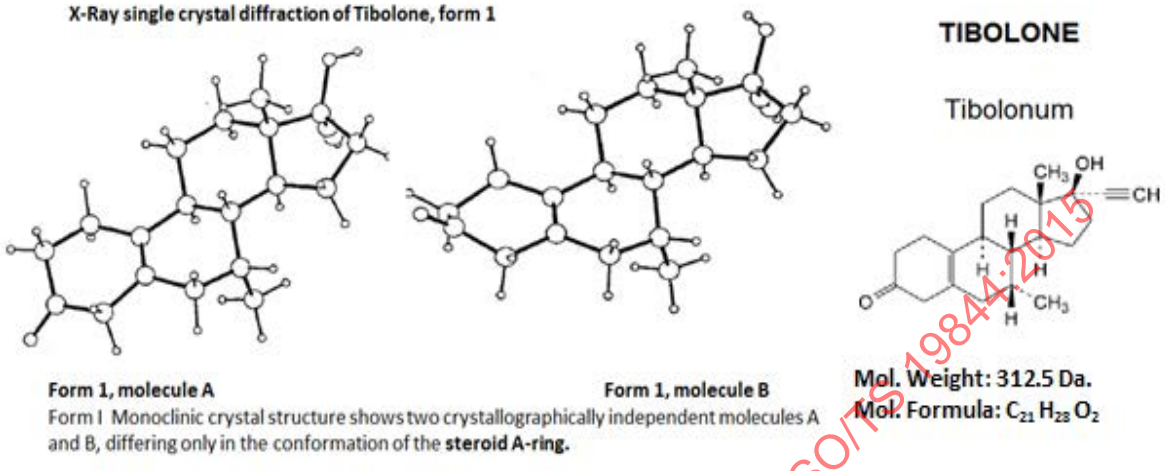


Figure B.7 — Representation of Tibolone monoclinic crystal structure and triclinic crystal structure

Tibolone, form 2 has a triclinic crystal structure with one molecule in the unit cell. This molecule **resembles B** in crystal form 1. **Form 2 has only one type of hydrogen bond.**

The company manufactures the API Tibolone only with the polymorphic form 1 because drug product batches with this form show much higher bioavailability than batches manufactured with polymorphic form 2 and form 1 shows more stability than form 2.

NOTE Source: Araujo, Gabriel *et al.* "Thermal studies on polymorphic structures of Tibolone", Journal of Thermal Analysis & Calorimetry; Oct 2010, Vol. 102 Issue 1, p233; and Gabriel Lima Barros de Araújo; Solid state characterization of tibolone polymorphs, Dissertation University São Paulo, 2009; http://www.teses.usp.br/teses/disponiveis/9/9139/tde-31032010-091615/es.php; (The results confirm that two pure polymorphic forms for Tibolone were crystallised: Form I, monoclinic, obtained by recrystallization from acetone and water and Form II, triclinic, recrystallized from toluene or hexane. Substantial differences were detected comparing the polymorphs with regard to their solid-state properties.)

Conclusion: The property Polymorphic form will be captured at the specified substance Group 1 information level. Tibolone, polymorph form 1 will have a different Specified substance Group _ID from Tibolone, polymorphic form 2.

B.4.5 Non-stoichiometric chemical substances

Non-stoichiometric chemical susbstances are substances having variable ratios among moieties; each moiety in itself has a discrete molecular structure. (E.g. The monosulfides of the transition metals are often nonstoichiometric. Iron(II) sulfide (the mineral Pyrrhotide) with a composition Fe(1-x)S (x = 0 to 0.2), in 5.1.5.

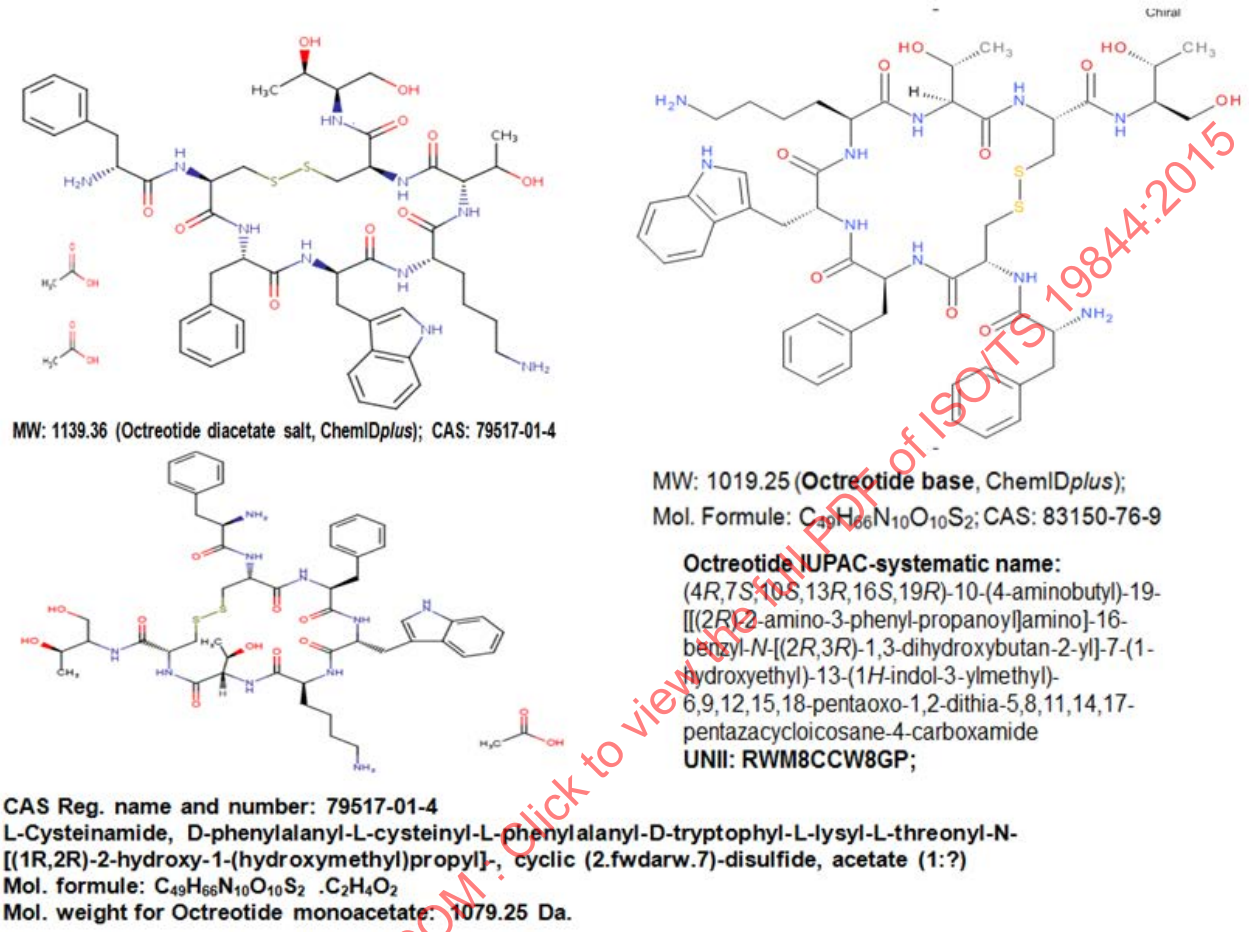
B.4.5.1 Example: USP monograph: Aluminium Chlorohydrex Propylene Glycol

Aluminium Chlorohydrex Propylene Glycol is a complex of aluminium chlorohydrate and propylene glycol in which some of the waters of hydration of the aluminium chlorohydrate have been replaced by propylene glycol. It contains the equivalent of not less than 90.0 percent and not more than 110.0 percent of the labelled amount of anhydrous aluminium chlorohydrate.

Molecular formula: $Al_y(OH)_{3y-z}Cl_z{\cdot}nH_2O{\cdot}mC_3H_8O_2$ $Al_2(H_2O)_{y-z}(OH)_{6-n}(Cl)_n(C_3H_8O_2)_z$

B.4.5.2 Example: Octeotride Acetate

An example for non-stoichiometric organic chemical substance is Octreotide acetate. The acetate moiety can vary by 1-2 molecules. In Figure B.8, Octreotide base **[$C_{49}H_{66}N_{10}O_{10}S_2$],** Mol. weight 1019.25, Octreotide monoacetate **[$C_{49}H_{66}N_{10}O_{10}S_2 . C_2H_4O_2$],** and Octreotide diactetate **[$C_{49}H_{66}N_{10}O_{10}S_2 . 2C_2H_4O_2$]** are shown obtained form FDA-SRS (base) and from Chem*ID*plus for the mono and diacetate salts.



MW: 1139.36 (Octreotide diacetate salt, ChemIDplus); CAS: 79517-01-4

MW: 1019.25 (**Octreotide base**, ChemIDplus);
Mol. Formule: $C_{49}H_{66}N_{10}O_{10}S_2$; CAS: 83150-76-9

Octreotide IUPAC-systematic name:
(4*R*,7*S*,10*S*,13*R*,16*S*,19*R*)-10-(4-aminobutyl)-19-[[(2*R*)-2-amino-3-phenyl-propanoyl]amino]-16-benzyl-*N*-[(2*R*,3*R*)-1,3-dihydroxybutan-2-yl]-7-(1-hydroxyethyl)-13-(1*H*-indol-3-ylmethyl)-6,9,12,15,18-pentaoxo-1,2-dithia-5,8,11,14,17-pentazacycloicosane-4-carboxamide
UNII: RWM8CCW8GP;

CAS Reg. name and number: 79517-01-4
L-Cysteinamide, D-phenylalanyl-L-cysteinyl-L-phenylalanyl-D-tryptophyl-L-lysyl-L-threonyl-N-[(1R,2R)-2-hydroxy-1-(hydroxymethyl)propyl]-, cyclic (2.fwdarw.7)-disulfide, acetate (1:?)
Mol. formule: $C_{49}H_{66}N_{10}O_{10}S_2 . C_2H_4O_2$
Mol. weight for Octreotide monoacetate: 1079.25 Da.

Figure B.8 — Structural representation of Octreotide Acetate

In Figure B.8, the coloured structures are in agreement with 4.8, but do not show the R,S-denominators. These are described in the systematic name.

B.4.6 Mixture substance

A mixture substance is a type of polydisperse substance that is a combination of single substances isolated together or synthesised/ obtained or produced in the same process. Two examples are provided:

B.4.6.1 Teicoplanin (EP)

Definition: Teicoplanin is a mixture of glycopeptides produced by certain strains of *Actinoplanes teichomyceticus sp.*; the fermentation product contains 6 principal components of the mixture; teicoplanin A_{2-1} to A_{2-5} and teicoplanin A_{3-1}. The chemical structure consists of 6 components of glycopeptides, which are composed of a hepta peptide core of 7 amino acids connected with 3 sugars N-acetylglucosamine, α-mannitose and the glucose substituted by different N-acylamino moieties.

Chirality: The configuration of the 7 amino acids is: 1R, 2R, 3S, 4R, 6S and 7S. Five formations of the six amide linkages are trans, the 5-6 linkage is cis. The mixture substance contains as active multiple active components two pairs of constitutional isomers ($A_{2\text{-}4/5}$) and ($A_{2\text{-}2/3}$), see Figure B.9.

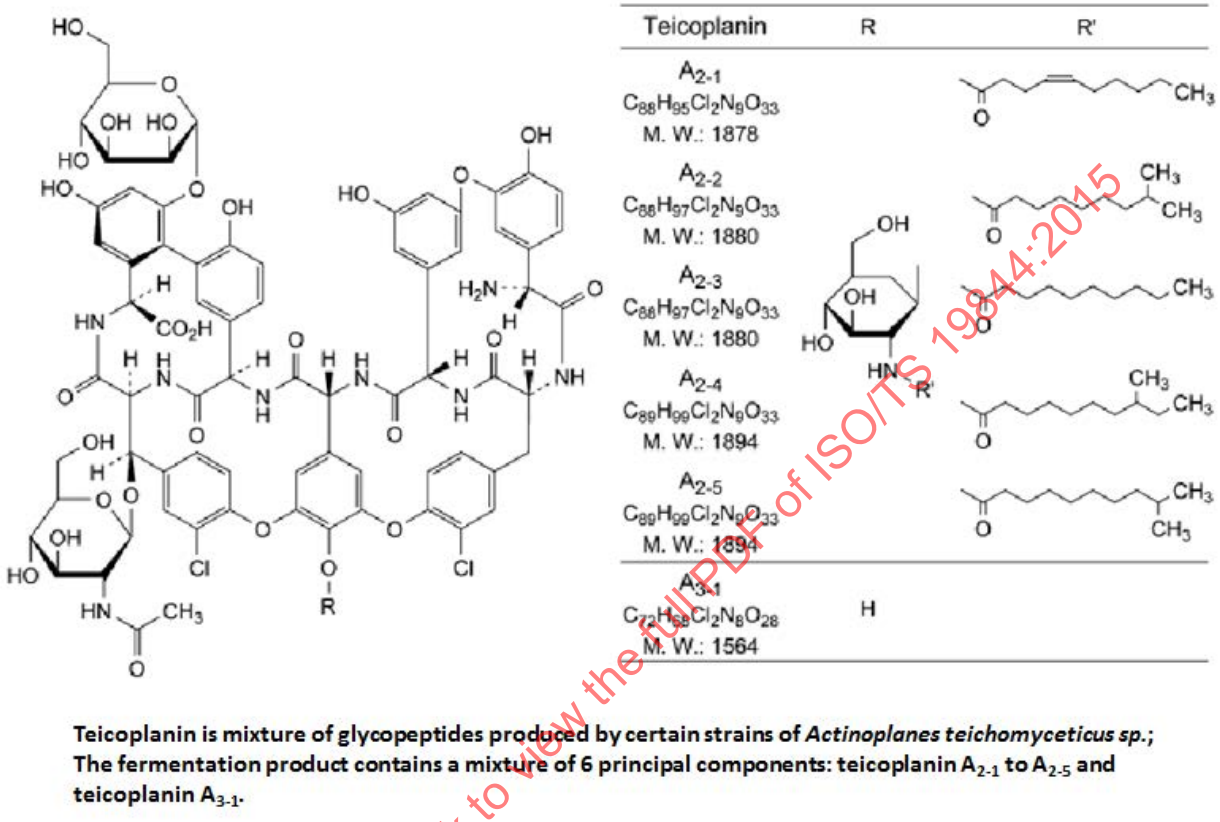


Figure B.9 — Structural representation of Teicoplanin (EP)

B.4.7 Multi substance material

Multi substance material is defined as single substances of diverse origin that are brought together and do not undergo a chemical transformation. It is considered as multi-substance material (specified substance Group 1) and not a mixture substance (e.g. Simethicone consist of two substances: dimethicone and silicon dioxide). These substances are not typically isolated together or synthesised together, but physically mixed together. They would be defined as a specified substance Group 1 substance.

B.4.7.1 Example: Paclitaxel-Albumin complex

Abraxane is presented as a sterile lyophilisate for suspension for injection. Before use the product has to be reconstituted with 0,9% sodium chloride solution to obtain a suspension containing 5 mg paclitaxel per ml. Paclitaxel is present in the form of non-covalently albumin-bound nanoparticles with a mean size of approximately 130 nm. Other ingredients include human albumin solution, water for injections and nitrogen. Only human albumin is left in the finished product.

Paclitaxel is a known active substance described in the EP and the USP. The chemical name of paclitaxel is: 5β,20-epoxy-1,2α,4,7β,10β,13α-hexahydroxytax-11-en-9-one 4,10-diacetate 2-benzoate 13-ester with (2*R*,3*S*)-N-benzoyl-3-phenylisoserine, see Figure 11. It is a white or almost white crystalline powder, practically insoluble in water, soluble in methanol and freely soluble in dichloromethane. Paclitaxel has 11 stereogenic centres and may exist in different crystalline forms. However only one crystal form of paclitaxel is found in the active substance used in the manufacture of the finished product.

In Abraxane the solubility problem of paclitaxel is solved by reducing its size at nano-scale level, thereby greatly increasing the surface area of the particles and improving dissolution. Human albumin functions as a surface-active polymer providing charge and steric stabilisation to the paclitaxel nanoparticles to prevent aggregation. Stabilisation is achieved by the fact that albumin adsorbs onto the surface of the paclitaxel nanoparticles, thus creating a sheet that functions as a surface-active polymer preventing aggregation of paclitaxel particles. The interaction between paclitaxel and human albumin is weak and both substances freely dissociate after reconstitution.

Conclusion: Paclitaxel as nanoparticles is stabilised by human albumin in a ratio 1:8 (w/w) based on stability studies by weak interaction between the nanoparticles and the human albumin, but the two substances are not typically isolated together in the same manufacturing process or form covalent bonds or are packed in a lattice or form a salt.

In terms of the ISO standard Paclitaxel, nanoparticles — Albumin "complex" is considered as a multi substance material consisting of two separate substances Paclitaxel and Albumin, having a weak interaction. The Paclitaxel-Albumin 1:8 ratio complex is captured at the specified substance Group 1 level. The structure of Paclitaxel and names are shown in Figure B.10.

In Figure B.11 the structure, names and molecular formula and molecular weight of Paclitaxel and Paclitaxel Poliglumex substances are presented.

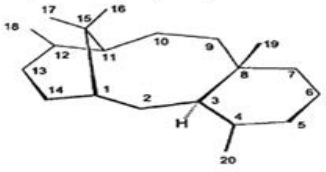
Paclitaxel Poliglumex is a covalent modification of the Paclitaxel molecule. Paclitaxel Poliglumex is captured at the substance level.

Paclitaxel Poliglumex is described according to the WHO Drug Information, Vol. 18, No. 3, 2004 Recommended INN: List 52 as: poly(L-glutamic acid) partly γ-esterified by (2R,3S)-3-benzamido-1-{[4,10β-bis(acetoxy)-2α-(benzoyloxy)-1,7β-dihydroxy-9-oxo-5,20-epoxytax-11-en-13α-yl]oxy}-1-oxo-3-phenylpropan-2-yl. The "coloured" structure is a structural representation of both the fragments. The "uncoloured" structure is the structural representation following the ISO/TS 19844 for polymers.

Paclitaxel Poliglumex is described in the FDA-SRS system by:
" BIODEGRADABLE POLYMER-DRUG CONJUGATE RESULTING FROM THE RANDOM CONDENSATION OF POLY-L-GLUTAMIC ACID AND PACLITAXEL CONTAINING ONE PACLITAXEL ESTER LINKAGE PER ELEVEN GLUTAMIC ACID UNITS (MW 48000)" ; **UNII:** TQ64FZ98ZN.

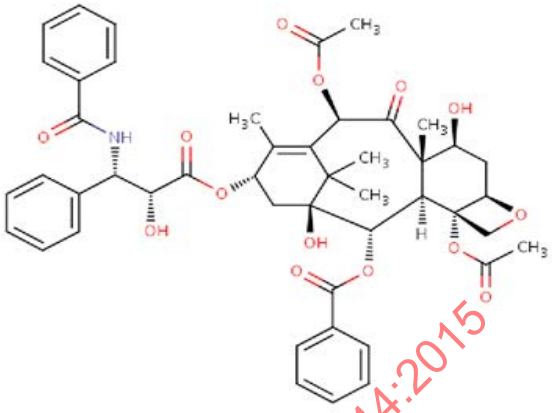
USAN-Name: Paclitaxel

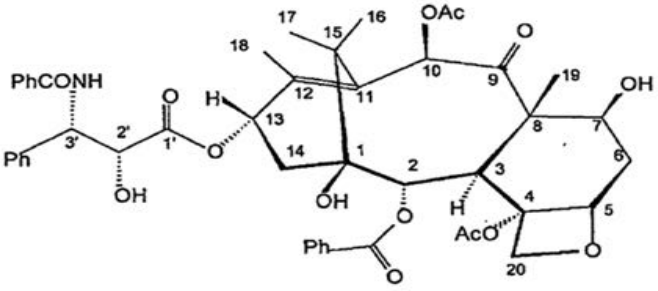
According to this system, Paclitaxel is "5β,20-epoxy-1,2α,4,7β,13α-hexa-hydroxytax-11-en-9-one 4,10-diacetate 2-benzoate 13-ester with (2R,3S)-N-benzoyl-3-phenylisoserine.



Formula I - Taxane skeleton

IUPAC, Numbering system according to Commission on the Nomenclature of Organic Chemistry (CNOC), *"Nomenclature of Organic Chemistry, Section F: Natural Products and Related Compounds"*, Eur. J. Biochem. 1978, 86, 1-8)





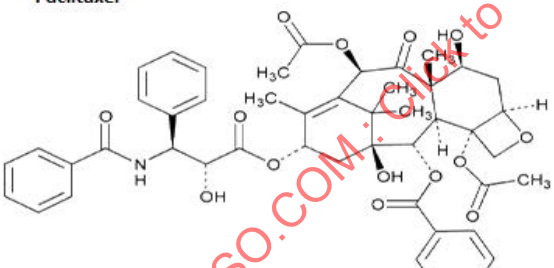
Systematic name:
IUPAC: (αR,βS)-β-(Benzoylamino)-α-hydroxy-Benzenepropanoic acid, (2aR,4S,4aS,6R,9S,11S,12S,12aR,12bS)-6,12b-bis(Acetyloxy)-12-(Benzoyloxy) 2a,3,4,4a,5,6,9,10,11,12,12a,12b-Dodecahydro-4,11-dihydroxy-4a,8,13,13-tetramethyl-5-oxo-7,11,methano-1H-cyclodeca[3,4]benz[1,2-b]oxet-9-yl-ester.

Mol. formula: $C_{47}H_{51}NO_{14}$; **Mol. weight:** 853.9.

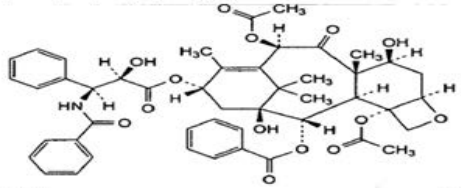
CAS Name Benzenepropanoic acid, .beta.-(benzoylamino)-.alpha.-hydroxy-, (2aR,4S,4aS,6R,9S,11S,12S,12aR,12bS)-6,12b-bis(acetyloxy)-12-(benzoyloxy)-2a,3,4,4a,5,6,9,10,11,12,12a,12b-dodecahydro-4,11-dihydroxy-4a,8,13,13-tetramethyl-5-oxo-7,11-methano-1H-cyclodeca[3,4]benz[1,2-b]oxet-9-yl ester, (.alpha.R,.beta.S)-
CAS number: 33069-62-4;

Figure B.10- Structural representation of the Paclitaxel skeleton and Paclitaxel

Paclitaxel



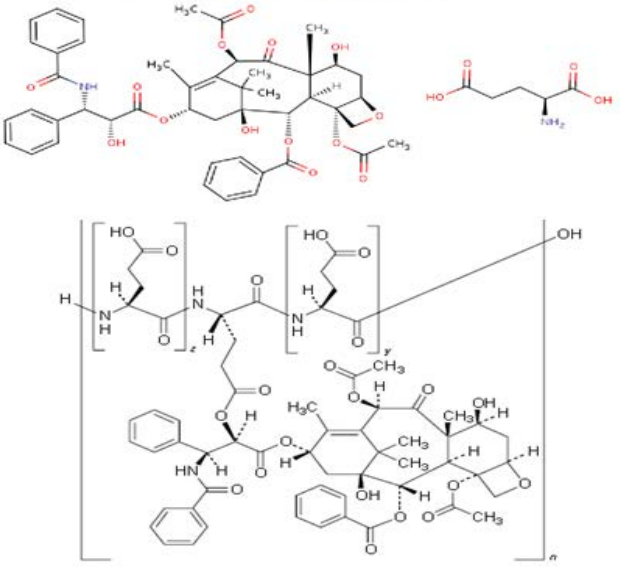
Structure USP: Benzenepropanoic acid, -(benzoylamino)- -hydroxy-, 6,12b-bis(acetyloxy)-12-(benzoyloxy)-2a,3,4,4a,5,6,9,10,11,12,12a,12b-dodecahydro-4,11-dihydroxy-4a,8,13,13-tetramethyl-5-oxo-7,11-methano-1*H*-cyclodeca[3,4]benz[1,2-b]oxet-9-yl ester, [2a*R*-[2a ,4 ,4a ,6 ,9 (*R**, *S**),11 ,12 ,12a ,12b]]-. **or**
(2a*R*,4*S*,4a*S*,6*R*,9*S*,11*S*,12*S*,12a*R*,12b*S*) 1,2a,3,4,4a,6,9,10,11,12,12a,12b-Dodecahydro-4,6,9,11,12,12b-hexahydroxy-4a,8,13,13-tetramethyl-7,11-methano-5*H*-cyclodeca[3,4]-benz[1,2-*b*]oxet-5-one 6,12b-diacetate, 12-benzoate, 9-ester with (2*R*,3*S*)-***N*-benzoyl-3-phenylisoserine**
Structure EP- monograph: 5β,20-Epoxy-1,7β-dihydroxy-9-oxotax-11-ene-2α,4,10β,13α-tetrayl 4,10-diacetate 2-benzoate 13-[(2*R*,3*S*)-3-(benzoylamino)-2-hydroxy-3-phenylpropanoate].



$C_{47}H_{51}NO_{14}$
[33069-62-4]

M, 854

Paclitaxel Poliglumex: BIODEGRADABLE POLYMER-DRUG CONJUGATE RESULTING FROM THE RANDOM CONDENSATION OF POLY-L-GLUTAMIC ACID AND PACLITAXEL CONTAINING ONE PACLITAXEL ESTER LINKAGE PER ELEVEN GLUTAMIC ACID UNITS (MW 48000); Mol. formula: $(C_{52}H_{56}N_2O_{16})n\,(C_5H_7NO_3)x$



WHO Drug Information, Vol. 18, No. 3, 2004 **Recommended** INN: List 52: **Paclitaxel Poliglumex** [USAN:INN]
poly(L-glutamic acid) partly γ-esterified by (2R,3S)-3-benzamido-1-[[4,10β-bis (acetoxy)-2α-(benzoyloxy)-1,7β-dihydroxy-9-oxo-5,20-epoxytax-11-en-13α-yl]oxy}-1-oxo-3-phenylpropan-2-yl;

Figure B.11 — Structural representation of Paclitaxel and Paclitaxel Poliglumex

B.5 Discussion of the key elements of a chemical substance

B.5.1 Identity of material

The identity, nomenclature and chemical structure of the substance type chemical substance which is the subject of an application for marketing authorisation, or is under investigation, or is already marketed, is characterised by 5 key elements mentioned below of which in B.3 the solid state form already has been discussed:

- Substance name and/or Company-code
- Substance structure
- Substance molecular formula
- Substance molecular weight
- Physical form

In the main body of ISO/TS 19844 the first two elements are extensively discussed in 4.5 (Substance names), 4.6 (Reference Sources), 4.7 (Substance Code), 4.8 (Structure representation) whereas abbreviations and molecular formula and molecular weight are described.

The structure representation of the chemical substances shown in the figures in B.3 of this annex is discussed for each figure with respect to 4.8. Additional information will be discussed in the next clauses.

B.5.2 Nomenclature

Regarding the nomenclature for chemical substances, a hierarchy for naming substances is used; **hierarchy of Official Name Types.**

a) **INN:** (Primary Official Name_Type, Official Name Status, see 4.5.4) following the WHO-application International Nonproprietary Name rules.

NOTE An INN for a new chemical entity does not routinely specify the stereoisomeric state of the molecule in the non proprietary name. If stereochemistry has been determined, then this information is presented in the chemical name(s) to identify the substance.

An INN can, therefore, represent the racemic mixture (e.g. ibuprofen), the *levo*-isomer (e.g. amifostine), or the *dextro*-isomer (e.g. butopamine).

b) **Other Official names and Pharmacopoeial names:** Typical nonproprietary names used in a given jurisdiction: USAN, BAN, MARTINDALE, JAN, EP, USP, JP, DAB.

c) **Systematic names:** IUPAC name, CAS Registry Name and definitions of a Pharmacopoeial Monograph.

d) **Codes**: In some cases a **Company code** can replace a name for a substance under development when name is not available.

B.5.3 Molecular formula

The molecular formula is directly derived from the molecular structure and is represented following the **Hill system.** The Hill system (or Hill notation) is a system of writing chemical formulas such that the number of carbon atoms in a molecule is indicated first, the number of hydrogen atoms next, and then the number of all other chemical elements subsequently, in alphabetical order. When the formula contains no carbon, all the elements, including hydrogen, are listed alphabetically. In the ISO/TS 19844 the molecular formula for an organic salt, solvate or hydrate will notate first the molecular formula of the Parent substance followed by the molecular formula of the salt, solvate or hydrate.

The molecular formula of Amlodipine besilate will be presented by both **$C_{20}H_{25}ClN_2O_5.C_6H_6O_3S$** and **$C_{26}H_{31}ClN_2O_8S$.** The dot between the two moieties is following the IUPAC recommendation. By the terms:

Molecular formula by moieties: $C_{20}H_{25}ClN_2O_5.C_6H_6O_3S$ and **Molecular formula** following the Hill system: $C_{26}H_{31}ClN_2O_8S$.

Inorganic acids and salts and metal salts are shown without charges or bonds: $HClO_4$ and $KMnO_4$ respectively.

If metal salts of inorganic acids include several metals, the symbols for the metals are shown in alphabetic order (e.g. K_2NaPO_4).

In salts of **inorganic acids**, the metal precedes the hydrogen (e.g. NaH_2PO_4). Molecules of water of crystallisation or of substances of solvation follow the formula of the salt. (e.g. $H_3PO_4.5H_2O$).

Hydrates and solvates of **organic acids**: Sodium acetate trihydrate will be represented as $C_2H_3NaO_2.3H_2O$. In metal salts of organic acids and in metal compounds of alcohols, phenols (and their sulfur, selenium and tellurium analogues), amines and amides, the metal symbol usually replaces the acid hydrogen.

In Non-cyclic linear structures like Sodium nitroprusside: $Na_2[Fe(CN)_5(NO)].2H_2O$, a non-cyclic structure is constructed in the following order:

a) symbol of the central atom placed on the left,

b) ionic ligands with cations first then anions,

c) neutral ligands.

B.5.4 Molecular weight

Molecular mass or **molecular weight** refers to the mass of a molecule and is also directly derived from the molecular structure or the molecular formula. It is calculated as the sum of the mass of each constituent atom multiplied by the number of atoms of that element in the molecular formula. For stoichiometric chemicals the molecular weight is calculated from the molecular formula using standard masses for each of the elements. In ISO/TS 19844 the molecular mass refers to the complete structure or a moiety or a fragment.

For Amlodipine besilate it is the sum of the molecular mass of the free base + the organic salt = 567.06 Da or 567.06 g/mol.

NOTE1 408.88 g/mol. (Parent Amlodipine), 158.18 g/mol. (Besilate).

NOTE 2 For a substance not described in a Pharmacopoeia (EP, USP, JP) a mass spectrum should be provided to substantiate the calculated molecular weight.

B.5.5 Substance structure

Subclause 4.8 describes that structural information is an essential element for chemical substances and for other types of substances that have structurally definable modifications. The representation should contain structural information in one or more of the standardised formats, 4.8.1 Structural Representation Type.

A FULL structure is a structure in which the complete connectivity of the substance is known and the substance is monodisperse.

A PARTIAL structure is a structure in which either complete connectivity or stereochemistry is not defined.

A REPRESENTATIVE structure is used when the connectivity of the underlying material is diverse and a single structure is needed to represent that material.

Sometimes the term “illustrative” or “incorrect” in this Annex B is used to indicate that some elements of a structure are not presented in accordance with the ISO/TS 19844 Implementation Guide.

The structure is “incomplete” to the definition of FULL structure and therefore called “illustrative” in the examples of Annex B.

B.5.5.1 Stereochemistry

The stereochemistry of the substance shall be indicated in the structure if it can be, as is stated in 4.8.4. Special cases of stereochemistry that cannot be indicated in the structure shall be described based on a controlled vocabulary. Figure B.12 shows a graphical representation (image) of a full structural representation type and the descriptors assigned in the structure: Racemic, Axial S and Axial R respectively.

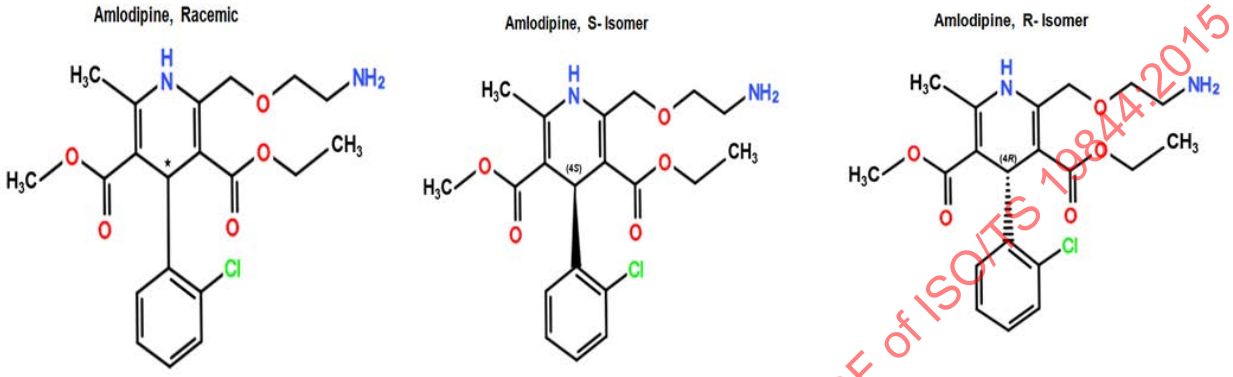


Figure B.12 — Full structural representation of Amlodipine

This means that the stereochemistry of mononuclear complexes is expressed by means of special descriptors formed from an abbreviated for the central atom geometry and the coordination number, see Figure B.12.

NOTE In a **tetrahedral molecular geometry**, a central atom is located at the centre with four substituents that are located at the corners of a tetrahedron. The bond angles are $\cos^{-1}(-1/3) \approx 109.5°$ when all four substituents are the same, as in CH_4 The perfectly symmetrical tetrahedron belongs to point group T_d, but most tetrahedral molecules are not of such high symmetry. Tetrahedral molecules can be chiral. See Figure B.13.

The ISO 19844 Substance Implementation Guide provides in 4.8.4 the following information:

The overall stereochemistry of the substance should be indicated in this field.

User Guidance	The stereochemistry of the substance shall be indicated in the structure and should be captured here. Special cases of stereochemistry that can’t be indicated in the structure shall be described based on a controlled vocabulary. Mixtures of stereoisomers shall be represented explicitly as a mixture of substances with absolute stereochemistry. In case the absolute stereochemistry is unknown the substance definition should be marked as “Incomplete”.
Example(s)	ABSOLUTE, AXIAL R, AXIAL S, SQUARE PLANAR 1, SQUARE PLANAR 2, SQUARE PLANAR 3, SQUARE PLANAR 4, TETRAHEDRAL, OCTAHEDRAL 12, OCTAHEDRAL 22, OCTAHEDRAL 21, RACEMIC, MIXED, ACHIRAL, EPIMERIC, UNKNOWN
Conformance	CONDITIONAL
Data Type	CD
Value Allowed	**AXIAL R, AXIAL S, SQUARE PLANAR 1, SQUARE PLANAR 2, SQUARE PLANAR 3, SQUARE PLANAR 4, TETRAHEDRAL, OCTAHEDRAL 12, OCTAHEDRAL 22, OCTAHEDRAL 21; RACEMIC, MIXED, ACHIRAL , ABSOLUTE , EPIMERIC, UNKNOWN,**
Business Rule(s)	When the Substance type (as defined in paragraph 4.3) is either chemical, protein, polymer or nucleic acid, this class becomes MANDATORY

NOTE Clarification of the Value Allowed in conformance table 4.8.4 , Source: NCI thesaurus: http://ncit60.nci.nih.gov/

Absolute **Definition:**	**Chiral** An object or molecule that is not superimposable on its mirror image. The classic description of chirality is the comparison of the hand in a mirror, because the mirror image is the opposite. A chiral molecule has one or more chiral centres
TETRAHEDRAL **Definition:**	**Tetrahedral Molecular Geometry** A molecule in which the central atom is bound to four atoms that form the vertices of a tetrahedron and are described by the chirality symbols (R) and (S)
Axial R **Definition:**	Chirality axis based stereochemistry where, using the Cahn-Ingold-Prelog (CIP) priority rules, the priorities (from highest to lowest) of the substituent atoms exist in a clockwise direction.
Axial S **Definition:**	Chirality axis based stereochemistry where, using the Cahn-Ingold-Prelog (CIP) priority rules, the priorities (from highest to lowest) of the substituent atoms exist in a counter clockwise direction
RACEMIC **Definition:**	**Racemate** An equimolar mixture of two enantiomers. A racemate is optically inactive.
EPIMERIC **Definition:**	**Epimeric Centres** A particular kind of diastereoisomers that differ in their stereochemistry at only one chiral centre; the stereochemistry at all other chiral centres is identical.
Cis-Trans Geometry	
SQUARE PLANAR 1 **Definition:**	**Square Planar 1 Molecular Geometry** Square planar molecular geometry where the Cahn-Ingold-Prelog (CIP) priority of the donor atom directly across from the priority 1 donor atom is also 1.
SQUARE PLANAR 2 **Definition:**	**Square Planar 2 Molecular Geometry** Square planar molecular geometry where the Cahn-Ingold-Prelog (CIP) priority of the donor atom directly across from the priority 1 donor atom is 2.
SQUARE PLANAR 3 **Definition:**	**Square Planar 3 Molecular Geometry** Square planar molecular geometry where the Cahn-Ingold-Prelog (CIP) priority of the donor atom directly across from the priority 1 donor atom is 3.
SQUARE PLANAR 4 **Definition:**	**Square Planar 4 Molecular Geometry** Square planar molecular geometry where the Cahn-Ingold-Prelog (CIP) priority of the donor atom directly across from the priority 1 donor atom is 4.
OCTAHEDRAL 12 **Definition:**	**Octahedral 12 Molecular Geometry (Trans Octahedral)** Octahedral molecular geometry where the Cahn-Ingold-Prelog (CIP) priority of the lowest priority donor atom directly across from (trans to) a priority 1 donor atom is also 1 and the priority of the donor atom across from (trans to) the highest priority atom in the plane perpendicular to this reference axis is 2.
OCTAHEDRAL 22 **Definition:**	**Octahedral 22 Molecular Geometry (trans Octahedral 22)** Octahedral molecular geometry where the Cahn-Ingold-Prelog (CIP) priority of the lowest priority donor atom directly across from (trans to) a priority 1 donor atom is 2 and the priority of the donor atom across from (trans to) the highest priority atom in the plane perpendicular to this reference axis is 2.
OCTAHEDRAL 21 **Definition:**	**Octahedral 21 Molecular Geometry (trans Octahedral 21)** Octahedral molecular geometry where the Cahn-Ingold-Prelog (CIP) priority of the lowest priority donor atom directly across from (trans to) a priority 1 donor atom is 2 and the priority of the donor atom across from (trans to) the highest priority atom in the plane perpendicular to this reference axis is 1.

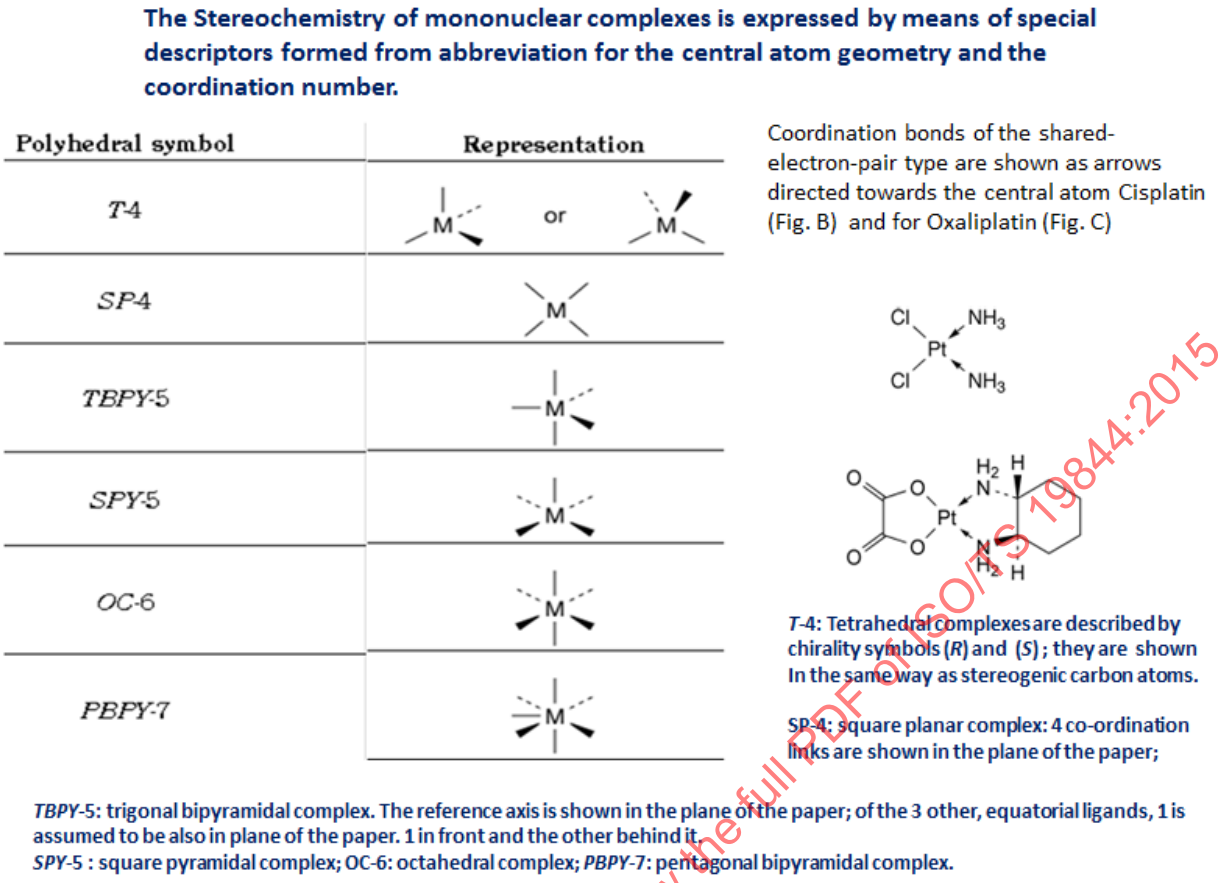


Figure B.13 — Stereochemistry of mononuclear complexes

B.5.5.2 General aspects of stereochemistry

Stereochemistry is concerned with the 3-dimensional arrangement of atoms in molecules, and stereoisomers are isomers with no difference in connectivity or bond multiplicity, but whose atomic special arrangements differ. A broken line denotes a bond projecting behind the plane of the paper and a filled wedge denotes one projecting in front of that plane. A line of normal thickness denotes a bond lying in the plane of the paper, see Figure B.14. Isomers can be classified in Stereoisomers (Enantiomers and Diastereomers), Geometric isomers, Conformers and Constitutional isomers. The definitions are described B.2. In the following sections Stereoisomers and Geometric isomers will be discussed.

B.5.6 Geometric isomerism

B.5.6.1 (*E*/*Z*) isomerism

Compounds containing carbon-carbon double bonds are represented as lines at an angle to each other. The representation of (*E*/*Z*) isomers shows that the (*Z*) configuration produces a bend in the chain whereas the (*E*) configuration does not. Figure B.14 presents example 1-Phenyl-1,3-pentadiene.

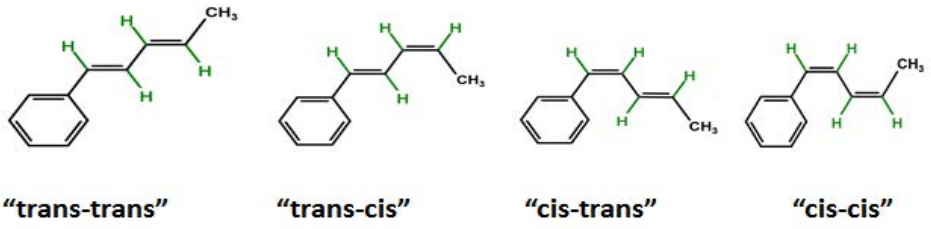


Figure B.14 — Stereochemistry of 1-Phenyl-1,3-pentadiene

It should be noted that the hydrogen atoms attached to the two carbon atoms forming the double bond are omitted in the case of (*E/Z*)-isomerism. The same convention is used for the isomers of imines / oximes and compounds containing several bonds, as in Figure B.15.

B.5.6.2 *Cis-Trans* Ring-isomerism

Cis-trans Ring-isomerism (chirality) is defined for a molecule with a plane of symmetry as the existence of *cis* and *trans* isomerism according to the arrangement of the substituents on a ring with respect to the plane of this ring, as in Figure B.15.

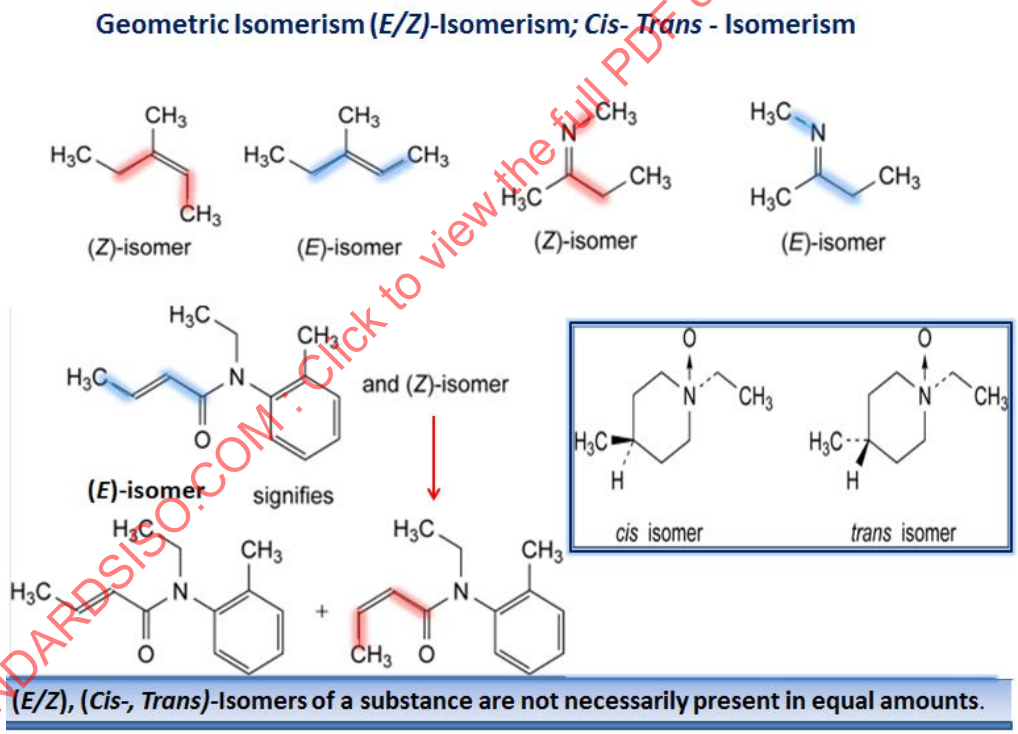


Figure B.15 — Geometric isomerism

B.5.7 Stereo-descriptors in systematic nomenclature: Substance with one centre of asymmetry

An enantiomer is one of a pair of molecular entities that are mirror images of each other and are non-superimposable. The descriptors (*R*) and (*S*) are symbols proposed by Cahn, Ingold and Prelog (CIP) and indicate the chiral absolute configuration around 4-coordinate (quadriligant) and 6-coordinate (sexiligant) stereogenic centres.

NOTE The CIP sequence rules are explained in B.2. **When the enantiomeric pair is of equal amounts the pair is called a racemic mixture or the substance is a 'racemate".** (E.g. Amlodipine, Figure B.12). The element group capturing Amlodipine besilate are described in B.6.1, tabular example 1.

B.5.8 Substance with two centres of Asymmetry, Epimers, Diastereomers

Epimers are diastereomers that have the opposite configuration at only 1 of 2 or more tetrahedral stereogenic centres present in the respective molecular entities. If a substance has two centres of asymmetry, there are four different stereoisomers. Two epimers of this substance have the same configuration at one centre of asymmetry and a different configuration of the other. Figure B.16 presents an example of the enantiomer and diasteriomers of CAS number 191226-98-9 with the CAS Registry name:

> Carbamic acid, N-[(1S,2R)-2-hydroxy-3-[(2-methylpropyl)[(4-nitrophenyl)sulfonyl]amino]-1-(phenylmethyl)propyl]-, 1,1-dimethylethyl ester.

NOTE At the right-top of the figure a section of the chromatogram is presented obtained by an HPLC-Chiral column experiment, showing that the properties of the diastereomers differ from the enatiomers so that both groups of stereoisomers could be separated.

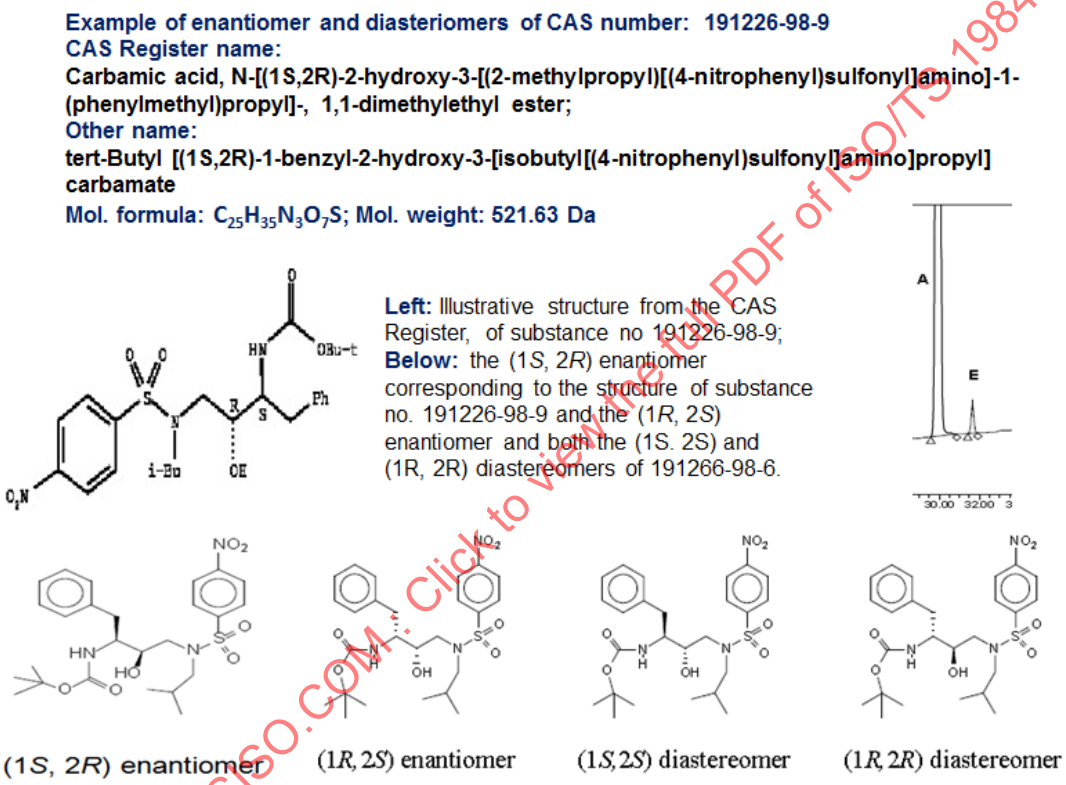


Figure B.16 — Structural representation of substances with two asymmetric centres

B.5.9 Anomers

A special case of epimerism is that of anomerism of sugars. Anomers are defined as diastereoisomers of glycosides, hemiacetals or related cyclic forms of sugars, or related molecules differing in configuration only at C-1 of an aldose, C-2 of a 2-ketose, etc.

In ISO/TS 19844, 4.8 the structural representation of glucose in the crystalline state is provided as well as the introduction of an anomeric C_1-atom after cyclisation of D-glucose and representation of D-glucose in the Haworth projection, as in Figure B.17. In Figure B.18 the structure of the disaccharide lactose is shown. The O-Glycosidic bond is a type of covalent bond that joins a carbohydrate (sugar) molecule to another group, which may be another carbohydrate. In the Haworth-projection C^*_1-OH group is *cis* to the C_5-CH2OH group resulting in the β-anomeric form.

In Figure B.19 the examples of anhydrous α-lactose (β-D-Galactose, α-D-Glucose connected by a β-glycosidic bond) and anhydrous β-lactose (β-D-Galactose, β-D-Glucose) will be presented in accordance with the ISO/TS 19844 Implementation Guide, 4.9 Example α-, and β-Glucose anhydrous. The examples provided in Figure B.19

are for the comparison of the Haworth presentation of structures to the "correct" presentations of the same substances. The structural representation as presented in Figure B.19 should be provided. Note that none of the electronic structural representation understands Haworth projections.

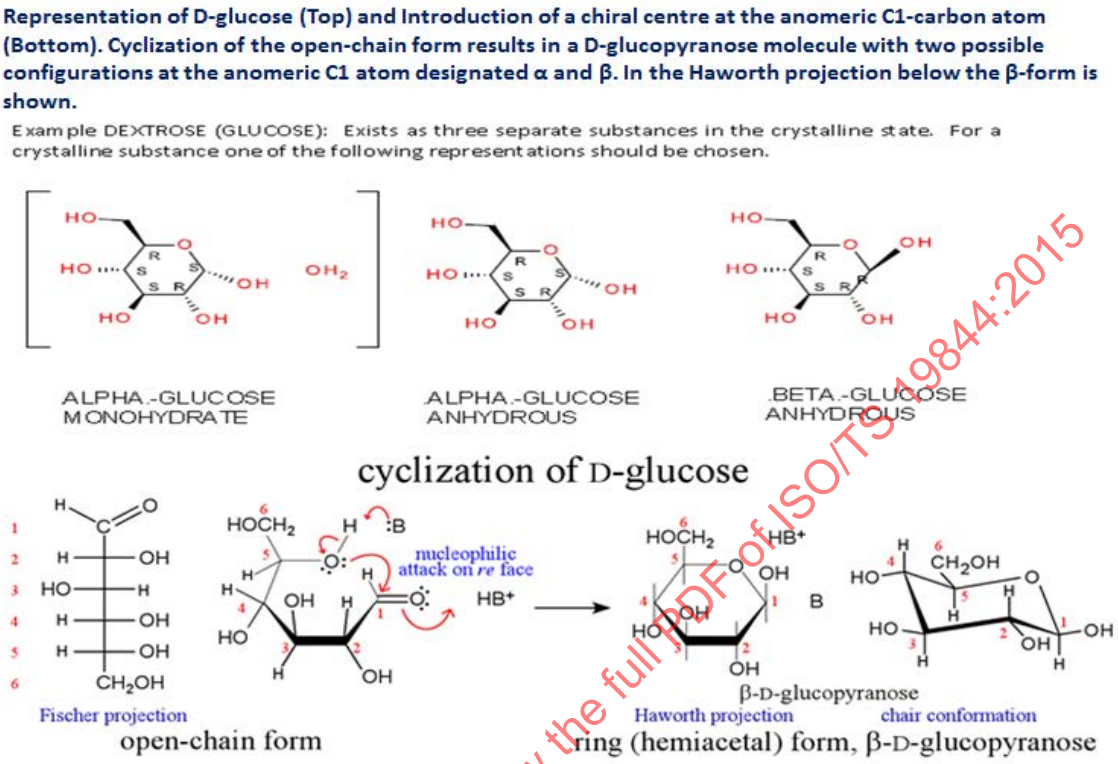


Figure B.17 — Structural representation of α- and β Glucose and α-Glucose monohydrate and in Haworth projection

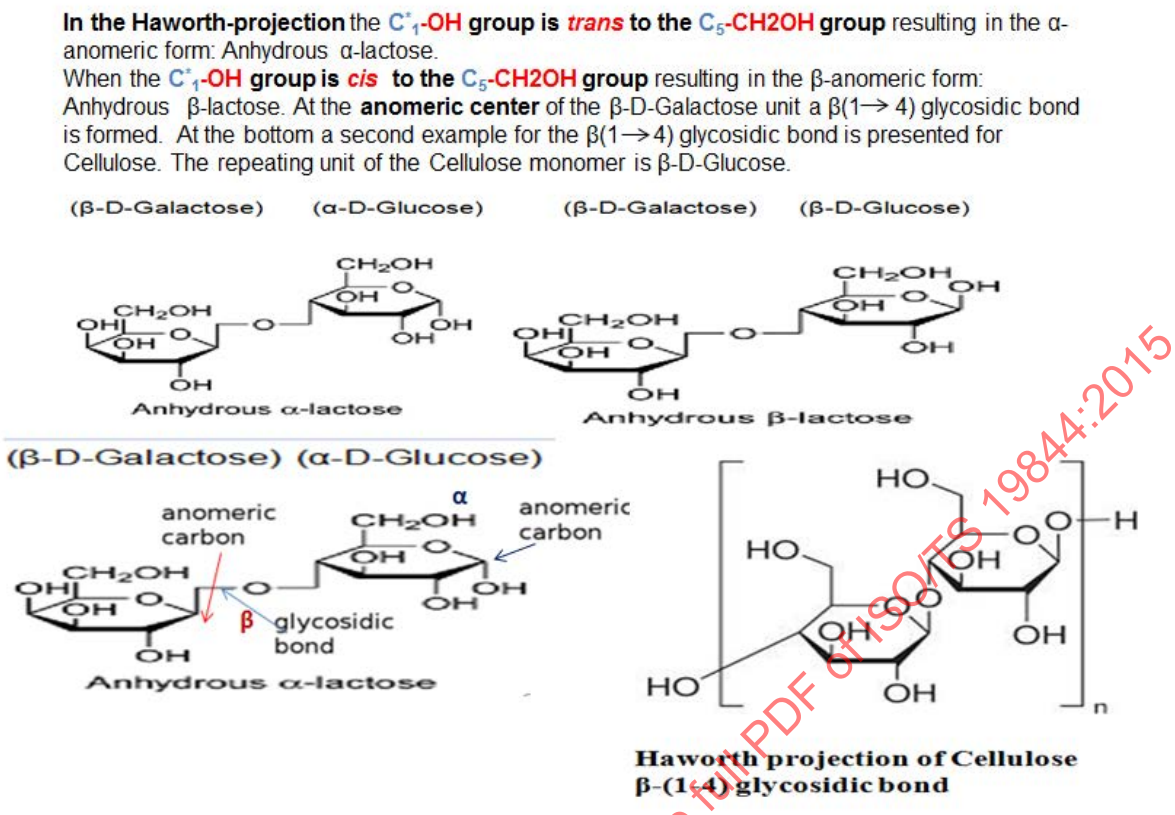


Figure B.18 — Haworth projection of α- and β Lactose anhydrous

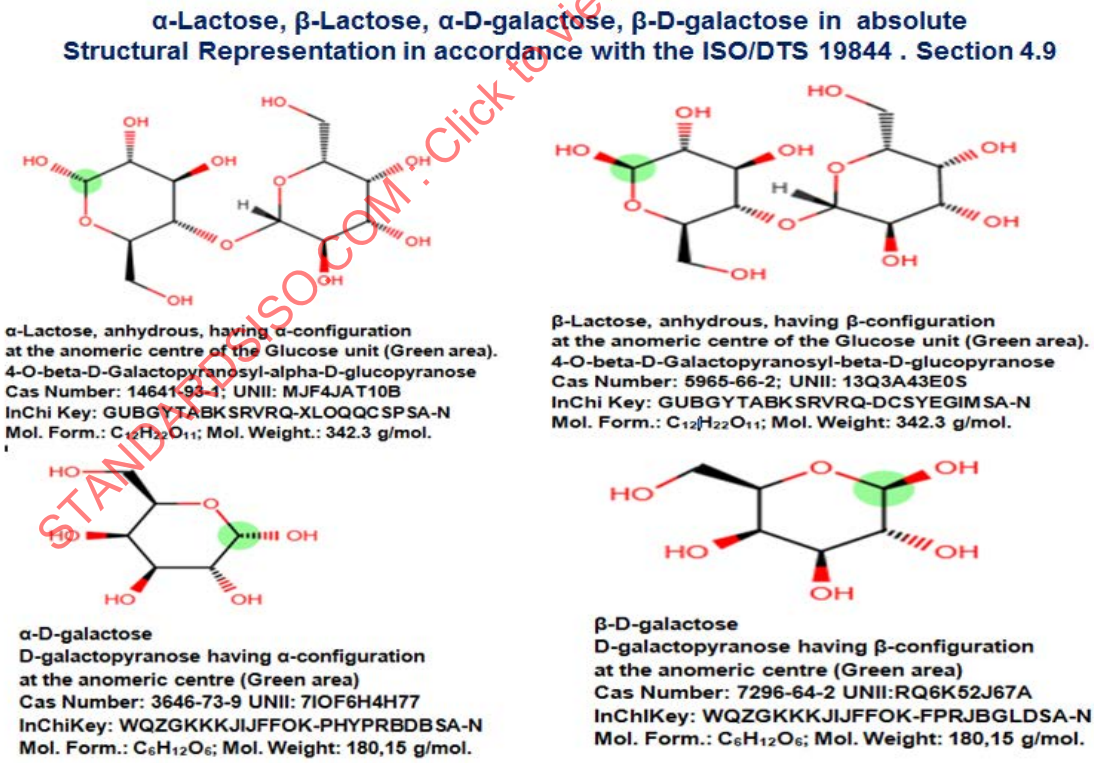


Figure B.19 — α- and β-Lactose, α-D- and β-D-Galactose in accordance with ISO/TS 19844

B.5.10Substance with more than two centres of Asymmetry (Mixture of stereoisomers)

In ISO/TS 19844, 4.8.4 the user guidance reads that mixture of stereoisomers shall be represented explicitly as a mixture of substances with absolute stereochemistry. In the case the absolute stereochemistry is unknown the substance definition should be marked as "Incomplete".

Next **example of Misoprostol** will provide the information needed to comply as a "complete" representation.

The EP, current edition, defines Misoprostol as:

Chemical name: Mixture of methyl 7-[(1RS, 2RS, 3RS)-3-hydroxy-2-[(1E,**4RS**)-4-hydroxy-4-methoxloct-1-enyl]-5-oxocyclopentyl] heptanoate **and** methyl 7-[1RS, 2RS, 3RS)-3-hydroxy-2-[(1E, **4SR**)-4-hydroxy-4-methyloct-1-enyl]-5-oxocyclopentyl] heptanoate.
The FDA-SRS refers to: (+/-)-METHYL (1R, 2R, 3R)-3-HYDROXY-2-((E)-(4RS)-4-HYDROXY-4-METHYL-1-OCTENYL)-5 OXOCYCLOPENTANEHEPTANOATE.
Other name: PROST-13-EN-1-OIC ACID, 11, 16-DIHYDROXY-16-METHYL-9-OXO-, METHYLESTER, (11.ALPHA. 13E)-(+/-)-;
IUPAC Name: methyl 7-[(1R, 2R, 3R)-3-hydroxy-2-[(E)-4-hydroxy-4-methyloct-1-enyl]-5-oxocyclopentyl] heptanoate+
Molecular formula: $C_{22}H_{38}O_5$;
Molecular weight: 382.5**;**
Unit: g/mol
CAS Registry number: 59122-46-2;
Code: UNII: 0E43V0BB57
Chirality: Misoprostol has four chiral centres. The presence of 2^4 = 16 stereoisomers is thus possible. Four of the 16 enantiomers compose Misoprostol [methyl (13E)-(±)-11, 16-dihydroxy-16-methyl-9-oxoprost- 13-en-1-oate]

The centre at **C_{16} is epimeric** and the centres at C_8, C_{11} and C_{12} are controlled relative to each other with cis substitution on the cyclopentanone ring. Misoprostol is therefore a mixture of four stereoisomers, Figure B.20.

NOTE When possible, the structure should contain the numbering of the Carbon atoms in case of complex structures. (e.g. difference in the conventional prostaglandin numbering of the carbons and the numbering used by IUPAC and EP, starting with the oxocyclopentyl ring (1 to 5), and counting the side chains separately.

INN: Misoprostol; Mol. form.: $C_{22}H_{38}O_5$; Mol. weight: 382.5 g/mol.; CAS: 59122-46-2
USAN: (±)-(11α, 13E)-11 ,16-dihydroxy-16-methyl-9-oxoprost-13-en-l-oic acid, methyl ester
EP: Mixture of methyl 7-[(1RS, 2RS, 3RS)-3-hydroxy-2-[(1E,4RS)-4-hydroxy-4-methoxloct-1-enyl]-5-oxocyclopentyl] heptanoate and methyl 7-[1RS, 2RS, 3RS)-3-hydroxy-2-[(1E, 4SR)-4-hydroxy-4-methyloct-1-enyl]-5-oxocyclopentyl] heptanoate

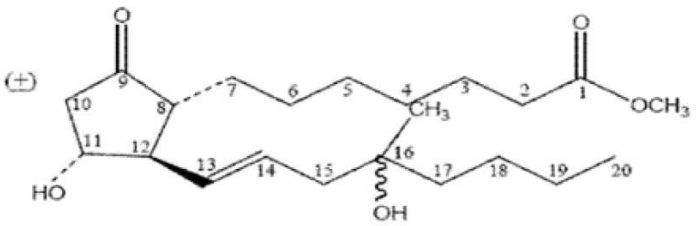
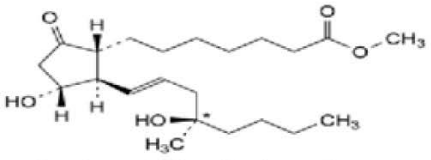
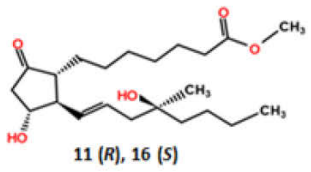


Fig. A, conventional prostoglandin numbering system

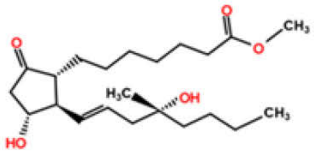


its epimer at C* and their enantiomers

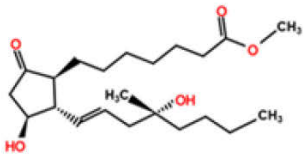
EP. Representation of the epimeric mixture at C* (4RS)(4SR)



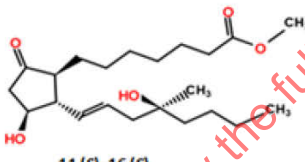
11 (*R*), 16 (*S*)



11 (*R*), 16 (*R*)



11 (*S*), 16 (*R*)



11 (*S*), 16 (*S*)

4 stereisomers:
Methyl 11(**R**), 16(**S**), 13 (*E*) - 11,16-dihydroxy-16-methyl-9-oxoporst-13-en-1-oate **(active)**
Methyl 11(**R**), 16(**R**), 13 (*E*) - 11,16-dihydroxy-16-methyl-9-oxoporst-13-en-1-oate
Methyl 11(**S**), 16(**R**), 13 (*E*) - 11,16-dihydroxy-16-methyl-9-oxoporst-13-en-1-oate
Methyl 11(**S**), 16(**S**), 13 (*E*) - 11,16-dihydroxy-16-methyl-9-oxoporst-13-en-1-oate

Figure B.20 — Structural representation of Misoprostol

B.5.11 Conclusion for the Key elements

In order to assign a Substance ID for a chemical substance, a complete covalent structure with all stereochemistry (*R* or *S* and *E* or *Z* and *sn*) defined is needed. The stoichiometry (mole ratio) of counterions or solvates present in the material should also be provided.

For all chemical substances a molecular representation (Mol file, SMILES, InChI, ChemDraw) of the substance should be provided with all stereochemistry assigned or sites of unknown stereochemistry identified. An actual image of the structure can be provided in an accompanied document (4.8). In addition to the structure the molecular formula, molecular weight, and systematic name should be provided at the structural representation.

B.5.12 Decision tree for a new Substance ID

Figure B.21 provides a decision tree for the conditions of elements needed which might help in the choice whether a new Identifier is needed or not.

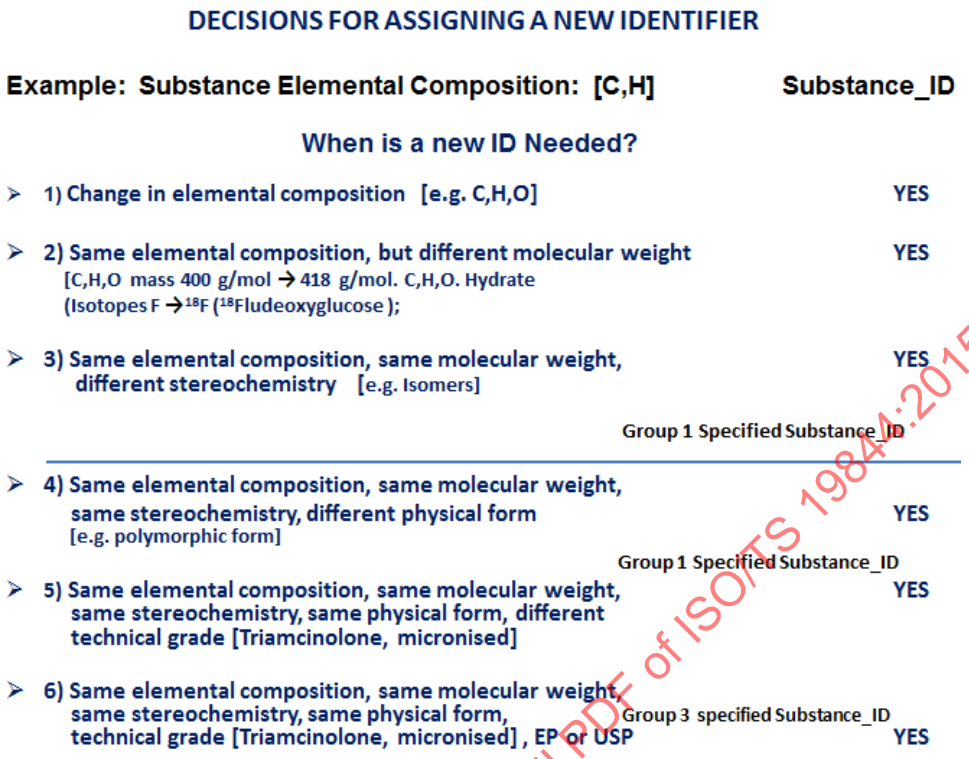


Figure B.21 — Decision tree for a new Substance ID

B.6Discussing other elements of importance regarding the characteristics of a substance

B.6.1 Introduction

At the ISO 11238 specified substance Group 3 Information level **vegetable oils** are captured as described in several Pharmacopoeial monographs e.g. Olive Oil, Virgin EP and Olive Oil, Refined EP Definition Olive Oil, Virgin: Fatty oil obtained by cold expression or other suitable mechanical means from the ripe drupes of *Olea europaea* L.

These types of substances are discussed in detail in **ANNEX E Structurally Diverse Substance, herbals.**

B.6.2 Naming Vegetable Oils

The **business rules** for naming of vegetable oils are described in ISO/TS 19844 Section Structurally Diverse Substance, herbals, 5.5.1 and in the section Source material (4.10).

NOTE The information model for Structurally Diverse substances has to be updated in the Main body document of the ISO/TS 19844 as by the latest version and in the ISO 11238 standard.

Since in the next version of the Implementation Guide a proposal will be made to include both the information for the herbal drug (a term used here in preference to herbal substance) and the herbal preparation at the substance information level. Additional herbal drug and herbal preparation information can be included in the Substance_type Structurally Diverse specified substance Group 1 information level. This is needed to capture the marker and signature substances at this level in order to comply with the Regulatory regulation.

Within this Annex, the terms Substance and specified substance are used as defined by ISO 11238. The terms herbal drug (herbal substance) and herbal preparation are used as defined by pharmacopoeias, and will both be captured at the Substance level. In addition, a more basic unit, comprising unprocessed, fresh material of a single species or infraspecies (occasionally a genus) + part, with no affiliation to any particular pharmacopoeia, will also be captured at the Substance level; for clarity, these will be referred to as Substance (fresh) in this Annex, The information model also accommodates fraction information at the substance level. The **element <Fraction>** will

be considered as **two herbal preparation_Types: <**Oil, Exudate, Juice> or <Extract>, each requiring different information.

Specified substance group 1 includes additional information for substances and multi-substance material. Additional information for herbal drugs and for herbal preparations *e.g*. additional information for an extract such as extraction solvent composition, herbal Drug/ herbal extract ratio will be captured at this level along with the constituents and marker or signature substances.

Specified substance group 2 will capture manufacturer information along with limited manufacturing information.

Specified substance group 3 will capture the grade of a given substance or specified substance group 1 as outlined in a Pharmacopoeia or imposed by a given regulatory agency.

The high-level information as is laid down in the ISO 112338 Substance standard is unchanged as is presented in Figure B.22.

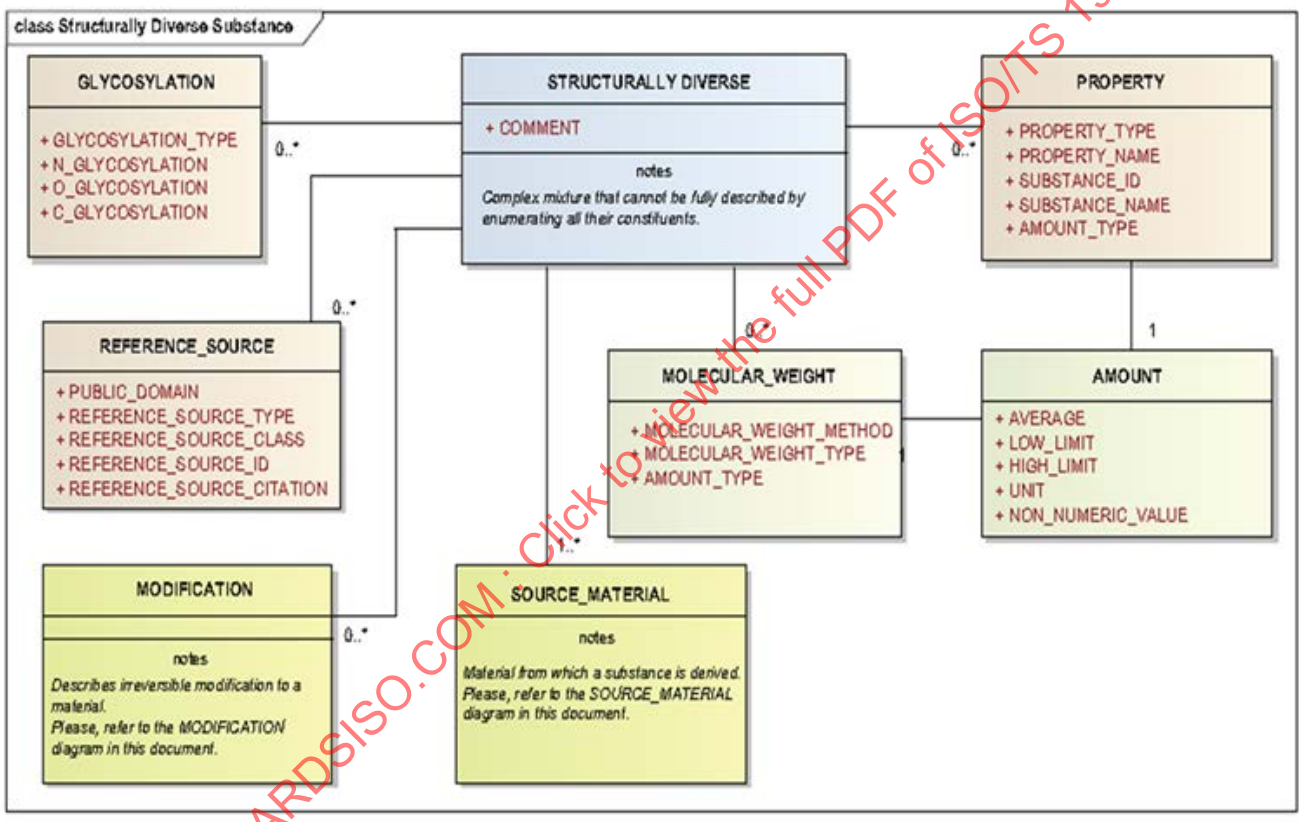


Figure B.22 — Information model for the unique identification of Substance Type: Structurally Diverse

In Figure B.23 the Substances subclasses are shown belonging to the Substance Type: Structurally Diverse.

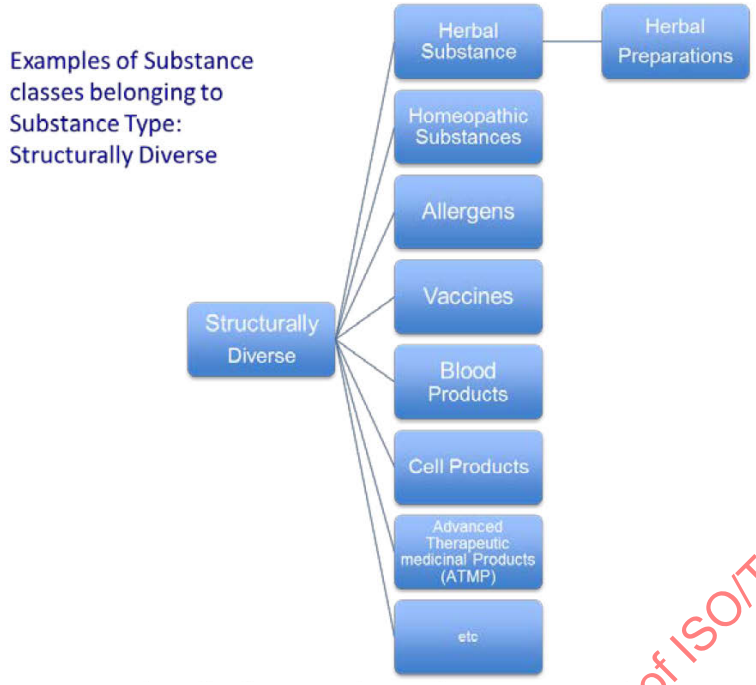


Figure B.23 — Examples of Substance classes belonging to Substance Type: Structurally Diverse

B.6.2.1 Example Olive Oil, Virgin

Olive Oil, Virgin EP is defined as a herbal preparation (because it has been obtained by expression). There is no defined herbal drug for this herbal preparation, but instead it's source material is the ripe drupes (fruit) of *Olea europaea* L.

For the purposes of the ISO 11238, the herbal preparation is defined as a Substance as [<Scientific genus/binomial/trinomial **without Author**>,<Part(s)>,<Fraction>], i.e. *Olea europaea*, Fruit, Oil.; Common name: Olive Oil, Virgin. The source material for the herbal preparation, with the **relationship** of **Parent substance,** giving the Substance Role the value <Parent>, is a separate Substance (fresh) defined as [<Scientific genus/binomial/trinomial **with Author**>,<Part(s)>], i.e. *Olea europaea* L., Fruit.

The **name** [<Scientific genus/binomial/trinomial **with Author**>,<Part(s)>] refers to a **Substance Name** in terms of the ISO 11238 that is not an official name in any pharmacopoeia or other legislation. The inclusion of the author is to ensure that it is not confused with the official names for herbal drugs and herbal preparations which are also defined as Substances by the ISO 11238. The scientific element of the name is usually a binomial (species), but may be a trinomial (infraspecies such as a variety or subspecies) or a genus, as appropriate, together with the author of the name. The Part element of the name should be chosen from a CV consisting of a restricted number of terms.

Taxonomic information, particularly the scientific name of a medicinal plant, is essential for defining herbal substances in medicinal products. The Kew Gardens Medicinal Plant Names Services (MPNS) should be the authoritative source for all scientific names and should also maintain a list of plant parts and fractions. A controlled vocabulary for medicinal plant taxonomy has been developed and will be maintained by the Kew Gardens Medicinal Plant Names Services. A submitter should choose the current accepted scientific name provided by MPNS.

The **name** [<Scientific genus/binomial/trinomial **without Author**>,<Part(s)>] refers to a **Substance Name** in terms of the ISO 11238 and is the equivalent to a herbal drug Name. There is no such Substance Name for Olive Oil in this example from the EP, although other pharmacopoeias might define a herbal drug for *Olea europaea*, Fruit.

The **name** [<Scientific genus/binomial/trinomial **without Author**>,<Part(s)>], [Fraction] refers to a **Substance Name** in terms of the ISO 11238 and is **equivalent** to an **herbal preparation Name.** The herbal preparation Name has two herbal preparation Name_Types; one for the Substance fraction: Oil, Exudate, Juice and one for the

Substance fraction: Extract. This is done to differentiate between the cardinalities of the herbal preparation_Types.

The herbal preparation-Substance (fresh) is a two way relationship. The herbal preparation_Type <Oil, Exudate, Juice> is prepared from **one** Substance (fresh).

Note that several herbal preparations, *e.g.* Juices, can be a mixture of separately obtained juices obtained from the herbal substances and mixed after expression. The final preparation is then a multi-substance material which is captured at the Specified substance Group 1 level.

The herbal preparation_Type <Extract> can be prepared from **one or more** herbal drugs and by **one or more** extraction steps and by **one or different** extraction solvent compositions. Only the first extraction solvent composition is used in the element solvent. Further details of the herbal preparation and herbal drug are described in the ANNEX, Structurally Diverse substances, herbals.

For the herbal preparation Olive Oil, Virgin the result of the information model will read:
Substance_ID: HJFTE78543 (Artificial ID); (UNII: 6UYK2W1W1E)
Substance Name (herbal preparation name): *Olea europaea*, Fruit, Oil
Substance Name_Type: Other
herbal preparation_Type: Oil [1)]
Substance Name: Olive Oil, Virgin
Substance Name_Type: Official (for further elements see tabular format)
Parent Substance_ID: OLIJF5643S
Parent Substance_Name: *Olea europaea* L., Fruit
NOTE 1) The value Oil is appropriate here. The herbal preparation_Type is valid for: Oil, Exudate, Juice with the cardinality (1,1).

The relationships between the ISO 11238, ISO/TS 19844 Structurally diverse description and the Substance (fresh) and Substance (herbal preparation) for Olive Oil, Virgin is represented in the Figure B.24.

For comparison, the relationships between the ISO 11238, ISO/TS 19844 Structurally diverse description, and the Substance (fresh), Substance (herbal drug) and Substance (herbal preparation) for Ginkgo biloba, Dry Extract is represented in the Figure B.25.

Further details and discussion is described in the Structurally Diverse substances, herbal ANNEX.

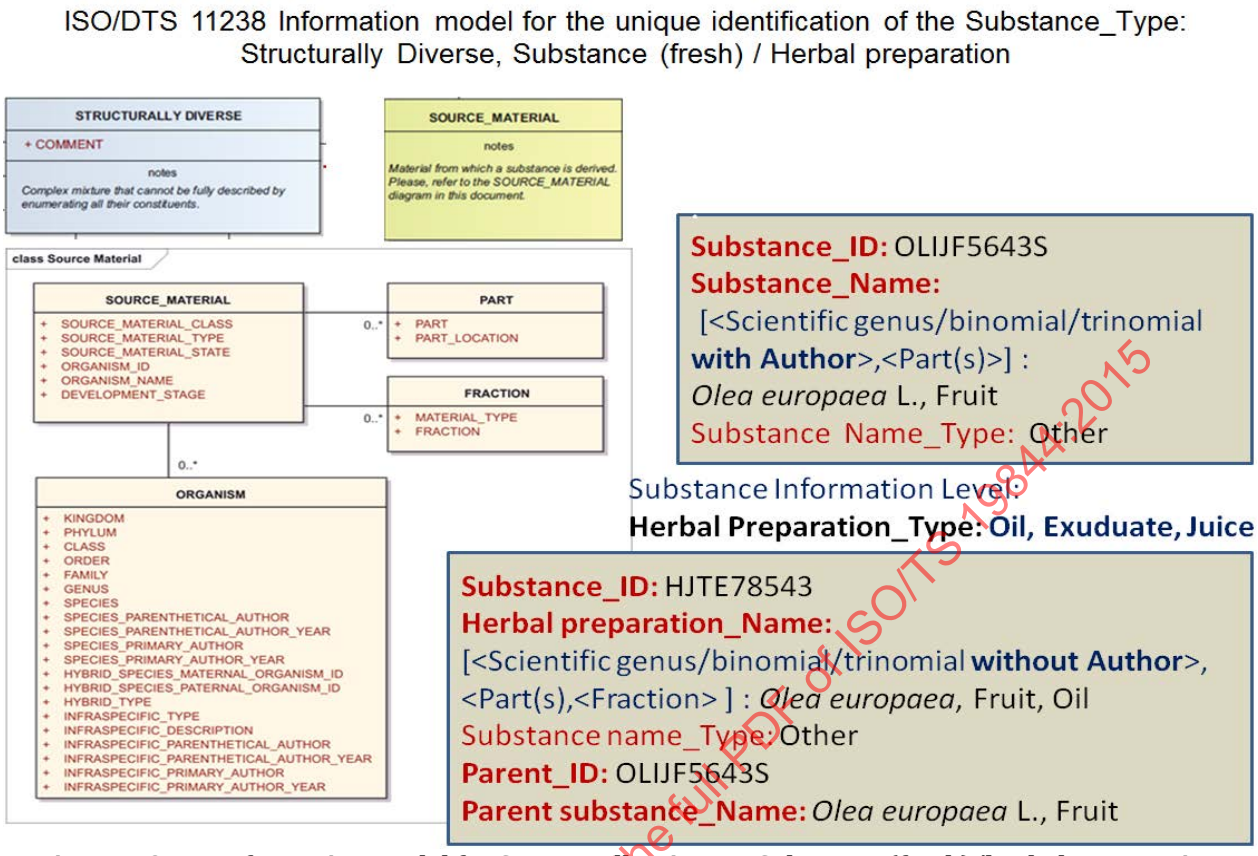


Figure B.24 — Information model for Structurally Diverse, Substance (fresh)/herbal preparation

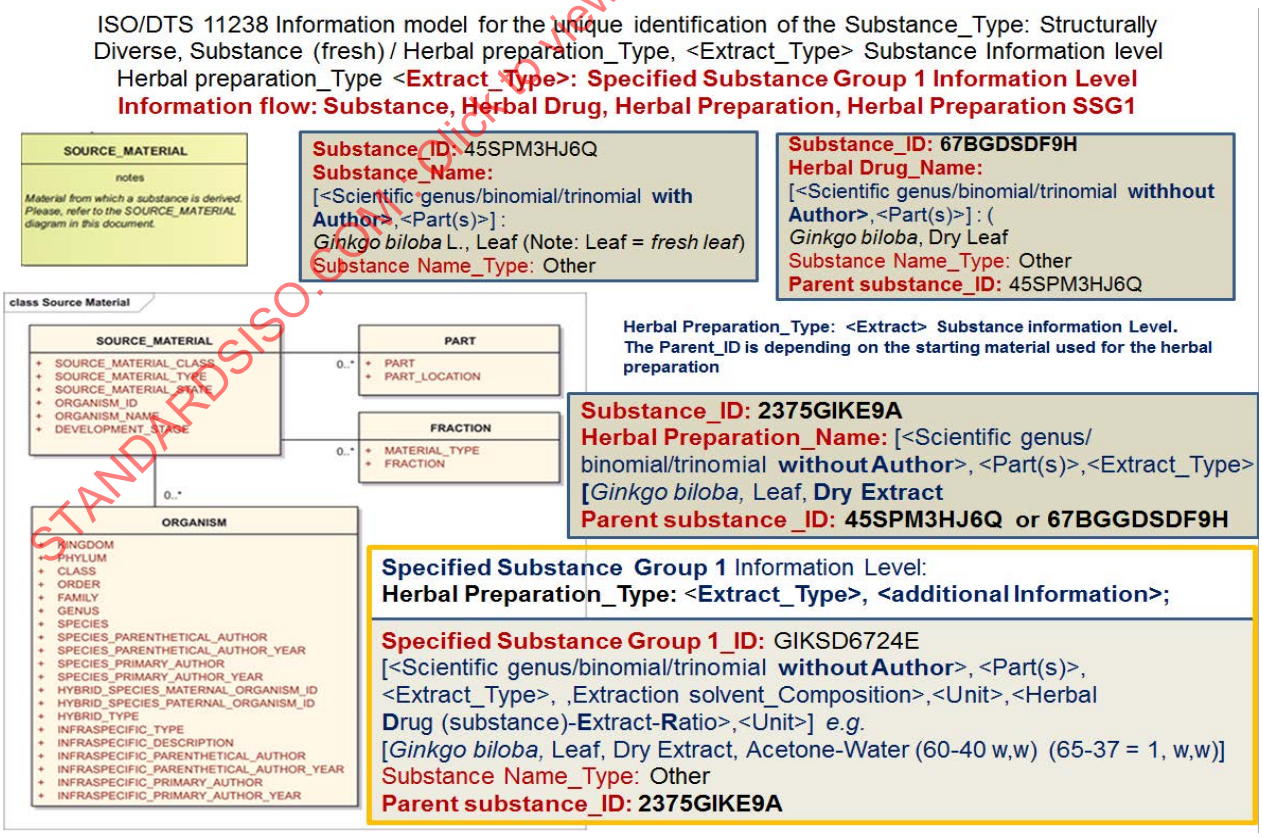


Figure B.25 — Information model for Structurally Diverse, Substance (fresh)/herbal preparation and herbal preparation at the specified substance Group 1 level

NOTE Note that at the Substance level the *Ginkgo biloba* L., Leaf is per definition always the fresh material. Consequently, *Ginkgo biloba*, dry leaf is a herbal drug and is considered to be captured at the Substance level and will therefore have a distinct Substance ID from the fresh leaf.

Captured at the Substance Level:

Ginkgo biloba L.: = Substance (fresh material)
Ginkgo biloba, Leaf: = herbal drug (dry material as defined by the pharmacopoeia)
Ginkgo biloba, Leaf, Dry Extract: = herbal preparation (as defined by the pharmacopoeia)
Captured at the specified substance Group 1 Level:
Ginkgo biloba, Leaf, Dry Extract, Acetone-Water (60-40 w,w) (65-37= 1, w,w) (as defined in the application dossier for registration).

Additional information for more elements at the Substance level for the Substance *Olea europaea* L., Fruit and the herbal preparation Olive Oil, Virgin is provided in Figure B.26.

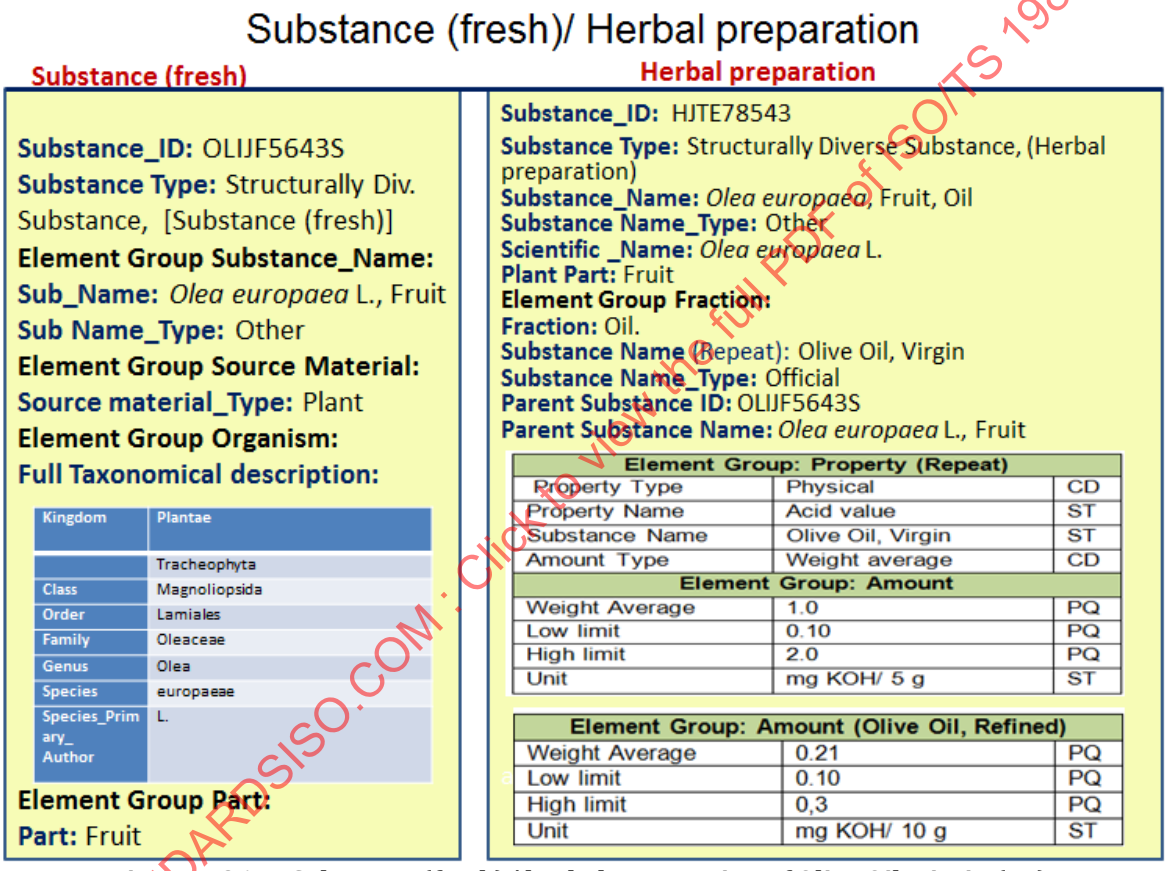


Figure B.26 — Substance (fresh)/ herbal preparation of Olive Oil, Virgin (EP)

B.6.2.2 Elements to be described at the specified substance Group 1 Information level

The element group Properties is envisioned to be allocated at the Substance or at the specified substance Group 1 information level for the description of the characteristics of the Substance (fresh), herbal drug and/or herbal preparation, Figure B.26.

Properties to be captured, related to liquids (oils) and the description can be found in B.2 of this ANNEX.

NOTE The specified substance Group 1 does not have a connection with the modification group. So it will be proposed to introduce the modification group at the specified substance Group 1 level in addition to its existing connection at the Substance level (Proposed Change to ISO 11238, in Figure B.27). Also the element group Property is updated with the Property parameter attribute.

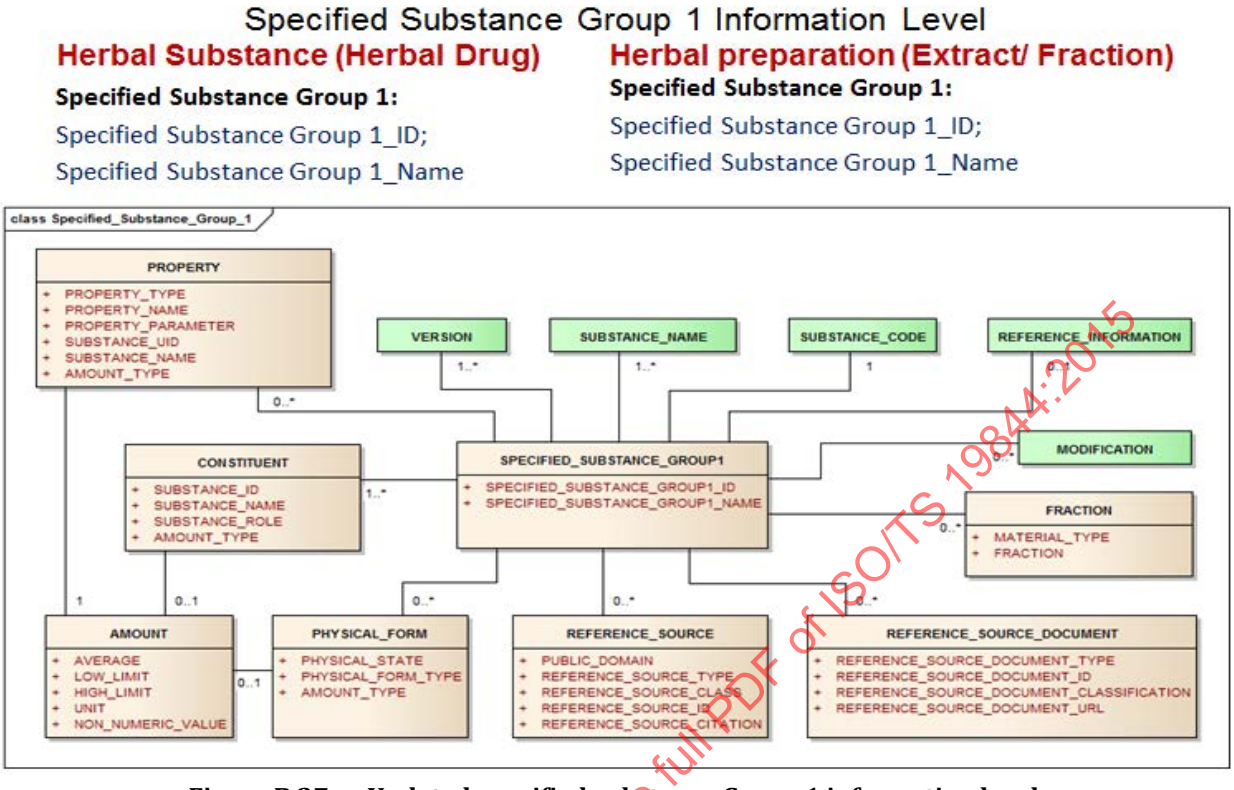


Figure B.27 — Updated specified substance Group 1 information level

NOTE 1 In Figure B.27 it is necessary to describe the herbal preparation "Extract" in more details compared to the herbal preparation "Oil" in order to comply to European legislation, since marker substances must be described and referred to for standardised and quantified herbal preparations. Since the constituent group is captured at the specified substance Group 1 information level this information is captured here and as a consequence a herbal preparation_Type <Extract> has a specified substance Group 1_ ID and Name. Further specifications described at the SSG1 level.

NOTE 2 Further information and examples are provided in ANNEX, Structurally Diverse, herbals.

B.6.2.3 Elements to be described at the specified substance Group 2 Information level

At specified substance Group 2 Information level Manufacturing information will be captured. For the **herbal drug** this information will tie the manufacturer which collects, processes (comminuted, cutting of the material) and stores the herbal drug.

For the **herbal preparation** the manufacturer of the processed herbal drug is captured. In this Substance Class the Element group Reference Source Document is included. The Reference Source Document Type provides the manufacturer to submit the Release specification of the manufactured preparation.

Only the Element Group Manufacturing is meant to be implemented yet up to the minimum information captured for the attributes in this element group, see Figure B.28. However in future versions of the Implementation Guide this element group can be extended up to a Manufacturing Class with the information model as is shown in Figure B.29. The element group Manufacturing information as a class is yet allocated at the specified substance Group 4 information level in the ISO 11238. This (extended) manufacturing information is shown only for informational purposes and is meant to capture the starting materials, impurities and the intended manufactured substance. The information model provides information up to the element group Manufacturing.

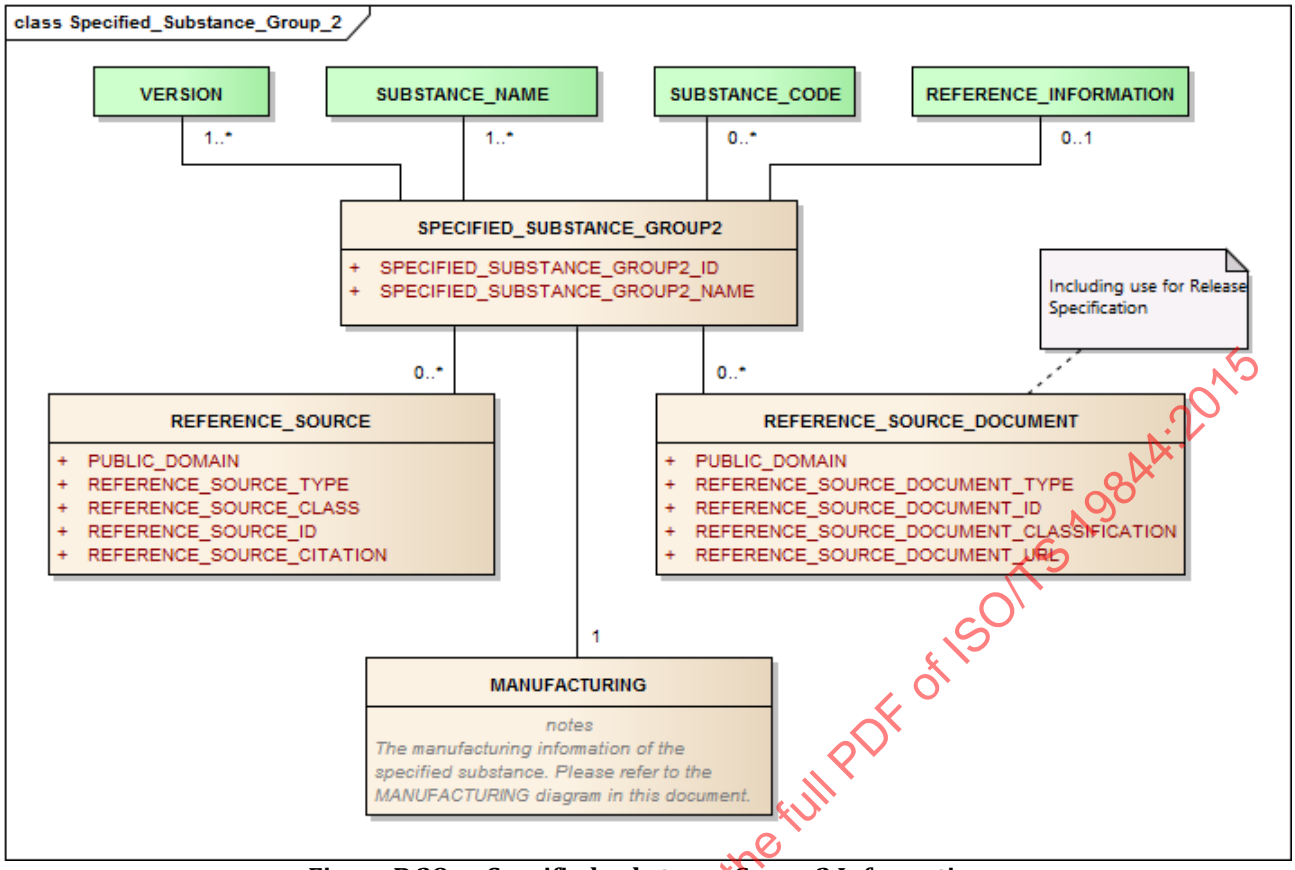


Figure B.28 — Specified substance Group 2 Information

Figure B.29 provides the information model for the element group Manufacturing, with extended information. Up to now the model captures only the Manufacturing element group in connection with Figure B.28.

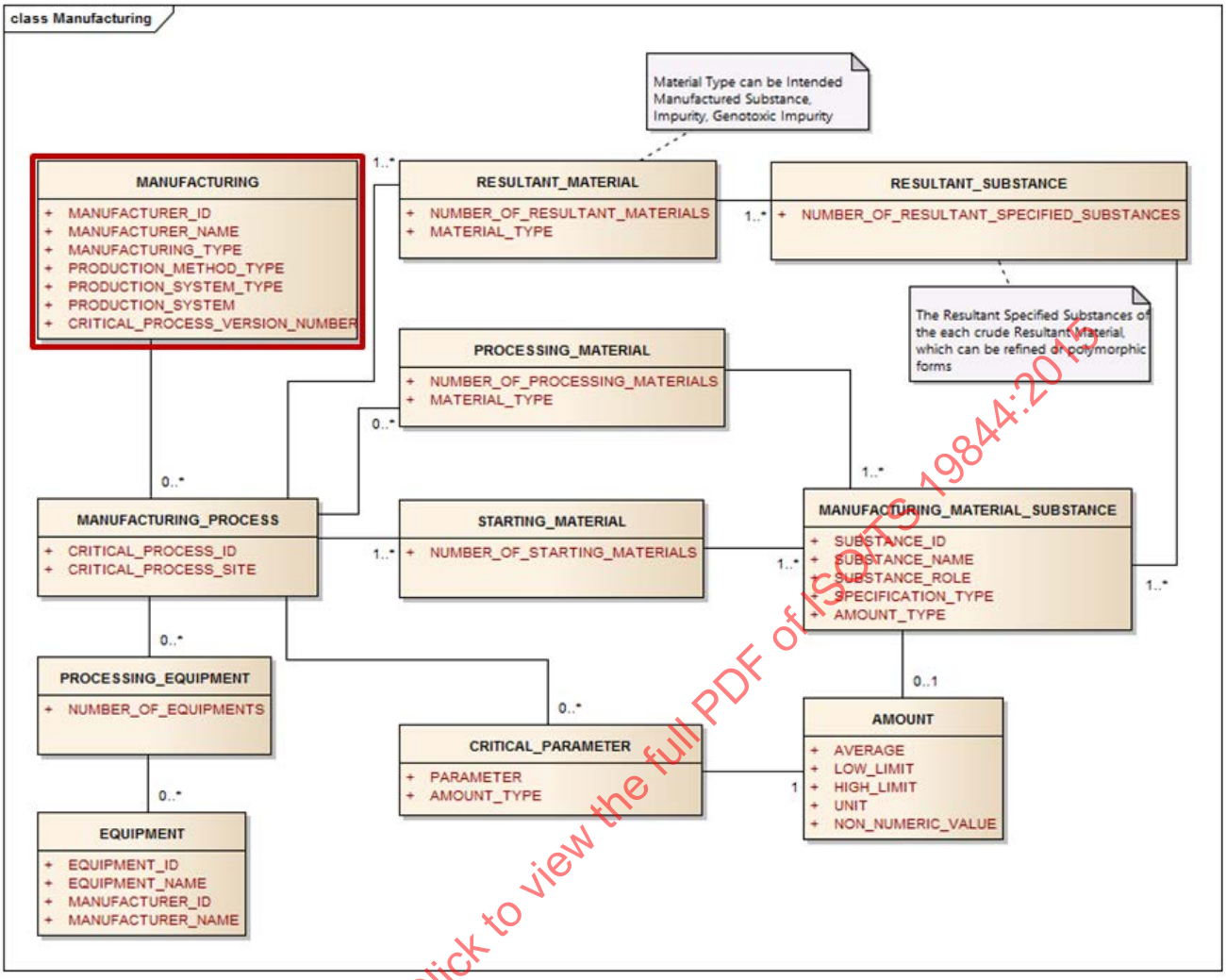


Figure B.29 — Specified substance Group 2 Information, element group Manufacturing (Red Box)

Figure B.29 specified substance Group 2 Information level, element group Manufacturing is captured in the red box.

B.6.2.4 Elements to be described at the specified substance Group 3 Information level

At specified substance Group 3 information level the Pharmacopoeial Grade of a (specified) substance is captured as a specification as for Olive Oil, Virgin with the Specified substance Group 3 name of Olive Oil, Virgin EP, in Figure B.30.

Specified substance group 3_ID: JFGDRW4526 (Artificial ID)
Specified substance group 3 name: Olive Oil, Virgin EP
Reference Source Type: EP
Parent substance ID: HJFTE78543 corresponding to the:
Parent substance: *Olea europaea*, Fruit, Oil (= herbal preparation) which is related in turn to the
Parent Substance: *Olea europaea* L., Fruit (= Substance (fresh)) and corresponding Parent substance ID (HJFTE78543).

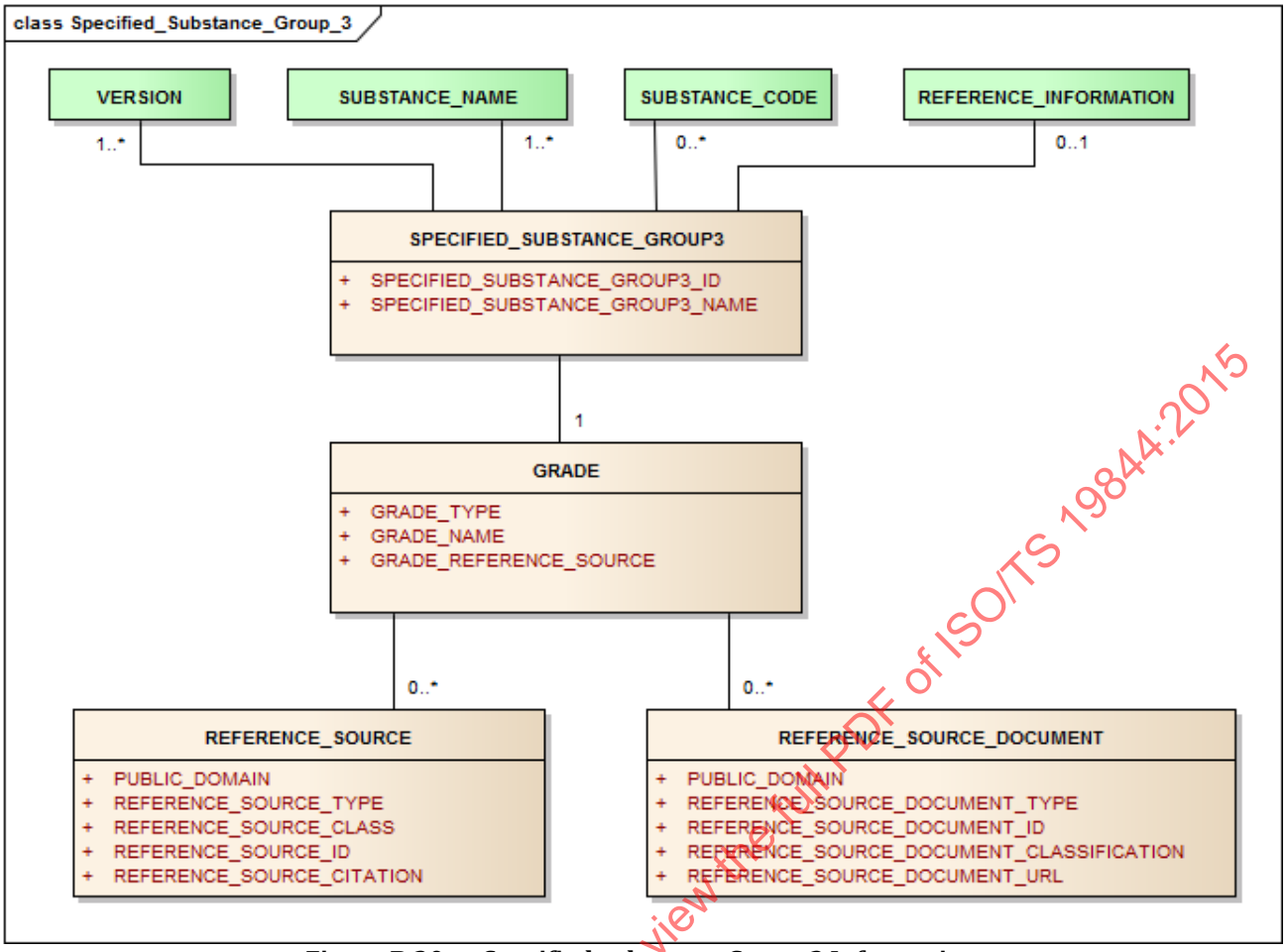


Figure B.30 — Specified substance Group 3 Information

In Figure B.30 the information is provided to capture the grade of a substance. This group is meant to be for Pharmacopoeial Grades (EP, USP) Each grade will have a separate specified substance Group 3 Identifier. However there are applications in which regulatory approval has been achieved for more than one grade and sometimes the sponsor will add an extra specification, such as particle size which is not captured in the Pharmacopoeial monograph. In this case the model provides in the Reference documentation type "In House" Grade in which a specification set can be laid down covering the EP, USP as well as the intended particle size.

B.6.2.5 Olive Oil, Refined

In the EP monograph Olive Oil, Refined the definition reads: Fatty oil obtained by refining of crude olive oil, obtained by cold expression or other suitable mechanical means from the ripe drupes of *Olea europaea* L.

This means that at the substance level the source of Olive Oil, Refined (EP) is captured by the Substance (fresh) *Olea europaea* L., Fruit, which is thus the same as for Olive Oil, Virgin (EP). For the Substance herbal preparation the fraction (Oil) will be captured as:

Substance ID: KHJGHR4523 (Artificial ID)
Substance name: *Olea europaea*, Fruit, **Oil Refined**
Substance Name Type: Other
Substance Name: Refined Olive Oil
Substance Name Type: Official
Parent substance ID: HJFTE78543 corresponding to: the
Parent substance *Olea europaea*, Fruit, Oil (= herbal preparation); Common name: Olive Oil, Virgin, and which in turn is related to the parent organism>
Parent Substance_Name:*Olea europaea* L., Fruit and

Parent substance_ID: OLIJF5643S (= Substance (fresh)).

NOTE The difference between the Virgin Olive Oil (EP) and the Refined Olive Oil (EP) is a reduction of the amount of free acids which will be captured as a modification. The result will be the difference in Grade, captured at specified substance Group 3 and probably a difference in manufacturing method captured in specified substance Group 2. There isn't any structural chemical modification process, such as (substitution, breaking covalent bonds, hydrogenation, etc.), but only a quantitative change in amount of constituents being expressed in the properties. These properties are captured in the element group Properties.

At the specified substance Group 3 level the Pharmacopoeial Grade of the substance is captured as a specification for Olive Oil, Refined (EP).

The specified substance Group 3 name for Olive Oil, Refined (EP), as in B.5.2.7.2, will be:
Specified substance Group 3 ID: SDAUT12543 (Artificial ID)
Specified substance group 3 name: Olive Oil, Refined EP
Reference Source Type: EP
Parent substance ID: KHJGHR4523 corresponding to the
Parent substance name: *Olea europaea*, Fruit, Oil Refined and is related to the
Parent substance name: *Olea europaea*, Fruit, Oil; Common name: Olive Oil, Virgin, with the
Parent substance ID: HJFTE78543 (= herbal preparation), which in turn is related by the relationship <parent organism> with the
Parent substance name: *Olea europaea* L., Fruit
Parent substance ID: OLIJF5643S (Substance (fresh)).

B.6.2.6 Properties to be captured, related to liquids (vegetable oils)

In B3 and B4 the key elements for defining the substance are: Structure, Molecular formula, Molecular weight and Solid state characterisation.

For this group of liquid substances, properties such as refractive index, solubility, viscosity, UV-absorption, specific gravity, saponification value, acid value, peroxide value and the composition of the mixture material, such as the mixture and relative quantity of the fatty-oil fraction and sterol-fraction of the oil, are essential in defining the specified substance, whether it is Virgin Olive Oil (EP), or Refined Olive Oil (EP), which is essential for use in parenteral preparations.

In Figures B.31 and B.32 these elements are discussed regarding the definition of a vegetable oil. In Figure B.31 the two EP grades are compared. The elements described in **red** are different for the grades, but important for whether the material can be used in a parenteral preparation.

In fact the two grades are different specified substances in the sense of ISO 11238, since the Refined Olive Oil (EP) is sourced from the Virgin Olive Oil (EP), but has been refined by a modification process in order to remove or reduce free acids.

In both the Virgin Olive Oil (EP) and the Refined Olive Oil (EP) the specifications for the fractions for fatty-acids and sterol-fraction are the same. Indeed not only the qualitative composition is important but also the quantitative composition in order to decide whether the substance is Crude Olive Oil or Refined Olive Oil. The difference between the two grades is managed by testing **quantitatively** the other properties of the constituents of the Olive Oil types by measuring the solubility, UV-absorbance, acid value and peroxide value.

Conclusion: To distinguish between Olive Oil (EP), Virgin and Olive Oil, Refined (EP), a Certificate of Analysis and a batch result should be provided to substantiate the submission for a specified substance Group 3 ID and as a consequence information should be provided up to the substance level (Source material), since quantitative results of the properties make the difference between the two types of Olive Oil.

In Figure B.32 a batch result is shown. In the **red box** the elements and values are captured demonstrating that the batch complies with the **quantitative** values of the properties requirements. The fatty-acid qualitative and quantitative composition as well as for the sterols composition show that the batch results comply with both the

Olive Oil types. The results in the red box are making the difference. They show by **quantitative values** that the sample complies with Olive Oil, Refined (EP).

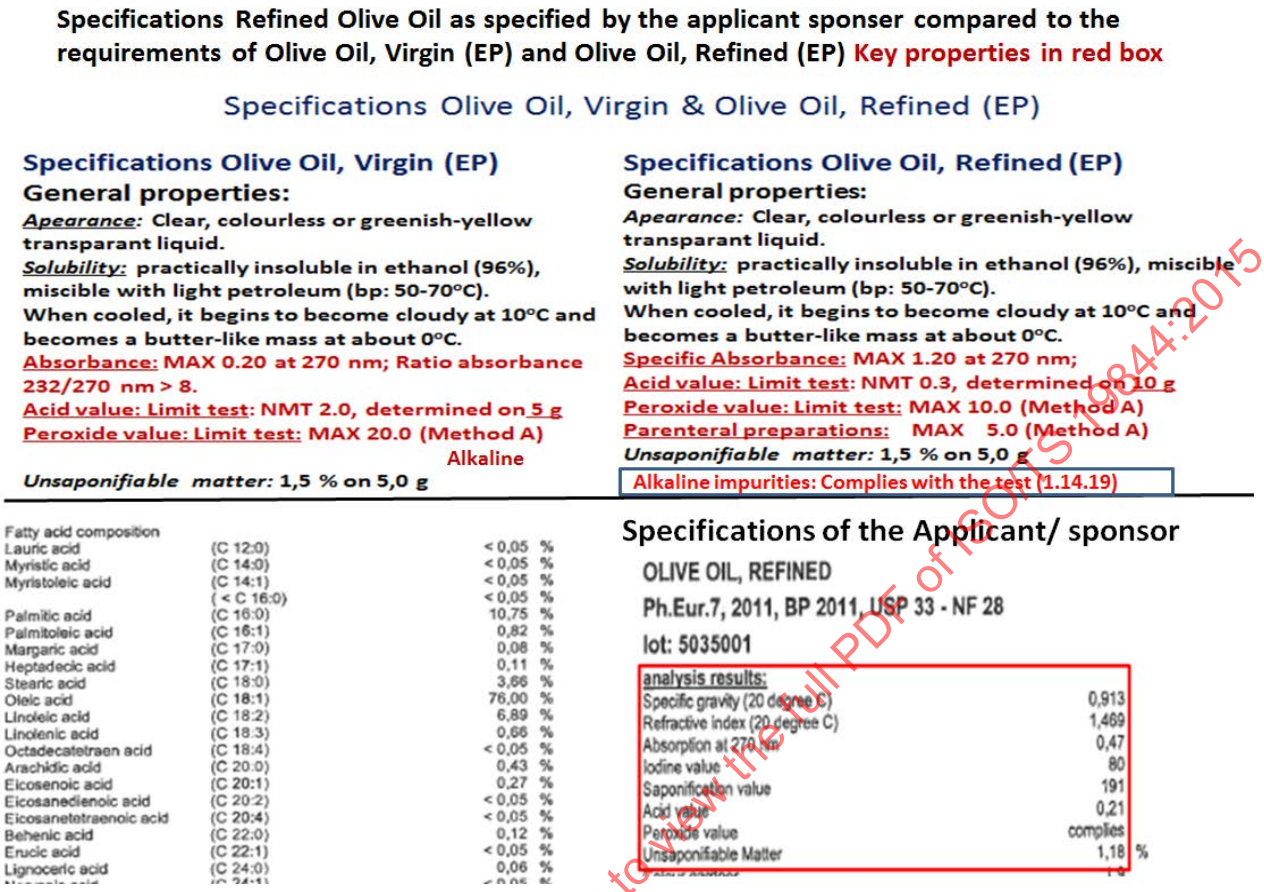
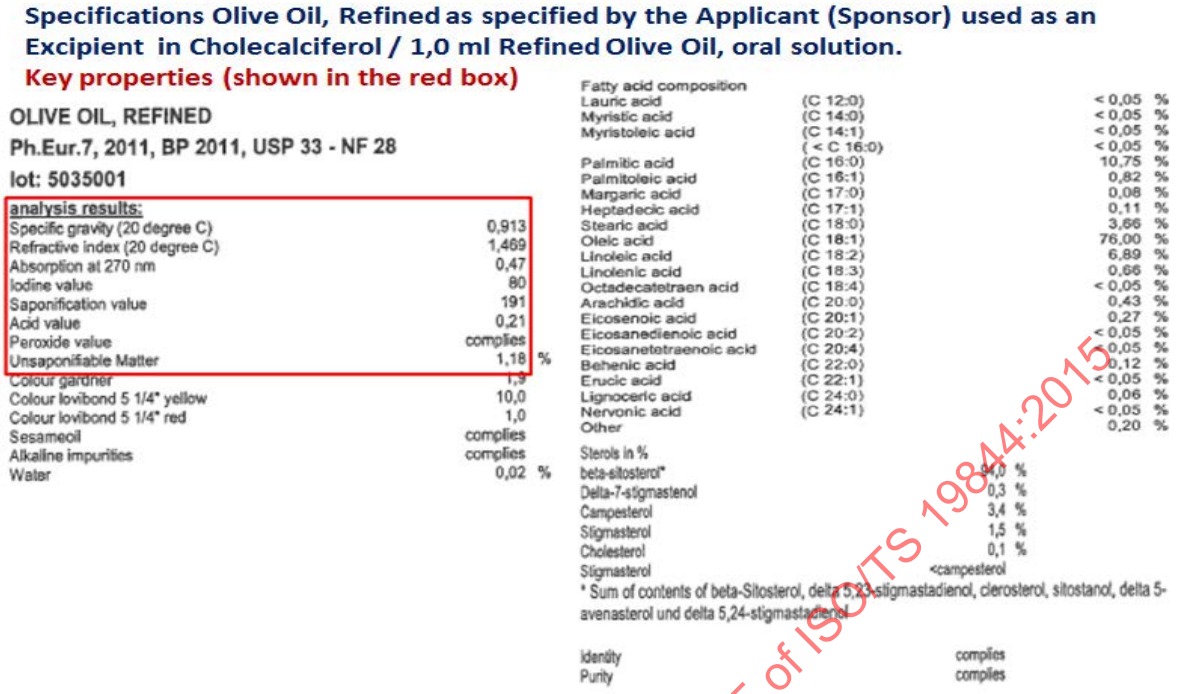


Figure B.31 — Specifications for Olive Oil (EP), Virgin and Olive Oil, Refined (EP)



Specifications Olive Oil, Refined as specified by the Applicant (Sponsor) used as an Excipient in Cholecalciferol / 1,0 ml Refined Olive Oil, oral solution.

Key properties (shown in the red box)

OLIVE OIL, REFINED

Ph.Eur.7, 2011, BP 2011, USP 33 - NF 28

lot: 5035001

analysis results:

Specific gravity (20 degree C)	0,913	
Refractive index (20 degree C)	1,469	
Absorption at 270 nm	0,47	
Iodine value	80	
Saponification value	191	
Acid value	0,21	
Peroxide value	complies	
Unsaponifiable Matter	1,18	%
Colour gardner	1,9	
Colour lovibond 5 1/4" yellow	10,0	
Colour lovibond 5 1/4" red	1,0	
Sesameoil	complies	
Alkaline impurities	complies	
Water	0,02	%

Fatty acid composition

Lauric acid	(C 12:0)	< 0,05 %
Myristic acid	(C 14:0)	< 0,05 %
Myristoleic acid	(C 14:1)	< 0,05 %
	(< C 16:0)	< 0,05 %
Palmitic acid	(C 16:0)	10,75 %
Palmitoleic acid	(C 16:1)	0,82 %
Margaric acid	(C 17:0)	0,08 %
Heptadecic acid	(C 17:1)	0,11 %
Stearic acid	(C 18:0)	3,66 %
Oleic acid	(C 18:1)	76,00 %
Linoleic acid	(C 18:2)	6,89 %
Linolenic acid	(C 18:3)	0,66 %
Octadecatetraen acid	(C 18:4)	< 0,05 %
Arachidic acid	(C 20:0)	0,43 %
Eicosenoic acid	(C 20:1)	0,27 %
Eicosanedienoic acid	(C 20:2)	< 0,05 %
Eicosanetetraenoic acid	(C 20:4)	< 0,05 %
Behenic acid	(C 22:0)	0,12 %
Erucic acid	(C 22:1)	< 0,05 %
Lignoceric acid	(C 24:0)	0,06 %
Nervonic acid	(C 24:1)	< 0,05 %
Other		0,20 %

Sterols in %

beta-sitosterol*	84,0 %
Delta-7-stigmastenol	0,3 %
Campesterol	3,4 %
Stigmasterol	1,5 %
Cholesterol	0,1 %
Stigmasterol	<campesterol

* Sum of contents of beta-Sitosterol, delta 5,23-stigmastadienol, clerosterol, sitostanol, delta 5-avenasterol und delta 5,24-stigmastadienol

Identity	complies
Purity	complies

Figure B.32 — Specifications for Olive Oil, Refined by the applicant (sponsor) documentation

B.6.2.7 Example: Identity of material, combining the elements for Olive Oil, Virgin (EP) and Olive Oil, Refined (EP)

B.6.2.7.1 Table: Olive Oil, Virgin (EP)

Element Name	Value	Data Format	Note #
Element Group: Substance (Fresh)			
Substance Type	Structurally Diverse, herbal Substance	CD	
Substance ID	OLIJF5643S (Artificial ID)	II	
Element Group: Substance Name			
Substance Name	*Olea europaea* L., Fruit	ST	1)
Substance Name Type	Other	CD	
Language	en	CD	
Element Group: Substance Name (Repeat)			
Substance Name (Scientific Name)	*Olea europaea* L.	CD	
Substance Name Type	Other	CD	
Language	en	CD	2)
Element Group: Source Material			
Source Material Class	Organic	CD	
Source Material Type	Plant	CD	
Source Material State	Live	CD	
Organism ID	LHJFG63421 (Artificial ID)	II	3)
Organism Name	*Olea europaea* L.	CD	3)
Development stage	Fruiting	CD	
Element Organism			
Kingdom	Plantae	CD	
Phylum	Tracheophyta	CD	
Class	Magnoliopsida	CD	
Order	Lamiales	CD	
Family	Oleaceae	CD	

Element Name	Value	Data Format	Note #
Genus	Olea	CD	
Species	Europaea	CD	
Species_Primary_ Author	L.	CD	
Element Group: Part			
Part	Fruit	CD	
Part location	Not applicable		
Element Group: Substance Structurally Diverse, herbal Preparation, Substance (Olive Oil, Virgin EP)			
Substance ID	HJFTE78543 (Artificial ID)	**II**	
Substance Type	Structurally Diverse	CD	
Element Group: Substance Name (herbal preparation)			
Substance Name	*Olea europaea*, Fruit, Oil	ST	
Substance Name Type	Other	CD	
Language	en	CD	
Element Group: Substance Name (Repeat)			
Substance Name	Olive Oil, Virgin		
Substance Name Type	Official	CD	
Language	en	CD	
Official Name Type	EP	CD	
Official Name Status	Primary	CD	
Official Name Domain	Drug	CD	
Official Name Jurisdiction	EU,US	CD	
Public Domain	Yes	BL	
Reference Source Type	EP	CD	
Reference Source Class	Official name source	CD	
Reference Source Citation	European Pharmacopoeia, monograph 07/2011:0518	ST	
Parent Substance ID	OLIJF5643S (Artificial ID)	II	
Parent Substance Name	*Olea europaea* L., Fruit	ST	
Substance Role	Parent	CD	
Element Group: Fraction			
Fraction	Oil	CD	
Element Group: Property			
Property Type	Physical	CD	
Property Name	Relative density	ST	
Property Parameter	Measured at 20^0 C.	ST	
Substance_ID	HJFTE78543	II	4)
Substance Name	Olive Oil, Virgin	ST	
Amount Type	Exact	CD	
Element Group: Amount			
Average [Numeric Value]	0.913	PQ	
Unit	Not applicable (dimensionless quantity)	ST	
Element Group: Property (Repeat)			
Property Type	Physical	CD	
Property Name	Acid value	ST	
Substance_ID	HJFTE78543	II	4)
Substance Name	Olive Oil, Virgin	ST	
Amount Type	Weight average	CD	
Element Group: Amount			
Average [Numeric Value]	1.0	PQ	
Low limit	0.10	PQ	

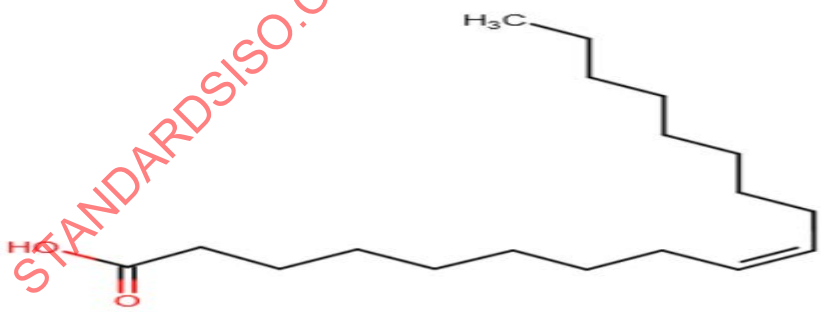
Element Name	Value	Data Format	Note #
High limit	2.0	PQ	
Unit	mg KOH/ 5 g	ST	
Element Group: specified substance Group 3 (Olive Oil, Virgin-EP)			
specified substance Group 3 ID	JFGDRW4526 (Artificial ID)	II	
Substance Type	specified substance Group 3	CD	
specified substance Group 3 Name	*Olea europaea*, Fruit, Oil EP	ST	
Parent Substance ID	HJFTE78543 (Artificial ID)	II	3)
Element Group: specified substance Group 3 Grade			
Grade Type	EP	CD	
Grade Name (= Name of the EP monograph)	Olive Oil, Virgin	ST	
Grade Reference Source	EP monograph 07/2011:0518	ST	

NOTE 1 The substance Name *Olea europaea* L., Fruit is equivalent with the herbal Substance (herbal drug) Name.

NOTE 2 The Scientific name has the Language value en. See 4.5.3 Language in the main body document.

NOTE 3 The Organism ID and Name should be provided by The Kew Gardens Medicinal Plant Names Services (MPNS).

NOTE 4 Oleic Acid should be registered in the database with the minimal information for the element group Structure.

Element Group: Structure			
Structural representation Type	Full	CD	
Structural representation Attachment	In image below	ED	
			
Stereochemistry	Achiral	CD	
Optical activity	None	ST	
Molecular formula	$C_{18}H_{34}O_2$.	ST	
Element Group: Molecular weight			
Molecular weight method	Calculated	CD	
Molecular weight Type	Number average	CD	
Amount Type	Exact	CD	
Element Group: Amount			
Average [Numeric Value]	282.5	PQ	

Unit	g/mol	CD	

B.6.2.7.2 Table Olive Oil, Refined (EP)

Element Group: Substance Structurally Diverse, herbal Preparation, Substance (Olive Oil, Refined)			
Substance ID	KHJGHR4523 (Artificial ID)	**II**	
Substance Type	Structurally Diverse	CD	
Element Group: Substance Name (herbal preparation)			
Substance Name	*Olea europaea* L., Fruit, Oil Refined	ST	
Substance Name Type	Other	CD	
Language	en	CD	
Element Group: Substance Name (Repeat)			
Substance Name	Refined Olive Oil	ST	
Substance Name Type	Other	CD	
Language	en	CD	
Element Group: Substance Name (Repeat)			
Substance Name	Olive Oil, Refined		
Substance Name Type	Official	CD	
Language	en	CD	
Official Name Type	EP	CD	
Official Name Status	Primary	CD	
Official Name Domain	Drug	CD	
Official Name Jurisdiction	EU,US	CD	
Public Domain	Yes	BL	
Reference Source Type	EP	CD	
Reference Source Class	Official name source	CD	
Reference Source Citation	European Pharmacopoeia, monograph 07/2011:1456	ST	
Parent Substance ID	OLIJF5643S (Artificial ID)	II	
Parent Substance Name	*Olea europaea* L., Fruit, Oil	ST	
Substance Role	Parent	CD	
Element Group: Fraction			
Fraction	Oil Refined	CD	
Element Group: Property			
Property Type	Physical	CD	
Property Name	Relative density	ST	
Property Parameter	Measured at 20^0 C.	ST	
Substance ID	KHJGHR4523	II	
Substance Name	Olive Oil, Refined	ST	
Amount Type	Exact	CD	
Element Group: Amount			
Average [Numeric Value]	0.913	PQ	
Unit	Not applicable (dimensionless quantity)	ST	
Element Group: Property (Repeat)			
Property Type	Physical	CD	
Property Name	Acid value	ST	
Substance ID	KHJGHR4523	II	
Substance Name	Olive Oil, Refined	ST	
Amount Type	Weight average	CD	
Element Group: Amount			
Average	0.21	PQ	
Low limit	0.10	PQ	

High limit	0,3	PQ	
Unit	mg KOH/ 10 g	ST	
Element Group: specified substance Group 3 (Olive Oil, Refined-EP)			
specified substance Group 3 ID	SDAUT12543 (Artificial ID)	II	
Substance Type	specified substance Group 3	CD	
specified substance Group 3 Name	*Olea europaea*, Fruit, Oil, Refined, EP	ST	
Parent Substance ID	KHJGHR4523 (Artificial ID)	II	
Element Group: specified substance Group 3 Grade			
Grade Type	EP	CD	
Grade Name	Olive Oil, Refined	ST	
Grade Reference Source	EP monograph 07/2011:1456	ST	

B.6.3 Castor Oil and related products

In the range from substances of vegetable origin somewhere the line has to be drawn whether these materials can be related to the Plant and Plant part and whether the substance is considered to be a mixture substance or another substance type e.g.:

- Castor Oil, Virgin
- Castor Oil, Hydrogenated
- Castor Oil, ethoxylated
- 12-Hydroxystearic acid
- Derivatives produced first by transesterification of the castor oil to methyl ricinoleate, followed by steam cracking to methyl undecylenate and n-heptaldehyde.

The two last examples are obvious: the Castor Oil is used as a starting material followed by a chain of chemical reactions and processes so that the relationship and the character of the Castor Oil material is no longer present in the processed material. These materials are then considered as single substances or as a mixture substance depending on the manufacturing process.

B.6.3.1 Castor Oil, Virgin (EP)

Castor Oil, Virgin is described in the EP as: Fatty oil obtained by cold expression from the seeds of *Ricinus communis* L. with the production specification that during the expression step, the temperature of the oil must not exceed 50 °C.

At the substance level the source of Castor Oil, Virgin (EP) is captured by the Substance (fresh) name *Ricinus communis* L., Seed.

The herbal preparation name for Castor Oil, Virgin (EP) will be:

Substance name: *Ricinus communis,* Seed, Oil with
Substance ID: JKGSDE6452 (Artificial ID); D5340Y2I9G (UNII);
Substance_Type: Other
Common Name: Castor Oil, Virgin,
Substance_Type: Official

B.6.3.2 Castor Oil, Hydrogenated (EP)

Castor Oil, Hydrogenated is described in the EP as: Fatty oil obtained by hydrogenation of Virgin Castor Oil. It consists mainly of the triglyceride of 12-hydroxystearic acid.

The Pharmacopoea refers to Virgin Castor Oil which connects the Plant and Part of the Plant name.

At the substance level the source of Castor Oil, Hydrogenated (EP) is captured by the Substance (fresh) *Ricinus communis* L., Seed.

The herbal preparation name for Castor Oil, Hydrogenated (EP) will be:

Substance Name: *Ricinus communis*, Seed, Oil Hydrogenated
Substance ID: CFDSW67453 (Artificial ID); ZF94AP8MEY(UNII)
Substance_Type: Other
Common Name: Castor Oil, Hydrogenated
Substance_Type: Official.

B.6.3.3 Relationship Castor Oil, Virgin and Castor Oil, Hydrogenated

Castor Oil, Hydrogenated is **not** like Refined Olive Oil, a special grade of Castor Oil.

Castor Oil has been modified (e.g. hydrogenated) by a chemical process which has changed chemically the major component, **Recinoleic acid** [(9*Z*,12*R*)-12-Hydroxyoctadec-9-enoic acid, (12-Hydroxyoleic acid), $C_{18}H_{34}O_3$. (M_r 298.5)] side chains of the triglyceride ester into **12-hydroxystearic acid**, [$C_{18}H_{36}O_3$,(M_w 300,4)] side chains, as in Figure B.33.

This process is to be considered as a chemical modification of the Castor Oil, Virgin into Hydrogenated Castor Oil. This process is a one step manufacturing process with Nickel as catalyst and will be captured in the element modification group at the substance level.

Based on the properties and chemically different main constituents Castor Oil, Virgin and Castor Oil, Hydrogenated are easily distinguished, as in Figure B.33.

NOTE Castor Oil shows optical activity and absorbance at 269 nm and the main constituent of the fatty acid composition is Ricinoleic acid.

Castor Oil, Hydrogenated does not have optical activity and absorbance and the main constituent of the fatty acid is the hydrogenated ricinoleic acid: 12-hydroxystearic acid. The other differences of the properties have more quantitative aspects, except for some of the minor qualitative different fatty acids constituents. This is another argument for the addition of the constituent group in the Structurally Diverse Information model.

B.6.3.4 Castor Oil, Ethoxylated

Castor Oil, Ethoxylated is the synonym for Macrogolglycerol Ricinoleate (EP). It is defined in the EP as mainly ricinoleyl glycerol ethoxylated with 30-50 molecules of ethylene oxide (nominal value), with small amounts of macrogol ricinoleate and of the corresponding free glycols. It results from the reaction of castor oil with ethylene oxide.

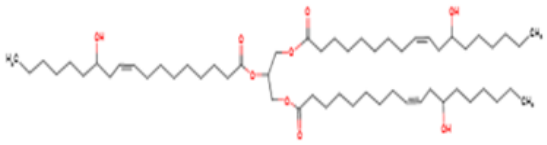
NOTE The EP definition does not speak of Virgin Castor oil, but of castor oil reaction with ethylene oxide:

Characteristics:
- *Appearance*: clear, yellow viscous liquid or semi-solid.
- *Solubility*: freely soluble in water, very soluble in methylene chloride, freely soluble in ethanol (96 %).
- *Relative density*: about 1.05.
- *Viscosity*: 500 mPa·s to 800 mPa·s at 25 °C.

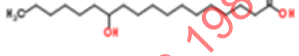
One of the major properties for the identification of this substance material is the viscosity. This property is a predominant property for the chemical Substance_type: "Polymer". In Figure A of Figure B.33 an illustrative representation is presented for Macrogolglycerol Ricinoleate. In Figure B.33 an illustrative representation is presented for the polymer when the polymeric reaction between Castor Oil and ethylene oxide is longer and of a different kind. The processed substance is a "Polymer".

Conclusion: Macrogolglycerol Ricinoleate (synonym Castor Oil, Ethoxylated) is not a modification of Castor Oil, Virgin but is a Substance, with the Substance_type Polymer, in view of the ISO-11238 and is captured at the Substance level having a Substance ID. Castor Oil and Ethylene oxide are both used as starting materials in a polymerizing process.

Figure B.34 shows the specifications for Castor Oil, Virgin and Castor Oil, Hydrogenated .

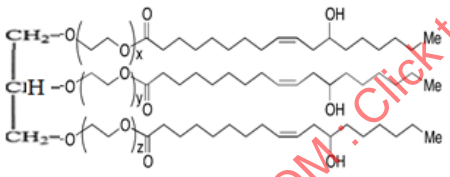


Castor Oil, Main constituent: UNII: D5340Y219G
Glycerol tri-Recinoleic acid ester. :
Mol. formula: $C_{57}H_{104}O_9$; Mol. weight: 933.4;
Side chain: (9*Z*,12*R*)-12-Hydroxyoctadec-9-enoic acid, (synomym: 12-Hydroxyoleic acid); $C_{18}H_{34}O_3$; (M_w 298.5).

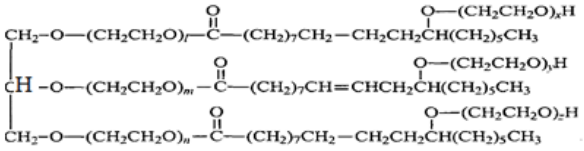


Castor Oil, Hydrogenated, Main Constituent : UNII: ZF94P8MEY
Glycerol tri-12-hydroxystearic acid ester;
Mol. formula: $C_{57}H_{110}O_9$, Mol. weight: 939.4:
Side chain: 12-Hydroxyoctadecanoic acid. (synomym: 2-hydroxystearic acid, $C_{18}H_{36}O_3$ (M_w 300,4)

Macrogolglycerol Ricinoleate (synonym: Castor Oil, Ethoxylated), contains mainly ricinoleyl glycerol ethoxylated with 30-50 molecules of ethylene oxide (nominal value), with small amounts of macrogol ricinoleate and of the corresponding free glycols. It results from the reaction of castor oil with ethylene oxide.



Castor Oil, Ethoxylated: figure A



Different polymerisation process: figure B

Figure B.33 — Structures of the major components of Castor Oil, Virgin, Castor Oil, Hydrogenated and Macrogolglycerol Ricinoleate (synonym Castor Oil, Ethoxylated)

Specifications Castor Oil, Virgin & Castor Oil, Hydrogenated (EP)

- Castor Oil, Virgin

Definition: Fatty oil obtained by cold expression from the seeds of *Ricinus communis* L.
Note: During the expression step, the temperature of the oil must not exceed 50 °C.
Properties tested:

- **Optical rotation**: + 3.5° to + 6.0°.
- **Absorbance:** at 269 nm, NMT 1.5
- Acid value: maximum 1.5
- Hydroxyl value:): minimum 160.
- Peroxide value: Maximum 10
- Unsaponifiable matter: Maximum 0.8 (5.0 g)
- **Composition of Fatty acid:**
- – *palmitic acid*: maximum 2.0 per cent;
- – *stearic acid*: maximum 2.5 per cent;
- – *oleic acid and isomers* : 2.5 per cent to 6.0 per cent;
- – *linoleic acid* : 2.5 per cent to 7.0 per cent;
- – *linolenic acid* : maximum 1.0 per cent;
- – *eicosenoic acid* : maximum 1.0 per cent;
- – ***ricinoleic acid* : 85.0 per cent to 92.0 per cent;**
- – *any other fatty acid*: maximum 1.0 per cent.

- Castor Oil, Hydrogenated

Definition: Fatty oil obtained by hydrogenation of Virgin Castor oil It consists mainly of the triglyceride of 12-hydroxystearic (12-hydroxyoctadecanoic) acid.

Properties tested:
- Acid value: NMT 4.0
- Hydroxyl value: 145 – 165
- Iodine value: NMT 5.0
- Alkaline impurities: Limit test
- **Composition of Fatty acid:**
- *palmitic acid*: not more than 2.0 per cent;
- *stearic acid*: 7.0 per cent to 14.0 per cent;
- *arachidic acid*: not more than 1.0 per cent;
- *12-oxostearic acid*: not more than 5.0 per cent;
- ***12-hydroxystearic acid*: 78.0 per cent to 91.0 per cent;**
- *any other fatty acid*: not more than 3.0 per cent.
- **Nickel** (2.4.31): maximum 1 ppm.

Figure B.34 — Specifications for Castor Oil, Virgin and Castor Oil, Hydrogenated (EP)

B.6.4 Properties to be captured, related to liquids (Gas), Nitrous oxide

Nitrous oxide (EP) is described in the EP monograph as a substance having a content of minimum 98.0 per cent V/V of N_2O in the gaseous phase sampled at 15°C. The substance can occur as a liquid or as a gas. Although the substance is stored liquefied under pressure in suitable containers complying with the legal regulations the physical form is to be considered as a gas and captured as such at the specified substance Group 1 level. Nevertheless the substance should comply with the specifications in the gaseous phase form as well as in the liquid phase form, as in 6.1.12 Physical Form and 6.1.12.1 Physical State.

B.6.4.1 Elements captured at the Substance level and specified substance Level

Element Name	Value	Data Format	Note #
Element Group: Substance			
Substance Type	Chemical	CD	
Substance ID	NGHD56234L (Artificial ID)	II	
Element Group: Substance Name			
Substance Name	Nitrous oxide	ST	
Substance Name Type	Official	CD	
Language	en	CD	
Official Name Type	EP	CD	
Official Name Status	Current	CD	
Official Name Domain	Drug	CD	
Official Name Jurisdiction	EU	CD	
Public Domain	Yes	BL	
Reference Source Type	EP	CD	
Reference Source Class	Official name source	CD	
Reference Source	European Pharmacopoeia	ST	1)

Element Name	Value	Data Format	Note #
Citation			
Reference Source URL	Not applicable		
Element Group: Substance Name (Repeat)			
Substance Name	Dinitrogen Oxide	ST	
Substance Name Type	Other	CD	
Language	en	CD	
Public Domain	Yes	BL	
Reference Source Type	Martindale	CD	
Reference Source Class	Literature	CD	
Reference Source URL	https://www.medicinescomplete.com/mc/martindale/current/3119q.htm?q=dinitrogen%20oxide&t=search&ss=text&p=1#_hit	ST	
Element Group: Substance Code			
Code	10024-97-2	ST	
Code System	CAS Registry	CD	
Code System ID	0049	CD	
Code System Status	Active	CD	
Element Group: Substance Code (Repeat)			
Code	K50XQU1029	ST	
Code System	FDA Substance Registration System (UNII)	CD	
Code System ID	0050	CD	
Code System Status	Active	CD	
Element Group: Substance Classification			
Domain	Human Pharmaceutical	CD	
Substance Classification	ATC	CD	
Substance Classification Code	N01AX13	ST	
Substance Classification Type	Other general anaesthetics	ST	
Public Domain	Yes	BL	
Reference Source Type	Web	CD	
Reference Source Class	Literature	CD	
Reference Source URL	http://www.whocc.no/atc_ddd_index/?code=N01AX13	ST	
Element Group: Structure			
Structural representation type	Full	CD	
Structural representation Attachment	See image below	ED	
			
Element Group: Structure (Repeat)			
Structural representation Type	InChI	CD	
Structural representation	InChI=1S/N2O/c1-2-3	ST	
Public Domain	Yes	BL	
Reference Source Type	Web	CD	

Element Name	Value	Data Format	Note #
Reference Source Class	Literature	CD	
Reference Source URL	http://webbook.nist.gov/cgi/cbook.cgi?ID=C10024972&Mask=1	ST	
Element Group: Structure (Repeat)			
Structural representation Type	MOL	CD	
Structural representation	See below:	ED	
-FDASRS-04301410562D 3 2 0 0 0 0 0 0 0 0999 V2000 -1.5958 -2.6833 0.0000 N 0 0 0 0 0 0 0 0 0 0 0 0 -0.8814 -2.2708 0.0000 N 0 0 0 0 0 0 0 0 0 0 0 0 -0.0625 -1.7167 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0 1 2 3 0 0 0 0 2 3 1 0 0 0 0 M CHG 2 2 1 3 -1 M END			
Public Domain	No	BL	
Reference Source Type	IND	CD	
Reference Source Class	FDA-SRS	CD	
Reference Source ID	222222	ST	
Element Group: Stereochemistry			
Stereochemistry	None	CD	
Optical activity	None	CD	
Molecular Formula	N_2O	ST	
Public Domain	Yes	BL	
Reference Source Type	EP	CD	
Reference Source Class	Official name source	CD	
Reference Source Citation	European Pharmacopoeia	ST	
Public Domain	Yes	BL	
Element Group: Chemical			
Stoichiometric	Yes	BL	
Element Group: Molecular weight			
Molecular weight method	Calculated	CD	
Moelecular weight Type	Number average	CD	
Amount Type	Exact	CD	
Element Group: Amount			
Average [Numeric Value]	44.01	PQ	
Unit	g/mol	CD	
Element Group: specified substance Group 1 (Properties to be described of the Gas)			
specified substance Group 1 ID	PHJSGT785G (Artificial ID)	II	
Substance Type	specified substance Group 1	CD	
specified substance	Nitrous oxide, gas	ST	

Element Name	Value	Data Format	Note #
Group 1Name			
Element Group: Constituent			
Substance ID	NGHD56234L (Artificial ID)	II	
Substance Name	Nitrous oxide	ST	
Substance Role	Parent Substance	CD	
Amount type	Volume Percent	CD	
Element Group: Physical form			
Physical state	Gas	CD	
Physical form type	None	BL	
Amount type	Volume percent v/v of Substance	CD	
Element Group: Amount			
Low limit	98.0	PQ	
Unit	per cent V/V	ST	
Property Parameter	Gaseous phase sampled at -15° C.	ST	2)
Element Group: Property (see figure)			
Physicochemical characterisation Solubility : 1 volume of nitrous oxide dissolves in 1,5 volumes of water. ◆ Physical characteristics — colourless, odourless, non flammable gas. ◆ Molecular mass — 44,013 ◆ Melting point under 101 kPa (1 atm) — -90,81°C ◆ Boiling point under 101 kPa (1 atm) — -88,46°C ◆ Critical temperature — 36,5°C ◆ Critical pressure — 7,271 Mpa (71,99 atm) ◆ Critical density — 452,5 Kg/m^3 ◆ Triple point — -90,80°C sous 87,56 kPa ◆ Density 0°C, P 31,29 bar (equilibrium) — liquid : 904,0 kg/m^3 gas : 85,76 kg/m^3			
Property Type	Physical	CD	
Property Name	Melting point	CD	
Property Parameter	Measured under 101 kPa	ST	
Amount Type	Exact	CD	
Element Group: Amount			
Average [Numeric Value]	-90.81	PQ	
Unit	Degree Celsius	CD	
Element Group: Property (Repeat)			
Property Type	Physical	CD	
Property Name	Boiling point	CD	
Property Parameter	Measured under 101 kPa	ST	
Amount Type	Exact	CD	
Element Group: Amount			
Average [Numeric Value]	-88.46	PQ	
Unit	Degree Celsius	CD	
Element Group: Property (Repeat)			
Property Type	Physical	CD	
Property Name	Critical temperature	CD	
Property Parameter	Measured under 101 kPa	ST	

Element Name	Value	Data Format	Note #
Amount Type	Exact	CD	
Element Group: Amount			
Average [Numeric Value]	36.5	PQ	
Unit	Degree Celsius	CD	
Element: Property (Repeat)			
Property Type	Physical	CD	
Property Name	Critical pressure	CD	
Amount Type	Exact	CD	
Element Group: Amount			
Average [Numeric Value]	71.99	PQ	
Unit	Atm	CD	
Element Group: Property (Repeat)			
Property Type	Physical	CD	
Property Name	Critical density	CD	
Amount Type	Exact	CD	
Element Group: Amount			
Average [Numeric Value]	452.5	PQ	
Unit	Kg/m^3	CD	
Element Group: Property (Repeat)			
Property Type	Physical	CD	
Property Name	Triple point	CD	
Property parameter	Under 87.56 kPa	ST	
Amount Type	Exact	CD	
Element Group: Amount			
Average [Numeric Value]	-90.80	PQ	
Unit	Degree Celsius	CD	
Element Group: Property (Repeat)			
Property Type	Physical	CD	
Property Name	Density	CD	
Property parameter	At 0^oC, 31.29 atm at equilibrium	ST	
Amount Type	Exact	CD	
Element Group: Amount			
Average [Numeric Value]	904.0	PQ	
Unit	Kg/m^3	CD	
Average [Numeric Value]	85.76	PQ	
Unit	Kg/m^3	CD	
Element Group: Property (Repeat)			
Property Type	Physical	CD	
Property Name	Solubility	CD	
Substance ID	059QF0KO0R (UNII)	II	
Substance Name	Water	ST	
Property parameter	1 Volume part 20 °C and at a pressure of 101 kPa	ST	
Amount Type	Number average	CD	
Element Group: Amount			
Average	1.5	PQ	
Unit	Volume Part	CD	

NOTE 1 European Pharmacopoeia Current Edition, (EP, monograph 01/2008:0416).

NOTE 2 List of properties valid for Nitrous oxide, gas.

Element Group: specified substance Group 2 (Nitrous oxide)			
specified substance Group 2 ID	KHJTYD674R (Artificial ID)	II	
Substance Type	specified substance Group 2	CD	
specified substance Group 2 Name	Nitrous oxide-Manufacturer XX	ST	
Parent Substance ID	NGHD56234L (Artificial ID)	II	3)
Element Group: Manufacturing			
Manufacturer ID	DUNS86745HF (Artificial ID)	II	
Manufacturer Name	Company XX	EN	
Manufacturing type	Manufacturer	CD	
Production Method Type	Synthetic	CD	
Production System Type	Continuous thermal decomposition	CD	
Production System	Chemical decomposition reaction decomposition of the ammonium nitrate: $NH_4N0_3 = N_20 + 2H_20$ + heat.	ST	
Critical Process Version Number	1	INT	4)

NOTE 3 This Parent Substance ID refers to the Substance Name: Nitrous oxide.

NOTE 4 The Critical Process Number Version Number would change if a substantial change in the (synthetic) manufacturing process occurred. These would include a change in the chemistry, addition of elimination of a purification step. The change in a version number should be tied to the production system type.

Element Group: specified substance Group 3 (Nitrous oxide)			
specified substance Group 3 ID	GHDRW765T (Artificial ID)	II	
Substance Type	specified substance Group 3	CD	
specified substance Group 3 Name	Nitrous oxide, EP	ST	
Parent Substance ID	NGHD56234L (Artificial ID)	II	5)
Element Group: Grade			
Grade Type	EP	CD	
Grade Name	Nitrous oxide	ST	
Reference Source Type:	EP	CD	
Reference Source Class	Official name source	CD	
Reference Source Citation	European Pharmacopoeia Current Edition, (EP, monograph 01/2008:0416)	CD	

NOTE 5 This Parent Substance ID refers to the Substance Name: Nitrous oxide.

B.7 Examples

This part of the document is meant to walk through the separate information levels and to discuss minimal information required in order to obtain an ISO Substance identifier.

This information must be seen in view of the substance characteristics as discussed in B3, B4 and B5.

For the Substance type Chemical Substances all physico-chemical properties of the Substance are placed at the specified substance Group 1 level, **unless** this property is essential for the defining of the substance and structural information is not enough in order to distinguish the substance from another substance.

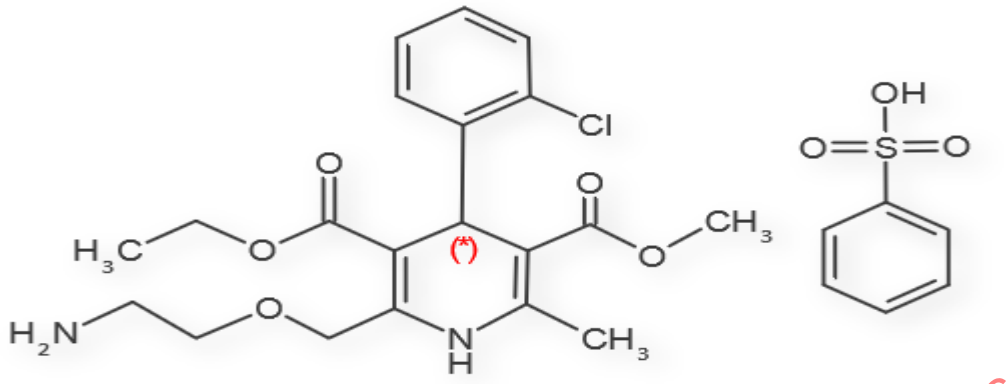
For the Substance type Chemical substances, in almost all situations the physical chemical properties prove to be non-defining elements at the Substance level, but are necessary to describe the character of the substance at the specified substance level and are in most cases mandatory information.

B.7.1 Example 1: Amlodipine besilate

Example: Identity of material, combining the elements for Amlodipine Besilate:

Element Name	Value	Data Format	Note #
Element Group: Substance			
Substance Type	Chemical	CD	
Substance ID	GHTD1234A (Artificial ID)	II	
Element Group: Substance Name			
Substance Name	Amlodipine besilate	ST	
Substance Name Type	Official	CD	
Language	en	CD	
Official Name Type	EP	CD	
Official Name Status	Primary	CD	
Official Name Domain	Drug	CD	
Official Name Jurisdiction	EU	CD	
Public Domain	Yes	BL	
Reference Source Type	EP	CD	
Reference Source Class	Official name source	CD	
Reference Source Citation	European Pharmacopoeia	ST	1)
Reference Source URL	Not applicable		
Element Group: Substance Name (Repeat)			
Substance Name	Amlodipine besylate	ST	
Substance Name Type	Official	CD	
Language	en	CD	
Official Name Type	USP	CD	
Official Name Status	Primary	CD	
Official Name Domain	Drug	CD	
Official Name Jurisdiction	US	CD	
Public Domain	Yes	BL	
Reference Source Type	USP	CD	
Reference Source Class	Official name source	CD	
Reference Source Citation	United States Pharmacopoeia	ST	1)
Element Group: Substance Name (Repeat)			
Substance Name	3-Ethyl 5-methyl (+-)-2-((2-aminoethoxy)methyl)-4-(o-chlorophenyl)-1,4-dihydro-6-methyl-3,5-pyridinedicarboxylate, monobenzenesulfonate	ST	
Substance Name Type	Systematic	CD	
Language	en	CD	
Public Domain	Yes	BL	
Reference Source Type	ChemID*plus*	CD	
Reference Source Class	Web	CD	
Reference Source URL	http://chem.sis.nlm.nih.gov/chemidplus/unii/864V2Q084H	ST	
Element Group: Substance Name (Repeat)			
Substance Name	3-Ethyl 5-methyl (4*RS*)-2-[(2-aminoethoxy)methyl]-4-(2-chlorophenyl)-6-methyl-1,4-dihydropyridine-3,5-dicarboxylate benzenesulphonate	ST	
Substance Name Type	Systematic	CD	
Language	en	CD	
Public Domain	Yes	BL	

Element Name	Value	Data Format	Note #
Reference Source Type	EP	CD	
Reference Source Class	Official name source	CD	
Reference Source Citation	EP monograph	ST	1.a)
Element Group: Substance Name (Repeat)			
Substance Name	3,5-Pyridinedicarboxylic acid, 2-[(2-aminoethoxy)methyl]-4-(2-chlorophenyl)-1,4-dihydro-6-methyl-, 3-ethyl 5-methyl ester, (±)-, monobenzenesulfonate	ST	
Substance Name Type	Systematic	CD	
Language	en		
Public Domain	Yes	BL	
Reference Source Type	USP	CD	
Reference Source Class	Official name source	CD	
Reference Source Citation	USP Monograph	ST	1.b)
Element Group: Substance Name (Repeat)			
Substance Name	A 1095	ST	
Substance Name Type	Company Code	CD	
Language	en		
Public Domain	No	BL	
Reference Source Type	IND	CD	
Reference Source Class	FDA Regulatory Submission	CD	
Reference Source ID	111111	ST	
Element Group: Substance Code			
Code	111470-99-6	ST	
Code System	CAS Registry	CD	
Code System ID	0049	CD	
Code System Status	Active	CD	
Element Group: Substance Code (Repeat)			
Code	864V2Q084H	ST	
Code System	FDA Substance Registration System (UNII)	CD	
Code System ID	0050	CD	
Code System Status	Active	CD	
Element Group: Substance Classification			
Domain	Human Pharmaceutical	CD	
Substance Classification	ATC	CD	
Substance Classification Code	C08CA01	ST	
Substance Classification Type	Selective Calcium channel blockers with mainly vascular effects, Dihydropyridine derivatives	ST	
Public Domain	Yes	BL	
Reference Source Type	WHO Collaborating Centre for Drug Statistics methodology (WHOCC)	CD	
Reference Source Class	Web	CD	
Reference Source URL	http://www.whocc.no/atc_ddd_index/?code=C08C	ST	
Element Group: Structure			
Structural representation Type	Full	CD	
Structural representation Attachment	See image below	ED	2)

Element Name	Value	Data Format	Note #
			
Element Group: Structure (Repeat)			
Structural representation Type	InChI	CD	
Structural representation	InChI=1S/C20H25ClN2O5.C6H6O3S/c1-4-28-20(25)18-15(11-27-10-9-22)23-12(2)16(19(24)26-3)17(18)13-7-5-6-8-14(13)21;7-10(8,9)6-4-2-1-3-5-6/h5-8,17,23H,4,9-11,22H2,1-3H3;1-5H,(H,7,8,9)	ST	
Public Domain	Yes	BL	
Reference Source Type	ChemID*plus*	CD	
Reference Source Class	Web	CD	
Reference Source URL	http://chem.sis.nlm.nih.gov/chemidplus/unii/864V2Q084H	ST	
Element Group: Structure (Repeat)			
Structural Representation Type	SMILES	CD	
Structural representation	CCOC(=O)C1=C(NC(=C(C1c2ccccc2Cl)C(=O)OC)C)COCCN.c1ccc(cc1)S(=O)(=O)O	ST	
Public Domain	Yes	BL	
Reference Source Type	ChemIDplus	CD	
Reference Source Class	Web	CD	
Reference Source URL	http://chem.sis.nlm.nih.gov/chemidplus/unii/864V2Q084H	ST	
Element Group: Structure (Repeat)			
Structural representation Type	MOL	CD	
Structural representation attachment	document TEXT	ED	
Public Domain	No	BL	
Reference Source Type	IND	CD	
Reference Source Class	FDA-SRS	CD	

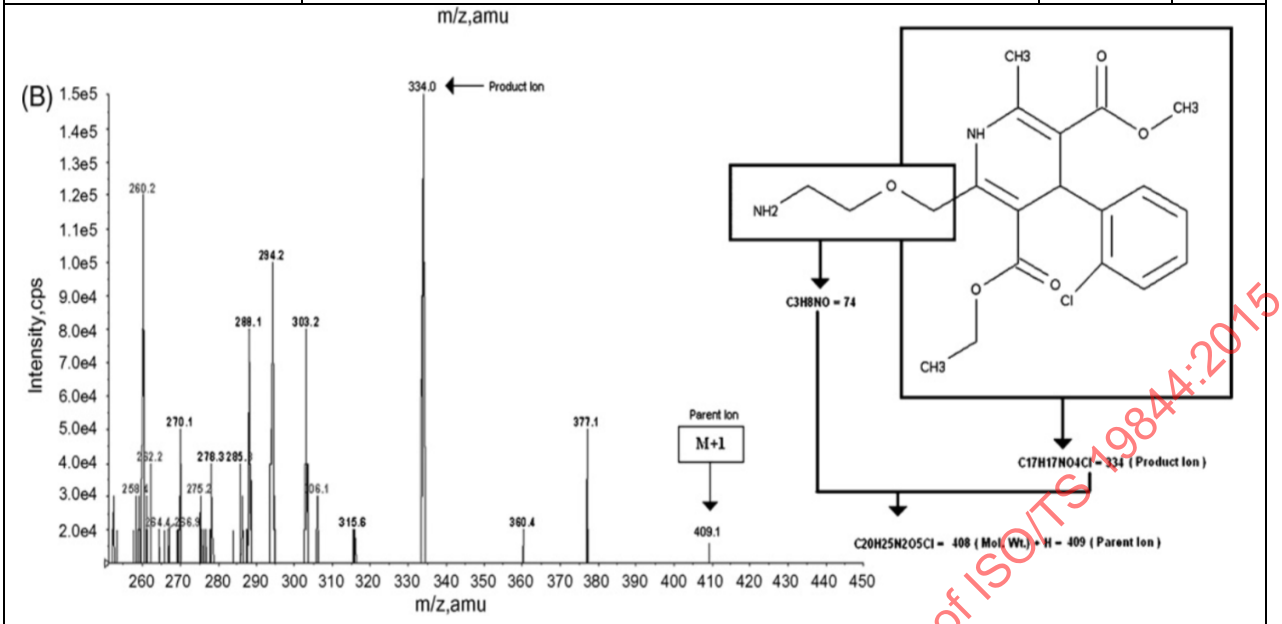
Element Name	Value	Data Format	Note #
Reference Source ID	222222	ST	
Element Group: Reference source Document (New Classes to be added in the ISO Standard in the next version; see 4.6.6. TS/ISO 19844 Implementation Guide)			
Reference Source Document	TEXT, in attachment	CD	
Reference Source Document Type	Regulatory submission	CD	
Structural representation Attachment:			

Symyx 06091422262D 1 1.00000 0.00000 0

38 39 0 0 0 0 0 0 0 0999 V2000
13.1372 -9.5449 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
11.8497 -9.5449 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
12.5081 -9.1824 0.0000 C 0 0 3 0 0 0 0 0 0 0 0 0
11.8497 -10.2824 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
13.1372 -10.2241 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
12.4997 -10.6616 0.0000 N 0 0 3 0 0 0 0 0 0 0 0 0
13.7580 -9.2074 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
11.2163 -9.1658 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
12.5081 -7.9449 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
13.1705 -7.5741 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
11.2163 -8.4199 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0
13.7580 -8.6741 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0
11.2122 -10.6616 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
14.3622 -9.6033 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0
10.5705 -9.5366 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0
13.7956 -10.6533 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
13.8080 -7.9324 0.0000 Cl 0 0 0 0 0 0 0 0 0 0 0 0
8.6580 -10.5908 0.0000 N 0 0 0 0 0 0 0 0 0 0 0 0

Element Name	Value	Data Format	Note #
11.8747 -7.5741 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
10.6330 -10.2699 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0			
13.1705 -6.8283 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
9.3038 -10.1949 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
9.9205 -9.1616 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
9.9205 -10.6616 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
14.9956 -9.1574 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
11.8747 -6.8283 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
9.2705 -9.5366 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
12.5330 -6.4449 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
16.6331 -8.2199 0.0000 S 0 0 3 0 0 0 0 0 0 0 0 0			
17.3247 -8.2199 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0			
15.9289 -8.2199 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0			
16.6331 -8.9241 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
16.6331 -7.5199 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0			
17.2455 -9.2741 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
16.0288 -9.2782 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
16.0288 -9.9824 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
17.2455 -9.9824 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
16.6372 -10.3324 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
24 20 1 0 0 0 0			
25 14 1 0 0 0 0			
26 19 2 0 0 0 0			
27 23 1 0 0 0 0			
28 26 1 0 0 0 0			
4 2 2 0 0 0 0			
28 21 2 0 0 0 0			

Element Name	Value	Data Format	Note #
2 3 1 0 0 0 0 **3 1 1 0 0 0 0** **4 6 1 0 0 0 0** **5 1 2 0 0 0 0** **6 5 1 0 0 0 0** **M END**			
Element Group: Structure (Repeat)			
Stereochemistry	Racemic	CD	
Optical activity	(+-)	CD	
Molecular Formula by Moieties	$C_{20}H_{25}ClN_2O_5$. $C_6H_6O_3S$	ST	2)
Molecular formula	$C_{26}H_{31}ClN_2O_8S$	ST	
Public Domain	Yes	BL	
Reference Source Type	USP	CD	
Reference Source Class	Official name source	CD	
Reference Source Citation	USP Pharmacopeia Current Edition	ST	
Public Domain	Yes	BL	
Element Group: Chemical			
Stoichiometric	Yes	BL	
Element Group: Molecular weight			
Molecular weight method	Calculated	CD	
Moelecular weight Type	Number average	CD	
Amount Type	Exact	CD	
Element Group: Amount			
Average [Numeric Value]	567.1 [567,06 = (408.88 + 158.18); rounded at 567,1]	PQ	3)
Unit	g/mol	CD	
Element Group: Reference Source Document			
Reference Source Document	Pdf. Document , see Figure 3	ED	
Amlodipine besylate mass spectrum obtained by LC-MS/MS, $[M+H]^+$ = 409,1 *m/z* Journal of Chromatography B. 873 (2008) 77-85			

Element Name	Value	Data Format	Note #
			
Reference Source Document Type	Journal article: MS-spectra Amlodipine salts	ST	
Reference Source Classification	Mass spectra	CD	
Reference Source URL	http://www.researchgate.net/profile/Amlan_Kanti_Sarkar/publication/230594874_Simultaneous_determination_of_metoprolol_succinate_and_amlodipinebesylate_in_human_plasma_by_liquid_chromatographytandem_massspectrometry_method_and_its_application_in_bioequivalence_study/links/0912f501b883d741d1000000.pdf	ST	
Element Group Structural Modification			
Structural Modification Type	Moiety	CD	
Element Group: Moiety			
Moiety Role	Salt	CD	
Moiety ID	685928Z18A (UNII)	ST	
Moiety Name	Benzenesulfonic acid	CD	
Amount type	Mole Ratio to Amlodipine	CD	
Element group: Amount			
Average	1	PQ	
NOTE **Element Group Relationship**: - The Active Moiety relationship is a one way relationship; - Parent molecule to a salt/ solvate is a two way relationship. - Both one-and two-way relationships are possible.			
Relationship	Active Moiety	CD	
Substance ID	PTYF75643O (Artificial ID), 1J444QC288 (UNII)	II	
Substance name	Amlodipine	ST	
Element Group: Substance Relationship (Repeat)			
Relationship	Parent (to Salt)	CD	
Substance ID	PTYF75643O (Artificial ID), 1J444QC288 (UNII)	II	

Element Name	Value	Data Format	Note #
Substance name	Amlodipine	ST	
Element Group: Substance Relationship (Repeat)			
Relationship	Salt	CD	
Substance ID	JUTYR9087S (Artificial ID), 685928Z18A (UNII)	II	
Substance name	Benzenesulphonic acid	ST	

NOTE 1 Comments made to the Business Rules in 4.6.2: When the substance is described in a Pharmacopoeial monograph, no reference has to be made to the version, i.e. reference is made to Ph. Eur, USP, or JP must comply to the latest (current) and valid edition. So if reference is made to the Ph. Eur, it means automatically Ph. Eur, 2015 version 8.3 up to July 2015.

1.a) Amlodipine besilate: EP, monograph 04/2012:1491
1.b) Amlodipine besylate: USP Monograph, *USP38–NF33* Page 2207.

NOTE 2 Comments made to the Business Rules in 4.8.6.and 5.1.4. The molecular formula 4.8.6 is specified in accordance with the Hill system. For a salt/hydrate the molecular formula of the base or the acid and the salt/hydrate part are separated by a dot per moiety, e.g. Desmopressin Monoacetate Trihydrate: **Molecular formule by moieties, 4.8.7:** $C_{46}H_{64}N_{14}O_{12}S_2 . C_2H_4O_2 . 3H_2O$.

NOTE 3 **The molecular weight is not a part of the Chemical Substance Class in the ISO 11238 Standard. This element must be included in the next version of the standard. See conceptual datamodel in B.0.1.**

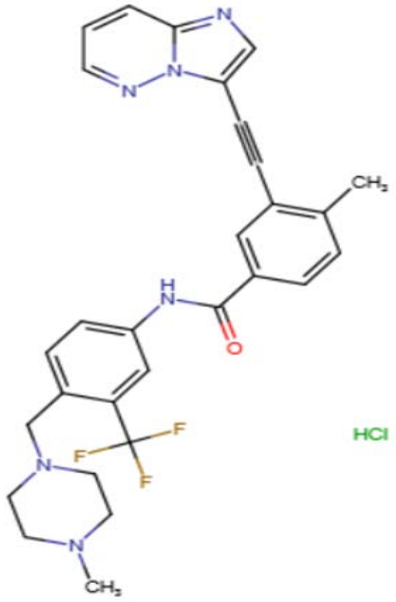
B.7.2 Example 2: Ponatinib hydrochloride

Ponatinib hydrochloride is not described in a Pharmacopoeia.

The following elements are captured at the **Substance level** for Ponatinib hydrochloride:

Element Name	Value	Data Format	Note #
Element Group: Substance			
Substance Type	Chemical	CD	
Substance ID	POYFG4531Y (Artificial ID)	II	
Element Group: Substance Name			
Substance Name	Ponatinib hydrochloride	ST	
Substance Name Type	Official	CD	
Language	en	CD	
Official Name Type	USAN	CD	
Official Name Status	Primary	CD	
Official Name Domain	Drug	CD	
Official Name Jurisdiction	US	CD	
Public Domain	Yes	BL	
Reference Source Type	Martindale	CD	
Reference Source Class	Official name source	CD	
Reference Source Citation	Martindale The complete drug reference	ST	
Reference Source URL	https://www.medicinescomplete.com/mc/martindale/current/ms-28485 - s.htm?q=%22ponatinib%20hydrochloride%22&t=search&ss=text&p=1#_hit		
Element Group: Substance Name (Repeat)			
Substance Name	3-(2-(Imidazo(1,2-b)pyridazin-3-yl)ethynyl)-4-methyl-N-(4-((4-methylpiperazin-1- yl)methyl)-3-	ST	

Element Name	Value	Data Format	Note #
	(trifluoromethyl)phenyl)benzamide monohydrochloride		
Substance Name Type	Systematic	CD	
Language	en	CD	
Public Domain	Yes	BL	
Reference Source Type	ChemID*plus*	CD	
Reference Source Class	Web	CD	
Reference Source URL	http://chem.sis.nlm.nih.gov/chemidplus/unii/96R6PU3D8J	ST	
Element Group: Substance Name (Repeat)			
Substance Name	Benzamide, 3-(2-imidazo[1,2-b]pyridazin-3-ylethynyl)-4-methyl-N-[4-[(4-methyl-1-piperazinyl)methyl]-3-(trifluoromethyl)phenyl]-,hydrochloride (1:1)	ST	
Substance Name Type	Other Name	CD	
Language	en	CD	
Public Domain	Yes	BL	
Reference Source Type	CAS	CD	
Reference Source Class	Other name source	CD	
Reference Source Citation	CAS Registry Name	ST	
Element Group: Substance Name (Repeat)			
Substance Name	XYWZ123	ST	
Substance Name Type	Company Code	CD	
Language	en		
Public Domain	No	BL	
Reference Source Type	Submitter	CD	
Reference Source Class	Regulatory submission	CD	
Reference Source ID	CBG345267C (Artificial ID)	II	
Element Group: Substance Code			
Code	1114544-31-8	ST	
Code System	CAS Registry	CD	
Code System ID	0049	CD	
Code System Status	Active	CD	
Element Group: Substance Code (Repeat)			
Code	96R6PU3D8J	ST	
Code System	FDA Substance Registration System (UNII)	CD	
Code System ID	0050	CD	
Code System Status	Active	CD	
Element Group: Substance Classification			
Domain	Human Pharmaceutical	CD	
Substance Classification	ATC	CD	
Substance Classification Code	L01XE24	ST	
Substance Classification Type	Protein kinase inhibitors	ST	
Public Domain	Yes	BL	
Reference Source Type	WHO Collaborating Centre for Drug Statistics methodology (WHOCC)	CD	
Reference Source Class	Web	CD	
Reference Source URL	http://www.whocc.no/atc_ddd_index/?code=L01XE24&showdescription=yes	ST	
Element Group: Structure			

Element Name	Value	Data Format	Note #
Structural representation Type	Full	CD	
Structural representation Attachment	In image below	ED	
			
Element Group: Structure (Repeat)			
Structural representation Type	InChI	CD	
Structural representation attachment	InChI=1S/C29H27F3N6O.ClH/c1-20-5-6-22(16-21(20)8-10-25-18-33-27-4-3-11-34-38(25)27)28(39)35-24-9-7-23(26(17-24)29(30,31)32)19-37-14-12-36(2)13-15-37;/h3-7,9,11,16-18H,12-15,19H2,1-2H3,(H,35,39);1H	ST	
Public Domain	Yes	BL	
Reference Source Type	ChemID*plus*	CD	
Reference Source Class	Web	CD	
Reference Source URL	http://chem.sis.nlm.nih.gov/chemidplus/unii/96R6PU3D8J	ST	
Element Group: Structure (Repeat)			
Structural Representation Type	SMILES	CD	
Structural representation attachment	Cc1ccc(cc1C#Cc2cnc3n2nccc3)C(=O)Nc4ccc(c(c4)C(F)(F)F)CN5CCN(CC5)C.Cl	ST	
Public Domain	Yes	BL	
Reference Source Type	ChemID*plus*	CD	
Reference Source Class	Web	CD	
Reference Source URL	http://chem.sis.nlm.nih.gov/chemidplus/unii/96R6PU3D8J	ST	
Element Group: Structure (Repeat)			

Element Name	Value	Data Format	Note #
Structural representation Type	MOL	CD	
Structural representation attachment	document TEXT, in attachment below	ED	
Public Domain	No	BL	
Reference Source Type	IND	CD	
Reference Source Class	FDA-SRS	CD	
Reference Source ID	33333	ST	
Element Group: Reference source Document (New Classes to be added in the ISO Standard in the next version; see 4.6.6. TS/ISO 19844 Implementation Guide)			
Reference Source Document	TEXT, in attachment	CD	
Reference Source Document Type	Regulatory submission	CD	
Structural representation Attachment:			

```
Symyx  06091422112D 1   1.00000    0.00000    0

40 43  0    0  0         999 V2000

 -2.2340   -0.1661    0.0000 N  0  0  0  0  0  0        0  0  0

 -1.5965    0.2006    0.0000 N  0  0  3  0  0  0        0  0  0

 -1.5965    0.9381    0.0000 C  0  0  0  0  0  0        0  0  0

 -2.2340    1.3048    0.0000 C  0  0  0  0  0  0        0  0  0

 -2.8716    0.9381    0.0000 C  0  0  0  0  0  0        0  0  0

 -2.8716    0.2006    0.0000 C  0  0  0  0  0  0        0  0  0

 -0.9048    1.1589    0.0000 N  0  0  0  0  0  0        0  0  0

 -0.4632    0.5714    0.0000 C  0  0  0  0  0  0        0  0  0

 -0.9048   -0.0453    0.0000 C  0  0  0  0  0  0        0  0  0

 -0.6840   -0.7327    0.0000 C  0  0  0  0  0  0        0  0  0

 -0.4382   -1.4453    0.0000 C  0  0  0  0  0  0        0  0  0

 -0.2174   -2.1327    0.0000 C  0  0  0  0  0  0        0  0  0

  0.4952   -2.2827    0.0000 C  0  0  0  0  0  0        0  0  0

  0.7160   -2.9953    0.0000 C  0  0  0  0  0  0        0  0  0

  1.4535   -3.1411    0.0000 C  0  0  0  0  0  0        0  0  0

  1.6743   -3.8536    0.0000 O  0  0  0  0  0  0        0  0  0

  1.9451   -2.5994    0.0000 N  0  0  0  0  0  0        0  0  0
```

Element Name	Value	Data Format	Note #
2.6576 -2.7494 0.0000 C 0 0 0 0 0 0 0 0 0			
2.9035 -3.4619 0.0000 C 0 0 0 0 0 0 0 0 0			
3.6159 -3.6077 0.0000 C 0 0 0 0 0 0 0 0 0			
3.8368 -4.2953 0.0000 C 0 0 3 0 0 0 0 0 0			
3.3451 -4.8619 0.0000 F 0 0 0 0 0 0 0 0 0			
4.0618 -5.0119 0.0000 F 0 0 0 0 0 0 0 0 0			
4.5535 -4.4703 0.0000 F 0 0 0 0 0 0 0 0 0			
4.1077 -3.0661 0.0000 C 0 0 0 0 0 0 0 0 0			
4.8201 -3.2161 0.0000 C 0 0 0 0 0 0 0 0 0			
5.3118 -2.6494 0.0000 N 0 0 3 0 0 0 0 0 0			
6.0493 -2.8203 0.0000 C 0 0 0 0 0 0 0 0 0			
6.5410 -2.2578 0.0000 C 0 0 0 0 0 0 0 0 0			
6.2951 -1.5703 0.0000 N 0 0 3 0 0 0 0 0 0			
6.8118 -1.0286 0.0000 C 0 0 0 0 0 0 0 0 0			
5.5827 -1.4203 0.0000 C 0 0 0 0 0 0 0 0 0			
5.0909 -1.9620 0.0000 C 0 0 0 0 0 0 0 0 0			
3.8868 -2.3536 0.0000 C 0 0 0 0 0 0 0 0 0			
3.1493 -2.2077 0.0000 C 0 0 0 0 0 0 0 0 0			
0.2243 -3.5369 0.0000 C 0 0 0 0 0 0 0 0 0			
-0.4882 -3.3869 0.0000 C 0 0 0 0 0 0 0 0 0			
-0.7090 -2.6744 0.0000 C 0 0 0 0 0 0 0 0 0			
-1.4465 -2.5286 0.0000 C 0 0 0 0 0 0 0 0 0			
3.5313 0.2188 0.0000 Cl 0 0 0 0 0 0 0 0 0			
1 2 1 0 0 0			
3 2 1 0 0 0			
3 4 1 0 0 0			
4 5 2 0 0 0			

Element Name	Value	Data Format	Note #
5 6 1 0 0 0			
6 1 2 0 0 0			
3 7 2 0 0 0			
7 8 1 0 0 0			
8 9 2 0 0 0			
2 9 1 0 0 0			
10 9 1 0 0 0			
11 10 3 0 0 0			
12 11 1 0 0 0			
13 12 2 0 0 0			
14 13 1 0 0 0			
15 14 1 0 0 0			
15 16 2 0 0 0			
15 17 1 0 0 0			
17 18 1 0 0 0			
18 19 1 0 0 0			
19 20 2 0 0 0			
20 21 1 0 0 0			
21 22 1 0 0 0			
21 23 1 0 0 0			
21 24 1 0 0 0			
20 25 1 0 0 0			
25 26 1 0 0 0			
26 27 1 0 0 0			
27 28 1 0 0 0			
28 29 1 0 0 0			
29 30 1 0 0 0			

Element Name	Value	Data Format	Note #
30 31 1 0 0 0 30 32 1 0 0 0 32 33 1 0 0 0 27 33 1 0 0 0 25 34 2 0 0 0 34 35 1 0 0 0 18 35 2 0 0 0 14 36 2 0 0 0 37 36 1 0 0 0 38 37 2 0 0 0 38 39 1 0 0 0 12 38 1 0 0 0 M END			
Element Group: Structure (Repeat)			
Stereochemistry	Achiral	CD	
Optical activity	None	CD	
Molecular Formula by moieties	$C_{29}H_{27}F_3N_6O.HCl$	ST	
Molecular formula	$C_{29}H_{28}FClN_6O$	ST	
Public Domain	Yes	BL	
Reference Source Type	ChemIDplus	CD	
Reference Source Class	Web	CD	
Reference Source URL	http://chem.sis.nlm.nih.gov/chemidplus/unii/96R6PU3D8J	ST	
Public Domain	Yes	BL	
Element Group: Chemical			
Stoichiometric	Yes	BL	
Element Group: Molecular weight			
Molecular weight method	Calculated	CD	
Molecular weight Type	Number average	CD	
Amount Type	Exact	CD	
Element Group: Amount			
Average [Numeric Value]	569.02	PQ	1)
Unit	g/mol	CD	
Element Group: Structural Modification			
Structural Modification Type	Moiety	CD	
Element Group: Moiety			
Moiety Role	Salt	CD	

Element Name	Value	Data Format	Note #
Moiety_ID	QTT17582CB (UNII)	II	
Moiety Name	Hydrochloric acid	CD	
Amount Type	Mole Ratio to Ponatinib	ST	
Element group: Amount			
Average Number	1	PQ	
NOTE Eelement Group Relationship: - The Active Moiety relationship is a one way relationship; - Parent molecule to a salt/ solvate is a two way relationship. - Both one-and two-way relationships are possible.			
Element Group: Substance Relationship			
Relationship	Active Moiety	CD	
Substance ID	POFG64523B (Artificial ID), 4340891KFS (UNII)	II	
Substance name	Ponatinib	ST	
Element Group: Substance Relationship (Repeat)			
Relationship	Parent (to Salt)	CD	
Substance ID	POFG64523B (Artificial ID), 4340891KFS (UNII)	II	
Substance name	Ponatinib	ST	
Element Group: Substance Relationship (Repeat)			
Relationship	Salt	CD	
Substance ID	HYDRO67542S (Artificial ID), QTT17582CB (UNII)	II	
Substance name	Hydrochloric acid	ST	
For the specified substance Group1 and 2 see page 72 — 74			

NOTE Since the Substance is not described in any Pharmacopoeia evidence of the **structure** should be **substantiated** by means of Mass spectrometry in order to give substantiate structure and molecular weight. The following information has been provided at the application, which will not be disclosed:

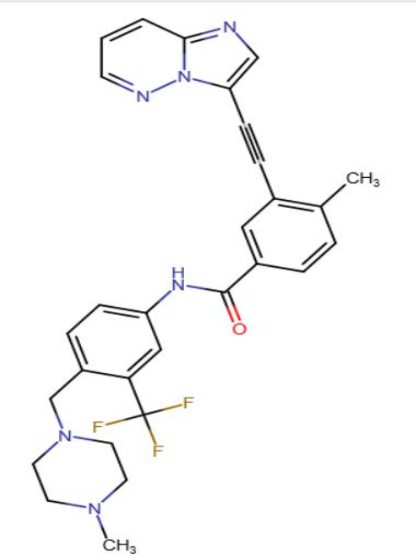
- **Provided information: Accurate Mass Determination:** A molecular ion m/z $[MH]^+$ = 533.2269 Da (corresponding to (ponatinib + H)$^+$ was observed for ponatinib HCl, consistent with the predicted molecular ion m/z $[MH]^+$ = 533.2271 Da;

- **Mass Spectrometry (MS-MS Fragmentation):** The Ponatinib HCl **MS^n** fragmentation pattern follows a well characterised pattern of bond breakage and elimination of labile chemical moieties. The ponatinib fragments observed are consistent with both known fragmentation routes and the chemical structure of ponatinib HCl.

The following elements are captured at the **Substance level** for **Ponatinib (base)**:

Element Name	Value	Data Format	Note #
Element Group: Substance			
Substance Type	Chemical	CD	
Substance ID	GHF26754YQ (Artificial ID)	II	
Element Group: Substance Name			
Substance Name	Ponatinib	ST	
Substance Name Type	Official	CD	
Language	en	CD	
Official Name Type	INN	CD	
Official Name Status	Primary	CD	
Official Name Domain	Drug	CD	
Official Name Jurisdiction	US, EU	CD	
Public Domain	Yes	BL	
Reference Source Type	INN	CD	
Reference Source Class	Official name source	CD	

Element Name	Value	Data Format	Note #
Reference Source Citation	WHO Drug information Vol. 25, no 3, 2011 Recommend INN: list 66	ST	1)
Reference Source URL	http://www.who.int/medicines/publications/druginformation/issues/RL_66.pdf	ST	
Element Group: Substance Name (Repeat)			
Substance Name	3-[2-(imidazo[1,2-*b*]pyridazin-3-yl)ethynyl]-4-methyl-*N*-{4-[(4-methylpiperazin-1-yl)methyl]-3-(trifluoromethyl)phenyl} benzamide	ST	
Substance Name Type	Systematic	CD	
Language	en	CD	
Public Domain	Yes	BL	
Reference Source Type	ChemID*plus*	CD	
Reference Source Class	Web	CD	
Reference Source URL	http://chem.sis.nlm.nih.gov/chemidplus/unii/4340891KFS	ST	
Element Group: Substance Name (Repeat)			
Substance Name	Benzamide, 3-(2-imidazo[1,2-b]pyridazin-3-ylethynyl)-4-methyl-N-[4-[(4-methyl-1-piperazinyl)methyl]-3-(trifluoromethyl)phenyl]-,	ST	
Substance Name Type	Other Name	CD	
Language	en	CD	
Public Domain	Yes	BL	
Reference Source Type	CAS	CD	
Reference Source Class	Other name source	CD	
Reference Source Citation	CAS Registry Name	ST	1.a)
Element Group: Substance Code (Repeat)			
Code	943319-70-8	ST	
Code System	CAS Registry	CD	
Code System ID	0049	CD	
Code System Status	Active	CD	
Element Group: Substance Code (Repeat)			
Code	4340891KFS	ST	
Code System	FDA Substance Registration System (UNII)	CD	
Code System ID	0050	CD	
Code System Status	Active	CD	
Element Group: Substance Classification			
Domain	Human Pharmaceutical	CD	
Substance Classification	ATC	CD	
Substance Classification Code	L01XE24	ST	
Substance Classification Type	Protein kinase inhibitors	ST	
Public Domain	Yes	BL	
Reference Source Type	WHO Collaborating Centre for Drug Statistics methodology (WHOCC)	CD	
Reference Source Class	Web	CD	
Reference Source URL	http://www.whocc.no/atc_ddd_index/?code=L01XE24&showdescription=yes	ST	
Element Group: Structure			
Structural representation Type	Full	CD	

Element Name	Value	Data Format	Note #
Structural representation Attachment	See image below	ED	2)
			
Element Group: Structure (Repeat)			
Structural representation Type	InChI	CD	
Structural representation	InChI=1S/C29H27F3N6O/c1-20-5-6-22(16-21(20)8-10-25-18-33-27-4-3-11-34-38(25)27)28(39)35-24-9-7-23(26(17-24)29(30,31)32)19-37-14-12-36(2)13-15-37/h3-7,9,11,16-18H,12-15,19H2,1-2H3,(H,35,39)	ST	
Public Domain	Yes	BL	
Reference Source Type	ChemID*plus*	CD	
Reference Source Class	Web	CD	
Reference Source URL	http://chem.sis.nlm.nih.gov/chemidplus/unii/4340891KFS	ST	
Element Group: Structure (Repeat)			
Structural Representation Type	SMILES	CD	
Structural representation attachment	Cc1ccc(cc1C#Cc2cnc3n2nccc3)C(=O)Nc4ccc(c(c4)C(F)(F)F)CN5CCN(CC5)C	ST	
Public Domain	Yes	BL	
Reference Source Type	ChemID*plus*	CD	
Reference Source Class	Web	CD	
Reference Source URL	http://chem.sis.nlm.nih.gov/chemidplus/unii/4340891KFS	ST	
Element Group: Structure (Repeat)			
Structural representation Type	MOL	CD	
Structural representation attachment	document TEXT	ED	

Element Name	Value	Data Format	Note #
Public Domain	No	BL	
Reference Source Type	IND	CD	
Reference Source Class	FDA-SRS	CD	
Reference Source ID	4444	ST	
Element Group: Reference source Document (New Classes to be added in the ISO Standard in the next version; see 4.6.6. TS/ISO 19844 Implementation Guide)			
Reference Source Document	TEXT, in attachment	CD	
Reference Source Document Type	Regulatory submission	CD	
Structural representation Attachment:			

```
 Marvin  06101411322D


39 43  0  0  0  0        999 V2000
   7.6413   -1.1539    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   7.0617   -1.7615    0.0000 N  0  0  0  0  0  0  0  0  0  0  0  0
   7.3375   -2.5327    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   6.7859   -3.1637    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   5.9586   -2.9720    0.0000 N  0  0  0  0  0  0  0  0  0  0  0  0
   5.4071   -3.6077    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   4.6079   -3.4395    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   4.3601   -2.6402    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   3.5328   -2.4765    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   2.9812   -3.0842    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   2.1820   -2.9159    0.0000 N  0  0  0  0  0  0  0  0  0  0  0  0
   1.6305   -3.5236    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   1.8782   -4.3229    0.0000 O  0  0  0  0  0  0  0  0  0  0  0  0
   0.8032   -3.3601    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   0.2516   -3.9676    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
  -0.5477   -3.7993    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
  -0.7953   -3.0001    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
  -1.6226   -2.8365    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
```

Element Name	Value	Data Format	Note #
-0.2439 -2.3924 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
0.5555 -2.5607 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
-0.4916 -1.6213 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
-0.7673 -0.8219 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
-1.0150 -0.0508 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
-0.5196 0.6410 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
-1.0150 1.3000 0.0000 N 0 0 0 0 0 0 0 0 0 0 0 0			
-1.7909 1.0523 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
-2.5060 1.4637 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
-3.2213 1.0523 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
-3.2213 0.2250 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
-2.5060 -0.1863 0.0000 N 0 0 0 0 0 0 0 0 0 0 0 0			
-1.7909 0.2250 0.0000 N 0 0 0 0 0 0 0 0 0 3 0 0			
3.2571 -3.8835 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
4.0562 -4.0470 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
4.3040 -4.8184 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
3.7524 -5.4540 0.0000 F 0 0 0 0 0 0 0 0 0 0 0 0			
4.5564 -5.6222 0.0000 F 0 0 0 0 0 0 0 0 0 0 0 0			
5.1080 -5.0147 0.0000 F 0 0 0 0 0 0 0 0 0 0 0 0			
5.7108 -2.2009 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
6.2625 -1.5933 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0			
1 2 1 0 0 0 0			
2 3 1 0 0 0 0			
2 39 1 0 0 0 0			
3 4 1 0 0 0 0			
4 5 1 0 0 0 0			
5 6 1 0 0 0 0			

Element Name	Value	Data Format	Note #
5 38 1 0 0 0 0			
6 7 1 0 0 0 0			
7 8 1 0 0 0 0			
7 33 2 0 0 0 0			
8 9 2 0 0 0 0			
9 10 1 0 0 0 0			
10 11 1 0 0 0 0			
10 32 2 0 0 0 0			
11 12 1 0 0 0 0			
12 13 2 0 0 0 0			
12 14 1 0 0 0 0			
14 15 1 0 0 0 0			
14 20 2 0 0 0 0			
15 16 2 0 0 0 0			
16 17 1 0 0 0 0			
17 18 1 0 0 0 0			
17 19 2 0 0 0 0			
19 20 1 0 0 0 0			
19 21 1 0 0 0 0			
21 22 3 0 0 0 0			
22 23 1 0 0 0 0			
23 24 2 0 0 0 0			
23 31 1 0 0 0 0			
24 25 1 0 0 0 0			
25 26 2 0 0 0 0			
26 27 1 0 0 0 0			
26 31 1 0 0 0 0			

Element Name	Value	Data Format	Note #
27 28 2 0 0 0 0 28 29 1 0 0 0 0 29 30 2 0 0 0 0 30 31 1 0 0 0 0 32 33 1 0 0 0 0 33 34 1 0 0 0 0 34 35 1 0 0 0 0 34 36 1 0 0 0 0 34 37 1 0 0 0 0 38 39 1 0 0 0 0 M END			
Element Group: Structure (Repeat)			
Stereochemistry	Achiral	CD	
Optical activity	None	CD	
Molecular Formula	$C_{29}H_{27}F_3N_6O$	ST	
Public Domain	Yes	BL	
Reference Source Type	ChemID*plus*	CD	
Reference Source Class	Web	CD	
Reference Source URL	http://chem.sis.nlm.nih.gov/chemidplus/unii/4340891KFS	ST	
Public Domain	Yes	BL	
Element Group: Chemical			
Stoichiometric	Yes	BL	
Element Group: Molecular weight			
Molecular weight method	Calculated	CD	
Moelecular weight Type	Number average	CD	
Amount Type	Exact	CD	
Element Group: Amount			
Average [Numeric Value]	532.56	PQ	
Unit	g/mol	CD	
Element group: Substance Relationship			
NOTE **Element Group Relationship:** - The Active Moiety relationship is a one way relationship; - Parent molecule to a salt/ solvate is a two way relationship. - Both one-and two-way relationships are possible.			
Element Group: Substance Relationship			
Relationship	Parent (to Salt)	CD	
Substance ID	POYFG4531Y (Artificial ID), 96R6PU3D8J (UNII)	II	

Element Name	Value	Data Format	Note #
Substance name	Ponatinib hydrochloride	ST	

Discussion on the specified substance Information level

The following elements are captured at the **Group 1 specified substance level for Ponatinib hydrochloride:**

Element Group: specified substance Group 1 (Properties to be described Ponatinib hydrochloride)			
specified substance Group 1 ID	PHIT2345RA (Artificial ID)	II	
Substance Type	specified substance Group 1	CD	
specified substance Group 1 Name	Ponatinib hydrochloride, crystalline, form 1	ST	
Element Group: Constituent			
Substance ID	POYFG4531Y (Artificial ID)	II	
Substance Name	Ponatinib hydrochloride	ST	
Substance Role	Parent Substance	CD	
Amount type	Non-numeric value	CD	
Element Group: Physical form			
Physical state	Solid	CD	
Physical form type	Anhydrous crystalline form 1	CD	1)
Amount type	Non-numeric value	CD	
Element Group: Property			
Property Type	Physical	CD	
Property Name	Solubility	ST	
Property Parameter	Measured at 20°C	ST	
Substance ID	YOW8V9698H (UNII)	II	
Substance Name	Dimethyl Sulfoxide	ST	
Amount Type	Exact	ST	
Element Group: Amount			
Average (Numeric Value)	80	PQ	
High value	90	PQ	
Low value	70	PQ	
Unit	mg/ml	CD	
Element Group: Property (Repeat)			
Property Type	Physical	CD	
Property Name	Solubility	ST	
Property Parameter	Measured at 20°C	ST	
Substance ID	Y4S76JWI15 (UNII)	II	
Substance Name	Methanol	ST	
Amount Type	Exact	ST	
Element Group: Amount			
Average (Numeric Value)	30	PQ	
Unit	mg/ml	CD	
Element Group: Property (Repeat)			
Property Type	Physical	CD	
Property Name	Melting point	CD	
Amount Type	Exact	CD	
Element Group: Amount			
Average (Numeric	263	PQ	

Value)			
High value	264	PQ	
Low value	262	PQ	
Unit	Degree Celsius	CD	
Element Group: Property (Repeat)			
Property Type	Physical	CD	
Property Name	pK_a	ST	
Property Parameter	Measured at pH range of 2.5 -10.5	ST	
Amount Type	Exact	CD	
Element Group: Amount			
Average (Numeric Value)	7.80	PQ	
Average (numeric Value)	2.77	PQ	
Unit	None	CD	
Element: Property (Repeat)			
Property Type	Physical	CD	
Property Name	UV absorption maxima	ST	
Amount Type	Exact	CD	
Element Group: Amount			
Average [Numeric Value]	258	PQ	
Average [Numeric Value]	301	PQ	
Average [Numeric Value]	321	PQ	
Unit	Nanometer	CD	

NOTE When the same manufacture/ or another manufacturer markets the Ponatinib hydrochloride, crystalline as a different polymorph (e.g. crystalline form 2) the Specified substance group 1 Name will change into Ponatinib hydrochloride, crystalline, form 2 with a different Specified substance group 1 ID: [FFFFGGG89].

The following elements are captured at the **specified substance Group 2 level** for Ponatinib hydrochloride, crystalline form 1.

Element Group: specified substance Group 2 (Ponatinib hydrochloride)			
specified substance Group 2 ID	KHJTYD674R (Artificial ID)	II	
Substance Type	specified substance Group 2	CD	
specified substance Group 2 Name	Ponatinib hydrochloride, crystalline form 1-Manufacturer YY	ST	
Parent Substance ID	PHIT2345RA (Artificial ID)	II	1)
Element Group: Manufacturing			
Manufacturer ID	DUNS543210 (Artificial ID)	II	
Manufacturer Name	Company YY	EN	
Manufacturing type	Manufacturer	CD	
Production Method Type	Synthetic	CD	
Production System Type	Chemical synthesis	CD	
Production System	Chemical synthesis from starting materials, purification steps and recrystallization to obtain crystalline form 1	ST	
Critical Process Version Number	1	INT	2)

NOTE 1 This Parent Substance ID refers to the specified substance Group 1Name: Ponatinib hydrochloride, crystalline form 1.

NOTE 2 The Critical Process Number Version Number would change if a substantial change in the (synthetic) manufacturing process occurred. These would include a change in the chemistry, addition of elimination of a purification step. **The change in a version number should be tied to the production system type. Proposed for change in the next version of the ISO 11238 Standard.**

NOTE 3 Since Ponatinib hydrochloride is not described in any pharmacopoeia the specified substance Group 3 information is not applicable.

B.7.3 Example 3: Benzathine Benzylpenicillin tetrahydrate, sterilised.

Introduction

This example discusses the preparation of the sterile substance according to the requirements of the USP with the starting materials Potassium Benzylpenicillin according to the requirements of the EP and Benzathine diacetate with *in house* specifications.

The following elements are captured at the **Substance level** for Benzathine, Benzylpenicillin tetrahydrate sterilised USP:

Element Name	Value	Data Format	Note #
Element Group: Substance			
Substance Type	Chemical	CD	
Substance ID	GHFTE5643B (Artificial ID)	II	
Element Group: Substance Name			
Substance Name	Penicillin G Benzathine	ST	
Substance Name Type	Official	CD	
Language	en	CD	
Official Name Type	USP	CD	
Official Name Status	Primary	CD	
Official Name Domain	Drug	CD	
Official Name Jurisdiction	US	CD	
Public Domain	Yes	BL	
Reference Source Type	USP	CD	
Reference Source Class	Official name source	CD	
Reference Source Citation	USP Pharmacopeia	ST	1)
Element Group: Substance Name (Repeat)			
Substance Name	4-Thia-1-azabicyclo[3.2.0]heptane-2-carboxylic acid, 3,3-dimethyl-7-oxo-6-[(phenylacetyl)amino-], 2[S-(2,5,6)]-, compd. with N,N¢-bis(phenylmethyl)-1,2-ethanediamine (2:1), tetrahydrate.	ST	
Substance Name Type	Systematic	CD	
Language	en	CD	
Public Domain	Yes	BL	
Reference Source Type	USP	CD	
Reference Source Class	Official Name Source	CD	
Reference Source Citation	USP Pharmacopeia	ST	
Element Group: Substance Name (Repeat)			
Substance Name	4-Thia-1-azabicyclo[3.2.0]heptane-2-carboxylic acid, 3,3-dimethyl-7-oxo-6-[(phenylacetyl)amino]-[2S-(2.alpha.,5.alpha.,6.beta.)]-, compd. with N,N'-bis(phenylmethyl)-1,2-ethanediamine (2:1), tetrahydrate	ST	
Substance Name Type	Other Name	CD	

Element Name	Value	Data Format	Note #
Language	en	CD	
Public Domain	Yes	BL	
Reference Source Type	CAS	CD	
Reference Source Class	Other name source	CD	
Reference Source Citation	CAS Registry Name	ST	
Element Group: Substance Name (Repeat)			
Substance Name	XYWZ123 (Artificial Company Code)	ST	
Substance Name Type	Company Code	CD	
Language	en		
Public Domain	No	BL	
Reference Source Type	Submitter	CD	
Reference Source Class	Regulatory submission	CD	
Reference Source ID	CBG564357C (Artificial ID)	ID	
Element Group: Substance Code			
Code	41372-02-5	ST	
Code System	CAS Registry	CD	
Code System ID	0049	CD	
Code System Status	Active	CD	
Element Group: Substance Code (Repeat)			
Code	RIT82F58GK	ST	
Code System	FDA Substance Registration System (UNII)	CD	
Code System ID	0050	CD	
Code System Status	Active	CD	
Element Group: Substance Classification			
Domain	Human Pharmaceutical	CD	
Substance Classification	ATC	CD	
Substance Classification Code	J01CE08	ST	
Substance Classification Type	Beta-lactamase sensitive penicillins	ST	
Public Domain	Yes	BL	
Reference Source Type	WHO Collaborating Centre for Drug Statistics methodology (WHOCC)	CD	
Reference Source Class	Web	CD	
Reference Source URL	http://www.whocc.no/atc_ddd_index/?code=J01CE08&showdescription=yes	ST	
Element Group: Structure			
Structural representation type	Full	CD	
Structural representation Attachment	In image 1 below	ED	

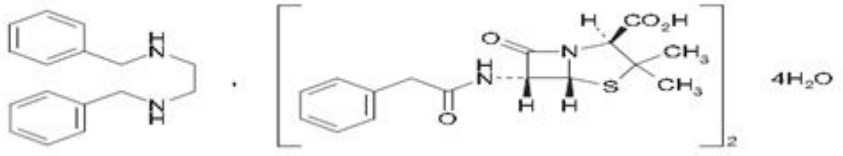
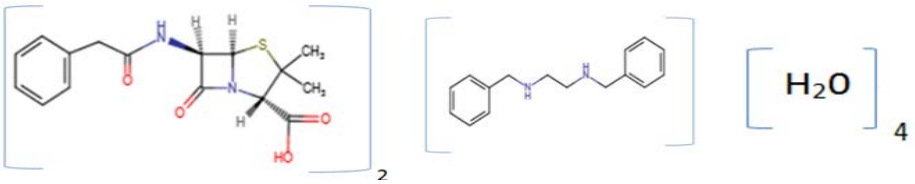


Image 1
Structure Representation in accordance with 4.9

Element Name	Value	Data Format	Note #
 **Image 2**			
Element Group: Structure (Repeat)			
Structural Representation Type	SMILES	CD	
Structural representation	CC1([C@@H](N2[C@H](S1)[C@@H](C2=O)NC(=O)Cc3ccccc3)C(=O)O)C.CC1([C@@H](N2[C@H](S1)[C@@H](C2=O)NC(=O)Cc3ccccc3)C(=O)O)C.c1ccc(cc1)CNCCNCc2ccccc2.O.O.O.O	ST	
Public Domain	Yes	BL	
Reference Source Type	ChemID*plus*	CD	
Reference Source Class	Web	CD	
Reference Source URL	http://chem.sis.nlm.nih.gov/chemidplus/rn/41372-02-5	ST	
Element Group: Structure (Repeat)			
Structural representation Type	MOL	CD	
Structural representation attachment	document TEXT	ED	
Public Domain	No	BL	
Reference Source Type	IND	CD	
Reference Source Class	FDA-SRS	CD	
Reference Source ID	44444	ST	
Element Group: Reference source Document (New Classes to be added in the ISO Standard in the next version; see 4.6.6. TS/ISO 19844 Implementation Guide)			
Reference Source Document	TEXT, in attachment	CD	
Reference Source Document Type	Regulatory submission	CD	

Structural representation Attachment:

```
-ISIS-  10311402042D

 43 45  0  0  1  0  0  0  0  0999 V2000
   3.4083   -6.7959    0.0000 N   0  0  3  0  0  0  0  0  0  0  0  0
   3.4000   -5.9709    0.0000 C   0  0  1  0  0  0  0  0  0  0  0  0
   2.5750   -6.7959    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   2.5750   -5.9709    0.0000 C   0  0  1  0  0  0  0  0  0  0  0  0
   4.1917   -7.0500    0.0000 C   0  0  2  0  0  0  0  0  0  0  0  0
   4.1833   -5.7125    0.0000 S   0  0  0  0  0  0  0  0  0  0  0  0
   4.6750   -6.3792    0.0000 C   0  0  3  0  0  0  0  0  0  0  0  0
   1.7500   -5.9709    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
   4.1875   -7.8709    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   1.0292   -5.5584    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   1.9917   -7.3834    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   3.4667   -8.2834    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   1.0333   -4.7334    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   0.3167   -5.9667    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
```

Element Name	Value	Data Format	Note #

```
    4.9000   -8.2834    0.0000 O  0  0  0  0  0  0  0  0  0  0  0  0
    5.3875   -5.9584    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
    5.3875   -6.7917    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   -0.4000   -5.5542    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   -0.4000   -4.7250    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   -1.1125   -5.9667    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   -1.8292   -5.5584    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   -1.1167   -4.3125    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   -1.8333   -4.7250    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
    3.3917   -5.1459    0.0000 H  0  0  0  0  0  0  0  0  0  0  0  0
    9.6917   -5.9918    0.0000 N  0  0  0  0  0  0  0  0  0  0  0  0
   11.8292   -5.5793    0.0000 N  0  0  0  0  0  0  0  0  0  0  0  0
    8.2667   -6.0001    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   13.2542   -5.5751    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   12.5417   -5.9876    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
    8.9792   -5.5876    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   11.1167   -5.9918    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   10.4042   -5.5793    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
    7.5500   -5.5876    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
    8.2667   -6.8293    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   13.9667   -5.9876    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   13.2500   -4.7543    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
    6.8375   -6.0001    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   14.6792   -5.5751    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   13.9583   -4.3418    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
    7.5500   -7.2418    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
   14.6750   -4.7501    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
    6.8375   -6.8293    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
    9.2375   -8.3751    0.0000 O  0  0  0  0  0  0  0  0  0  0  0  0
  5  1  1  0  0  0  0
  6  2  1  0  0  0  0
  7  5  1  0  0  0  0
  4  8  1  1  0  0  0
  5  9  1  6  0  0  0
 10  8  1  0  0  0  0
 11  3  2  0  0  0  0
 12  9  2  0  0  0  0
 13 10  2  0  0  0  0
 14 10  1  0  0  0  0
 15  9  1  0  0  0  0
 16  7  1  0  0  0  0
 17  7  1  0  0  0  0
 18 14  1  0  0  0  0
 19 18  2  0  0  0  0
 20 18  1  0  0  0  0
 21 20  2  0  0  0  0
 22 19  1  0  0  0  0
 23 22  2  0  0  0  0
  2 24  1  6  0  0  0
  3  4  1  0  0  0  0
  6  7  1  0  0  0  0
 21 23  1  0  0  0  0
  2  1  1  0  0  0  0
  3  1  1  0  0  0  0
 26 31  1  0  0  0  0
```

Element Name	Value	Data Format	Note #
27 30 1 0 0 0 0 28 29 1 0 0 0 0 29 26 1 0 0 0 0 30 25 1 0 0 0 0 31 32 1 0 0 0 0 32 25 1 0 0 0 0 33 27 2 0 0 0 0 34 27 1 0 0 0 0 35 28 2 0 0 0 0 36 28 1 0 0 0 0 37 33 1 0 0 0 0 38 35 1 0 0 0 0 39 36 2 0 0 0 0 40 34 2 0 0 0 0 41 39 1 0 0 0 0 42 37 2 0 0 0 0 40 42 1 0 0 0 0 41 38 2 0 0 0 0 4 2 1 0 0 0 0 M STY 2 1 MUL 2 MUL M SLB 2 1 1 2 2 M SAL 1 15 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 M SAL 1 9 16 17 18 19 20 21 22 23 24 M SPA 1 15 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 M SPA 1 9 16 17 18 19 20 21 22 23 24 M SDI 1 4 -2.3000 -8.8000 -2.3000 -3.8500 M SDI 1 4 6.3500 -3.8500 6.3500 -8.8000 M SMT 1 2 M SAL 2 1 43 M SPA 2 1 43 M SDI 2 4 8.7500 -9.0500 8.7500 -7.9000 M SDI 2 4 10.1500 -7.9000 10.1500 -9.0500 M SMT 2 4 M END			
Element Group: Structure (Repeat)			
Stereochemistry	Chiral	CD	
Optical activity	(-)	CD	
Molecular Formula by moieties	$[C_{16}H_{20}N_2]$.$[C_{16}H_{18}N_2O_4S]_2$. $4H_2O$	ST	
Molecular formula	$C_{48}H_{64}N_6O_{12}S_2$	ST	
Public Domain	Yes	BL	
Reference Source Type	FDA-SRS	CD	
Reference Source Class	Web	CD	
Reference Source URL	http://fdasis.nlm.nih.gov/srs/ProxyServlet?mergeData=true&objectHandle=DBMaint&APPLICATION_NAME=fdasrs&actionHandle=default&nextPage=jsp/srs/ResultScreen.jsp&TXTSUPERLISTID=RIT82F58GK&QV1=PENICILLIN+G+BENZATHINE+%5BORANGE+BOOK%5D	ST	
Public Domain	Yes	BL	
Element Group: Chemical			
Stoichiometric	Yes	BL	
Element Group: Molecular weight			
Molecular weight method	Calculated	CD	
Moelecular weight	Number average	CD	

Element Name	Value	Data Format	Note #
Type			
Amount Type	Exact	CD	
Element Group: Amount			
Average (Numeric Value)	981,20	PQ	1)
Unit	g/mol	CD	
Element Group: Structural Modification			
Structural Modification Type	Moiety	CD	
Element Group: Moiety			
Moiety Role	Salt	CD	
Moiety_ID	C659VZ7P7T (UNII)	ST	
Moiety Name	Benzanthine	CD	
Amount Type	Mole Ratio to 2 molecules Penicillin G	ST	
Element group: Amount			
Average Number	1	PQ	
Element Group: Moiety			
Moiety Role	Hydrate	CD	
Moiety_ID	059QF0KO0R (UNII)	ST	
Moiety Name	Water	CD	
Amount Type	Mole Ratio to Penicillin G	ST	
Element group: Amount			
Average Number	2	PQ	
NOTE 1 **Element Group Relationship:** - The Active Moiety relationship is a one way relationship. - Parent molecule to a salt/ solvate is a two way relationship. - Both one-and two-way relationships are possible.			
Element Group: Substance Relationship			
Relationship	Active Moiety	CD	
Substance ID	PENC56432TY (Artificial ID), Q42T66VG0C (UNII)	II	
Substance name	Penicillin G	ST	
Element Group: Substance Relationship (Repeat)			
Relationship	Parent (to Salt)	CD	
Substance ID	PENC56432TY (Artificial ID), Q42T66VG0C (UNII)	II	
Substance name	Penicillin G	ST	
Element Group: Substance Relationship (Repeat)			
Relationship	Parent (to hydrate)	CD	
Substance ID	WATHJT6453 (Artificial ID), 059QF0KO0R (UNII)	II	
Substance name	Water	ST	

B.7.4 specified substance Group 2 information level

The following elements are captured at the **Group 2 Substance level** for Benzathine, Benzylpenicillin tetrahydrate, sterilised (USP):

Element Group: specified substance Group 2 (Benzathine Benzylpenicillin tetrahydrate)			
specified substance Group 2 ID	BENZ97653M (Artificial ID)	II	
Substance Type	specified substance Group 2	CD	

specified substance Group 2 Name	Benzathine Benzylpenicillin tetrahydrate -Manufacturer ZZ	ST	
Parent Substance ID	GHFTE5643B (Artificial ID)	II	1)
Element Group: Manufacturing			
Manufacturer ID	DUNS543210 (Artificial ID)	II	
Manufacturer Name	Company ZZ	EN	
Manufacturing type	Manufacturer	CD	
Production Method Type	Fermentation and Chemical synthesis	CD	
Production System Type	Chemical synthesis	CD	
Production System	Fermentation, Chemical synthesis from starting materials, purification steps and recrystallization to obtain crystalline form.	ST	
Critical Process Version Number	1	INT	2)

NOTE 1 This Parent Substance ID refers to the Substance Name: Penicillin G Benzathine.

NOTE 2 The Critical Process Number Version Number would change if a substantial change in the (synthetic) manufacturing process occurred. These would include a change in the chemistry, addition of elimination of a purification step. **The change in a version number should be tied to the production system type.**

B.7.5 specified substance Group 3 information level

The following elements are captured at the **specified substance Group 3 level** for Benzathine Benzylpenicillin tetrahydrate, sterilised (USP):

Element Group: specified substance Group 3 (Benzathine Benzylpenicillin anhydrous, sterilised (USP)			
specified substance Group 3 ID	JKHYR7865P (Artificial ID)	II	
Substance Type	specified substance Group 3	CD	
specified substance Group 3 Name	Benzathine Benzylpenicillin tetrahydrate, sterilised-USP	ST	
Parent Substance ID	GHFTE5643B (Artificial ID)	II	1)
Element Group: Grade			
Grade Type	USP	CD	
Grade Name	Benzathine Benzylpenicillin tetrahydrate sterilised-USP	ST	
Reference Source Type:	USP, Monograph *USP38–NF33* Page 4781	ST	
Reference Source Class	Official name source	CD	
Reference Source Citation	http://www.uspnf.com/uspnf/pub/index?usp=38&nf=33&s=1&officialOn=August 1, 2015	CD	

NOTE 1 This Parent Substance ID refers to the Substance Name: Penicillin G Benzathine.

B.8Radiopharmaceutical substance

B.8.1 Introduction

The ISO 11238 Substance Standard describes a nuclide and defines the isotope element group based on the type of the underlying substance and not a type of the substance in and of itself.

Characteristics for each nuclide could at least include half-life, energy of emission, parent and daughter nuclides. For this version of the ISO Chemical Substance ANNEX regarding minimal requirements the Isotope information is included and **not** half-life and unit. The information model for structure and isotope is shown in Figure B.35: