DRAFT INTERNATIONAL STANDARD ISO/DIS 11238

ISO/TC **215** Secretariat: **ANSI**

Voting begins on: **2010-09-23** Voting terminates on: **2011-02-23**

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION • МЕЖДУНАРОДНАЯ ОРГАНИЗАЦИЯ ПО СТАНДАРТИЗАЦИИ • ORGANISATION INTERNATIONALE DE NORMALISATION

Health Informatics — Identification of medicinal products — Data elements and structures for unique identification and exchange of regulated information on substances

Informatique de santé — Identification des produits médicaux — Éléments de données et structures pour l'identification unique et l'échange d'informations réglementées concernant les substances

ICS 35.240.80

ISO/CEN PARALLEL PROCESSING

This draft has been developed within the International Organization for Standardization (ISO), and processed under the **ISO-lead** mode of collaboration as defined in the Vienna Agreement.

This draft is hereby submitted to the ISO member bodies and to the CEN member bodies for a parallel five-month enquiry.

Should this draft be accepted, a final draft, established on the basis of comments received, will be submitted to a parallel two-month approval vote in ISO and formal vote in CEN.

In accordance with the provisions of Council Resolution 15/1993 this document is circulated in the English language only.

Conformément aux dispositions de la Résolution du Conseil 15/1993, ce document est distribué en version anglaise seulement.

To expedite distribution, this document is circulated as received from the committee secretariat. ISO Central Secretariat work of editing and text composition will be undertaken at publication stage.

Pour accélérer la distribution, le présent document est distribué tel qu'il est parvenu du secrétariat du comité. Le travail de rédaction et de composition de texte sera effectué au Secrétariat central de l'ISO au stade de publication.

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.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

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.

ISO 11238 was prepared by Technical Committee ISO/TC 215, *Health informatics*, and by Technical Committee CEN/TC 251, *Health informatics* and in collaboration and with the co-operation of the Clinical Data Interchange Standards Consortium (CDISC), Health Level Seven (HL7) and the International Health Terminology Standards Development Organisation (IHTSDO).

Introduction

This standard was developed in response to a worldwide demand for internationally harmonised specifications for medicinal products. It is one of a group of five standards which together provide the basis for the unique identification of medicinal products. The group of standards comprises:

- ISO/DIS 11615 Health Informatics - Identification of medicinal products - Data elements and structures for the unique identification and exchange of regulated medicinal product information

- ISO/DIS 11616 Health informatics - Identification of medicinal products - Data elements and structures for the unique identification and exchange of regulated pharmaceutical medicinal product

- ISO/DIS 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/DIS 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.

To meet the primary objectives of the regulation of medicines and pharmacovigilance it is necessary to reliably exchange medicinal product information in a robust and reliable manner. The IDMP standards therefore support the following interactions:

— Regulator to regulator e.g. European Medicines Agency to the US Food and Drug Administration (FDA) or vice versa

— Pharmaceutical company to regulator e.g. Pharma Company A to Health Canada

— Sponsor of clinical trial to regulator e.g. University X to Austrian Medicines Agency

— Regulator to other stakeholders e.g. UK Medicines Health Regulatory Agency (MHRA) to National Health System (NHS)

— Interaction of regulator with worldwide-maintained data sources e.g. Pharmaceutical and Medical Device Agency (PMDA) and the assignment of a new substance identifier.

The necessary messaging specifications are included as an integral part of the IDMP standards to secure the interactions above.

Unique identifiers produced in conformance with the IDMP standards are aimed to support applications where it is necessary to reliably identify and trace the use of medicinal products and the materials within medicinal products.

The standard allows for the establishment and maintenance of a system that is capable of assigning unique identifiers for all substances in medicinal products or in packaging materials in which medicinal products are contained. The standard describes general rules for defining and distinguishing substances and a high level model for the structuring of information for substances and specified substances. A detailed element-by-element schema will be developed in an implementation guide for the use of this standard and in a messaging standard being developed by HL-7. The implementation of this standard will result in a strong non-semantic unique identifier for every substance present in a medicinal product. The use of the identifier is essential for the description of substances in medicinal products on a global scale. This standard does not involve developing nomenclature for substances or specified substances but common and official substance names in current use can be mapped to each identifier.

This standard provides high level models for the organization and capture of data, examples using a detailed information model are provided in Annex C, these examples are for informative purposes only. Implementation guides and HL7 messaging will be developed that will provide detailed schema for the capture and transmission of information on substances and specified substances. It is anticipated that a single global maintenance organization will be responsible for the generation of identifiers for every substance and such organization would retain the defining elements upon which the substance identifier was based. At the specified substance level a more regional approach may be necessary because of the proprietary nature of much of the information captured at this level.

There are many terms in use to describe basic concepts in the regulatory and pharmaceutical standards development domain for different purposes and in different contexts. The terms and definitions described in this standard shall be applied for the concepts which are required to uniquely identify, characterise and exchange regulated medicinal products and associated information.

The terms and definitions adopted in this standard are intended to facilitate the interpretation and application of legal and regulatory requirements but they shall be without prejudice to any legally binding document. In case of doubt or potential conflict, the terms and definitions contained in legally binding documents shall prevail.

Health Informatics — Identification of medicinal products — Data elements and structures for unique identification and exchange of regulated information on substances

1 Scope

This document provides an information model to define and identify substances within medicinal products or used for medicinal purposes, including dietary supplements, food and feed additives and cosmetics. The document references other standards and external terminological resources that are applicable to this standard.

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/DIS 11615, *Health informatics — Identification of Medicinal—Data elements and structures for the unique identification and exchange of regulated medicinal product information*

ISO/DIS 11616, *Health informatics — Identification of Medicinal Products — Data elements and structures for the unique identification and exchange of regulated pharmaceutical medicinal product*

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

ISO/HL7 27951:2009, *Health informatics — Common terminology services, release 1*

ISO/DIS 27953*, Health Informatics – Pharmacovigilance – Individual Case Safety Report*

ISO 2382-4:1999, *Information technology - Vocabulary - Part 4: Organization of data*

ISO 1087-1:2000, *Terminology work - Vocabulary - Part 1: Theory and application*

ISO 1087-2:2000, *Terminology work - Vocabulary - Part 2: Computer applications*

ISO/IEC 7064:2003, *Information technology - Security techniques - Check character systems*

ISO 8601:2004, *Data elements and interchange formats - Information interchange - Representation of dates and times*

ISO 3166-1 *alpha-2 codes as defined in ISO 3166-1, part of the ISO 3166 (ISO), to represent countries, dependent territories, and special areas of geographical interest*

IUPAC, *Compendium of Chemical Terminology version 2.1.5*

ISBT 128, *Standard Technical Specification version 3.6*

HL7 Common Product Model, *Substance Model (POCP_MT080100) and Substance Specification Model (POCP_MT090100) [*6]

HL7 Version 3 Standard, *Common Terminology Services HL7 Draft Standard for Trial Use DSTU Release 2, 14 October 2009, HL7 Inc.*

ISO/HL7 27951:2009, *Health informatics — Common terminology services, release 1*

HL7 Version 3 Standard, *Common Clinical Product Model, Release 1, Last Ballot: Normative Ballot 1 - January 2009*

HL7 Version 3 Standard*, Common Product Model CMETS, Release 10, DSTU Ballot 3 - January 2010*

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use Draft Consensus Guideline — *Data Elements and Standards for Drug Dictionaries — M5 Revision 4, 2 February 2007*

3 Terms, definitions and abbreviations

3.1 Terms and definitions

For the purposes of this document, the following terms, definitions and abbreviations apply.

3.1.1
active marker
constituents or groups of constituents of an herbal substance, herbal preparation or a herbal medicinal product, that are of interest for control purposes, and which are generally accepted to contribute to the therapeutic activity

NOTE Active markers are not equivalent to analytical or signature markers that serve solely for identification or control purposes.

3.1.2
analytical data
set of elements to describe and capture methods and reference material used to determine purity, potency, or identity in a specified substance

3.1.3
chemical
type of substance defined by a single molecular structure that is not a protein or nucleic acid

NOTE A chemical is generally considered a “small” molecule, salt, solvate or ion, with a single definitive or representative structure.

3.1.4
component
role in the description of a specified substance indicates the substance or specified substance is an intended constituent

EXAMPLE Dimethicone and silicon dioxide are the components in simethicone. Human Insulin and protamine are the components in Human Insulin Isophane.

NOTE Components are used to describe the substances and specified substances that form an intermediate product.

3.1.5
composition stoichiometry
quantitative relationships between the chemical elements or moieties that make up a substance. Materials with different composition stoichiometry shall be defined as different substances

EXAMPLE SODIUM PHOSPHATE, DIBASIC, HEPTAHYDRATE and SODIUM PHOSPHATE, DIBASIC, DIHYDRATE are different substances.

3.1.6
constituent
substance or specified substance present within a specified substance . Constituents shall have a role and amount associated with them

NOTE Constituents can be impurities, degradents, active marker or signature substances, or single substances mixed together to form a product. It will also allow the capturing of specifications on material which could include limits on specific impurities or related substances.

3.1.7
controlled vocabulary
A finite set of values that represent the allowed values for a data item. These values may be codes, text, or numeric

[CDISC Clinical Research Glossary V8.0, 2009]

3.1.8
copolymer
polymer with more than one type of single repeating unit linked through covalent bonds

NOTE Copolymers are obtained by copolymerization or sequential polymerization of two monomers. Copolymers can be random, alternating, block or graft.

3.1.9
critical processing
manufacturing steps necessary for production of a specified substance

3.1.10
DCF
French approved name

3.1.11
degree of polymerization
number of single repeating units in a polymer or polymeric block or chain

NOTE Applies to both homopolymers and block copolymers where it refers the degree of polymerisation within a block.

3.1.12
diverse origin
refers to materials that are not isolated together or the result of the same synthetic process

EXAMPLE Material derived from the root and flower of the same plant would be defined as two substances since are not physically connected. Material isolated from the stem and leaf of plant which are physically connected would be a single substance.

NOTE Basis upon which a material is defined as a single substance or as multiple substances. Material isolated from non-contiguous part of a biological matrix would be defined as separate substances.

3.1.13
enhancer
cis-acting sequence of DNA that increases the utilization of (some) eukaryotic promoters and can function in either orientation and in any location (upstream or down stream) relative to the promoter

3.1.14
established name
adopted name / approved name / common name / nonproprietary name / proper name / name given by an official registration authority

NOTE Typical registration authorities include BAN, DCF, INCI, INN, JAN, , and USAN. Controlled vocabulary will be developed as required.

3.1.15
fraction
distinct portion of material derived from a complex matrix, the composition of a fraction is always different from that of the antecedent material from which it is derived

EXAMPLE Sera to Immunoglobulins to IGG is an example of recursive fractionation.

NOTE Concept used to describe source material. Concept is recursive in that a subsequent fraction can be derived from an antecedent fraction which is implied in the order of listing.

3.1.16
gene
basic unit of hereditary material that encodes and controls the expression of a protein or protein subunit

3.1.17
gene element
individual elements within a gene such as promoter, enhancer, silencing or coding sequences

3.1.18
glycosylation
enzymatic process that links saccharides or oligosaccharides to proteins, lipids, or other organic molecules

3.1.19
glycosylation type
defining element for glycoproteins, captures significant differences in glycosylation between classes of organism

NOTE 1 Controlled vocabulary will be developed as required.

NOTE 2 Distinguishes the pattern of glycosylation within a class of organisms, current glycosylation types include yeast, fungal, plant, insect, avian, mammalian, human

3.1.20
grade
set of specifications indicating the quality of a specified substance

3.1.21
homopolymer
polymer containing a single structural repeating unit

3.1.22
intermediate product
multiple substances and/or specified substances of diverse origin that are brought together to form a product used in the formulation of medicinal products

EXAMPLE Materials such as Human Insulin Isophane, Aluminum Lakes, Nicotine Polacrilex, and Phosphate Buffered Saline are all intermediate products

NOTE Intermediate products are specified substances. Any medicinal product used to formulate another medicinal products could also be considered an intermediate product.

3.1.23
isotopes
nuclides having the same atomic number but different atomic mass numbers. Radionuclides or nuclides with a non-natural isotopic ratio shall be represented in the structural representation with the nuclide number displayed. Natural abundance isotopes shall be represented by a plain elemental symbol.

EXAMPLE ^{13}C refers to a carbon atom

3.1.24
manufacturing
element set to describe the company that produces and the critical processes that result in a given specified substance

3.1.25
material
any real matter

3.1.26
microheterogeneity
refers to proteins isolated together that contain minor differences in glycosylation, or post-translational modification, or sequence heterogeneity. Microheterogeneity of proteins will not be captured in either substances or specified substance.

NOTE Microheterogeneity consists of variability in the type of glycosylation (biantennary, triantennary), extent of glycosylation at a given site (site occupancy), sequence heterogeneity due to polymorphism in source material, translation errors or variable proteolytic processing. .

3.1.27
medicinal product
any substance or combination of substances, which may be administered to human beings or animals for treating or preventing disease, with the view to making a medical diagnosis or to restore, correct or modify physiological functions

[ENV 13607, ENV 12610]

NOTE 1 A medicinal product may consist of one or several Pharmaceutical Products.

NOTE 2 In certain jurisdictions a medicinal product may also be defined as any substance or combination of substances which may be used to make a medical diagnosis.

3.1.28
mixture substance
substance that is a combination of single substances isolated together or produced in the same synthetic process. Substances of diverse origin that are brought together to form a product will not be defined as a mixture substance.

EXAMPLE Gentamicin would be defined as a mixture substance of Gentamicin C1A, Gentamicin C1, and Gentamicin C2. Glyceryl monoesters could be defined as a mixture substance of two single substances which differ in the position of esterification. Simethicone which consists of dimethicone and silicon dioxide would not be defined as a mixture substance since these are diverse materials brought together to form a product.

NOTE Impurities, related substances, or degradents are not considered single substances within a mixture substance. Mixtures will also be allowed when substance ambiguity exists in authoritative sources such as pharmacopeias.

3.1.29
moiety
entity within a substance that has a complete and continuous molecular structure. When the strength of a product is based on a moiety, the moiety should be defined in a consistent manner across all products. In order to have a definitive moiety, the free acid and free base should be used as the moiety upon which strength should be based.

NOTE Within this document moiety will be used in the context of non-stoichiometric chemicals. Moieties in general are also be single substances, ions, or solvate molecules.

3.1.30
molecular fragment
portion of a molecule that has one of more sites of attachment Molecular fragments will be used in the description of polymers to represent substituents and may be used to represent modifications to a given substance

3.1.31
molecular structure
unambiguous representation of the arrangement of atoms. For the purposes of defining substances, the three dimensional conformations will normally not be captured. Individual conformations or conformers of substances would only be captured in either a general sense for proteins (i.e. denatured) or when a given rotation about a single bond is restricted in such a way that the two different conformers are isolatable from each other and do not interconvert at room temperature. (i.e. substituted biphenyls).

NOTE This representation should be generally translatable into a graphical representation.

3.1.32
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. The type of molecular weight should always be captured.

NOTE For polymers there are several different types of molecular weight (i.e. weight average, number average etc.).

3.1.33
nucleic acid
substance defined by a linear sequence of nucleotides linked through phosphate esters

NOTE The type of nucleic acid (RNA, DNA) will also be captured. Oligonucleotides and gene elements (i.e. promoters, enhancers, coding sequences, and silencers) would also be defined as nucleic acid substances.

3.1.34
parent organism
organism from which biological source material is derived

3.1.35
part
anatomical origin and location of source material within an organism

3.1.36
pharmaceutical product
qualitative and quantitative composition of the product as administered to the patient in line with the regulated product information

NOTE 1 A medicinal product may contain one or more pharmaceutical products.

NOTE 2 In many instances the pharmaceutical product is equal to the manufactured item. However, there are instances where the manufactured item(s) must undergo a transformation before being administered to the patient (as the pharmaceutical product) and the two are not equal.

3.1.37
physical form
physical state, either gas, liqiuid, or solid, and the type of organization for solid matter

NOTE Solids can be either crystalline or amorphous, polymorphism can also be captured

3.1.38
polydispersity
measure of the range of molecular masses in a polymeric substance

NOTE The polydispersity index of polymers is typically calculated by the ratio of weight average molecular weight to number average molecular weight

3.1.39
polymer
type of substance that is inherently heterogeneous that contains structural repeat units linked by covalent bonds

NOTE Proteins and nucleic acid with defined sequence will not be defined as a polymer.

3.1.40
post-translational modification
modifications of a protein that typically occur in vivo during or after translation

NOTE Post-translational modification will be described within the structural representation and not as a modification of a protein

3.1.41
processing material
material essential to the manufacturing process that is not incorporated into the resultant material

3.1.42
protein
substance defined by one or more linear sequences of natural amino acid residues. Site of glycosylation, disulfide bonds, and general type of glycosylation (yeast, plant, insect, mammalian, human) will also be captured as defining elements of a protein. A peptide of greater than three amino-acids in length will be described as a protein. A graphical molecular structure will be created for all peptides of 15 residues or less.

3.1.43
protein sequence
the order and identity of amino acids within a protein subunit

NOTE Protein sequences will be represented by a single letter Dayhoff codes and listed from the N-terminus to the C-terminus.

3.1.44
protein subunit
linear sequence of amino acid residues connected through peptide bonds. Subunits, if repeated in a protein will be captured

EXAMPLE Monoclonal antibodies will typically consist of four subunits

3.1.45
resultant material
material that is the result of a critical process

NOTE Resultant material may be the starting material of the next process or the final material or actual specified substance.

3.1.46
salt
ionic substances formed from the neutralization reaction of an acid and base. Salts are ionic compounds composed of cations (positive ions) and anions (negative ions)

3.1.47
single substance
substance that can be described by a single representation or set of descriptive elements

NOTE Racemates, and substances with unknown or mixed chirality can be defined as single substances because a single structural representation can be generated

3.1.48
solvate
substance formed through association of a solvent molecule (i.e. water, alcohol) with another moieity. Solvates can be either stoichiometric or non-stoichiometric and are predominately present in the solid form of substances

3.1.49
source material
material from which a substance is derived. Source material shall be defined based on taxonomic and anatomical origins

NOTE This class is used to define structurally diverse and polymer substances isolated from biological matrices.

3.1.50
specified substance
concept to further specify substances or describe intermediate products

NOTE Information needed to further specify a substance. This could include grade, units of measure, physical form, constituents, manufacturer, and critical manufacturing processes (i.e. extraction, synthetic, recombinant processes).

3.1.51
starting material
material that is modified through a manufacturing process

3.1.52
stoichiometric
substances that contains moieties in simple integral ratios. 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 Chemicals have defined stoichiometry when the ration of all moieties, (ion, counter ion and solvate) can be represented as simple integral ratios.

3.1.53
stereochemistry
relative spatial arrangement of atoms within molecules

3.1.54
structurally diverse
substance that cannot be defined as a chemical, protein, nucleic acid, or polymer. Structurally diverse substances are defined using a set of elements based on immutable properties of a given material. Modifications that irreversibly alter the structure of the material, distinctive physical properties or are added to a material (gene in gene therapy substances) are captured in the definition of structurally diverse substance. Fractions derived from source material (oils and juices) are also captured in the definition

NOTE For materials of biological origin this typically would be the family, genus, species, and part from which a material is derived.

3.1.55
substance
any matter that has discrete existence, whose origin may be biological, mineral or chemical. Substances can be either single substances or mixture substances. Single substances shall be defined using a minimally

sufficient set of data elements divided into five types, chemical, protein, nucleic acid, polymer, and structurally diverse. Substances may be salts, solvates, free acids, free bases, mixtures of related compounds that are either isolated or synthesized together. Pharmacopeial terminology and defining characteristics will be used when available appropriate. Defining elements are dependent on the type of substance.

NOTE Discrete existence refers to the ability of a substance to exist independently of any other substance. Substances can either be well-defined entities containing definite chemical structures, synthetic (i.e. isomeric mixtures) or naturally-occurring (i.e. conjugated estrogens) mixtures of chemicals containing definite molecular structures, or materials derived from plants, animals, microorganisms or inorganic matrices for which the chemical structure may be unknown or difficult to define. Substances may be salts, solvates, free acids, free bases, mixtures of related compounds that are either isolated or synthesized together.

3.1.56
substituent
molecular fragment attached to a structural repeat unit of a polymer that typically replaces a hydrogen atom

NOTE This information shall be captured as part of the structural repeat unit when the position of substitution is fully occupied. When occupancy of a site is incomplete, the amount of a substitiuent or substituents shall be specified.

3.1.57
tautomers
chemical substances with the same elemental composition that are capable of facile interconversion between two or more structural forms

NOTE Typically interconversion involves the migration of a hydrogen atom between two adjacent atoms (i.e. keto-enol). It is anticipated that a single tautomeric form will be associated with each substance and detailed rules will be developed within the implementation guide to indicate the tautomeric form associated with each chemical substance.

3.1.58
taxonomy
information class needed to describe the origin of source material in substances isolated from biological matrices. Taxonomic information is captured to the species level for all substances, if such information is available and the source material is consistently derived from the species. Intraspecific information (i.e. subspecies, strain, or variety) shall be captured when there are consistent differences in either content or function.

3.2 Abbreviated terms

3.2.1
ACS
American Chemical Society

3.2.2
BAN
British Approved Name

3.2.3
EP
European Pharmacopiea

3.2.4
INCI
International Nomenclature of Cosmetic Ingredients

3.2.5
INN
International Nonproprietary Name, also known as rINN, recommended International Nonproprietary Name or pINN, proposed International Nonproprietary Name

3.2.6
JAN
Japanese approved name

3.2.7
JP
Japanese Pharmacopeia

3.2.8
UCUM
Unified Code for Units of Measure

3.2.9
USAN
United States Adopted Names

3.2.10
USP
United States Pharmacopeia

4 Requirements

4.1 Concepts required for the unique identification and description of substances

The standard defines the concepts required for the unique identification of substances at an international level, whenever such recognition is required. Such identification shall be based on the following principles:

- A Substance shall generally be defined based on what something is and not on how it is made or used;
- A Substance shall be defined based on immutable properties independent of physical form, grade or level or purity;
- Substances can be single molecular entities or mixtures of single molecular entities either synthesised or isolated together;
- To avoid ambiguity and facilitate implementation a mixture shall be defined as a combination of single substances;
- Substances shall not be diverse material brought together to form a medicinal or intermediate product.

EXAMPLE Simethicone would not be a substance since it consists of dimethicone and silicon dioxide which are diverse material brought together to form a product.

- Complex materials from biological matrices that cannot be defined by a limited number of chemical structures will be defined based on taxonomic, anatomical and component type;
- Materials containing interactions of an indefinite nature and indefinite stoichiometry shall not be defined as substances.

NOTE Materials that contain moieties that interact with polymers, complex matrices, or cyclodextrins will typically not be defined as substances. Simple polymeric salts such as sodium polystyrene sulfonate would be defined as a single substance.

EXAMPLE nicotine polacrilex will be defined as two distinct substances; nicotine and polacrilex; Human insulin isophane would also be defined as two distinct substances protamine and human insulin. Nicotine polacrilex and human insulin isophane however could be defined as single specified substances. Liposomal doxorubicin would be defined as a specified substance that contains doxorubicin and the components that make up the liposome.

Substances shall be defined based on one or more of the following types:

- Chemical
- Protein
- Nucleic Acid
- Polymer
- Structurally Diverse

All types of substances shall have the ability to capture established names, synonyms, isotopic information, reference information.

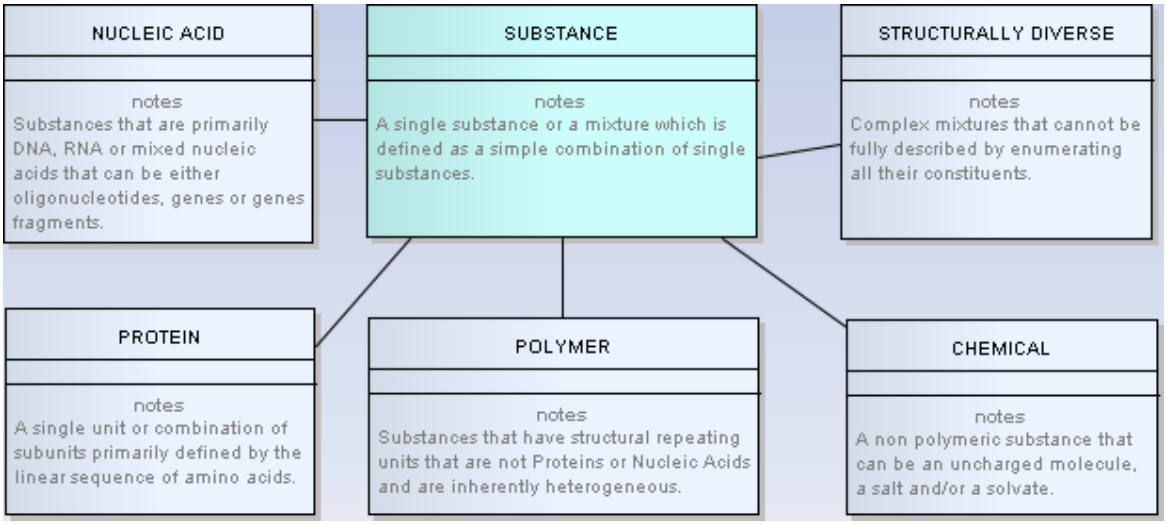


Figure 1 High level information model of substances

4.2 Concepts required for the description of specified substances

A specified substance shall capture more detailed characteristics of single substances.

A specified substance shall define intermediate products that contain multiple substances or specified substances of diverse origin.

Characteristics include: physical form; constituents (components, marker substances, additives, impurities, degradents); grade; manufacturing information, including information manufacturer, critical production processes; additional properties both physical and biological; units of measure and the analytical method used to measure strength or potency.

4.3 Naming of Substances

At least one name or company code shall be associated with each substance.

If the name is an established name, the registration authority, the language and jurisdiction in which the name is used shall be identified.

This standard shall not prescribe or favour any official or systematic nomenclature nor provide any means for the development of nomenclature.

NOTE It is anticipated that every substance will have a name in English. Additional synonyms can be associated with a substance. Translations of English names to other language can also be provided. Language and jurisdiction will be described using official ISO terminology.

The information model for the class name is shown in Figure 2.

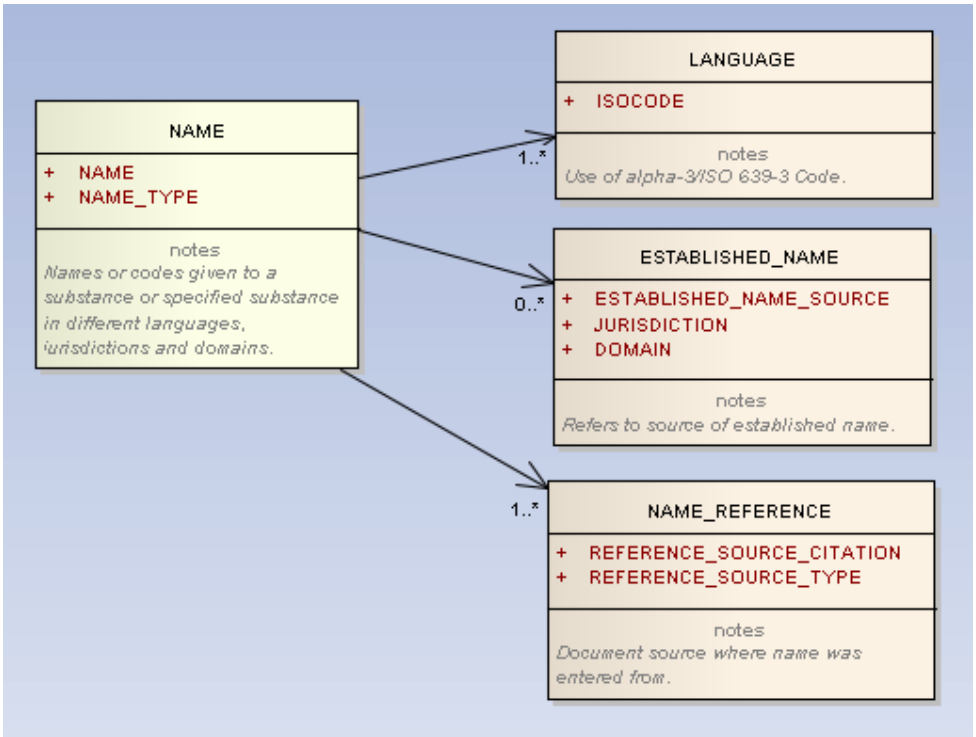


Figure 2 Information model for substance names

4.4 Requirements for Unique Identifier

Each substance and specified substance shall have only one permanently associated unique identifier that shall not indicate the order of submission to the system.

The unique identifier shall be non-semantic, random, fixed length with an internal integrity check.

The unique identifiers shall be publicly available and their use royalty free.

A unique identifier shall be assigned to both approved and investigational substances.

NOTE 1 A variety of chemical and biological nomenclature systems have been developed that describe the pharmacological actions of drugs. Functional naming systems such as INN or USAN have a great deal of value in either describing molecular structure or the biological actions of a substance. However, a unique identifier based on such classification systems would result in greater maintenance requirements since any classification scheme requires a great deal of expertise and controlled terminology.

NOTE 2 Once a substance is defined and assigned a unique identifier it is essential that this identifier is permanently associated with the substance. A substance shall only have one unique identifier. This will necessitate the generation of detailed rules to define substances that will be presented in an implementation guide.

NOTE 3 A major purpose of the unique identifier is its use in electronic data systems. An identifier of fixed length with an internal integrity check would facilitate the use of the identifier and help identify errors that may occur in data systems that use the identifier.

The information model for the relationship between names, unique identifier, substance and specified substance is shown in Figure 2.

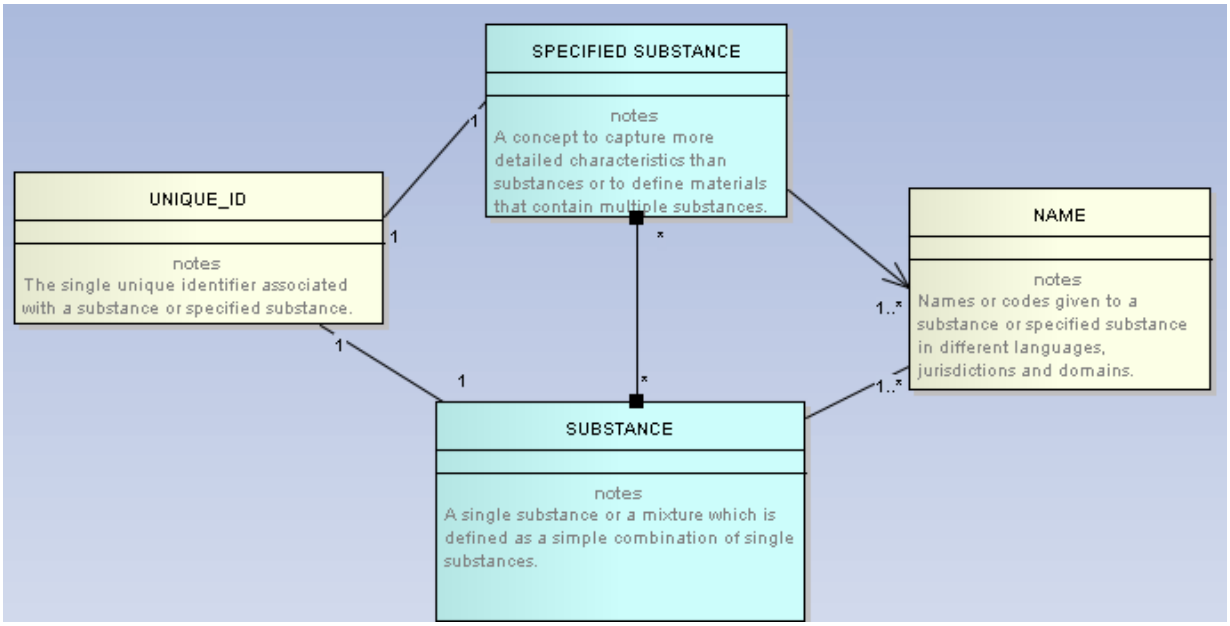


Figure 3 Information model for substance, specified substances, names and identifiers

4.5 Types of Substances

4.5.1 Mixture substance

Mixtures substances shall be described as simple combinations of single substances that are either isolated together or are the result of the same synthetic process. For mixtures derived from natural sources, the source material from which the mixture was derived should be identified.

Mixture substances shall not be combinations of diverse material brought together to form a product.

There shall be three types of mixture substance: “All Of” in which all of the single substances are required to be present; “Any Of” in which one or more of the single substances are required to be present; “One Of” in which only one of the single substances is present. The standard shall indicate if a single substance is always required to be present.

The relative amount of each single substance shall not be captured and the standard shall not allow mixture substances of mixture substances.

If it is possible to represent a substance as a single substance or a mixture substance the substance shall be represented as a single substance.

NOTE Racemic substances will be represented as single substances since they can be represented with a single structural diagram and distinguished from chiral substances.

4.5.2 Element Sets Common to Multiple Types of Substances

4.5.2.1 Isotope

Radionuclides, and other non-naturally abundant nuclides shall be defined as isotopes and associated with characteristics using a controlled terminology derived from an internationally recognized reference source.

The presence of isotopes shall also be indicated in structural representations.

Radiopharmaceuticals shall be defined based on the type of substance of the underlying pharmaceutical.

NOTE Characteristics for each nuclide could include half-life, type and energy of emission, parent and daughter nuclides.

EXAMPLE Yttrium Y 90 Iritumomab Tiuxetan will be described as a protein substance. Thyroxine I-131 would be described as a chemical substance.

4.5.2.2 Modification

Irreversible changes in the underlying molecular structure of protein, nucleic acid, or structurally diverse substances shall be described as a modification of the antecedent material. The substituent entity in polymers shall be used to capture the modifications of most polymeric material.

Modification of a chemical substance shall result in a new chemical substance.

NOTE Modifications of a chemical substance are inherently captured in the structural representation.

Irreversible changes in underlying structure of polymers shall only be captured using modification elements when other elements are not sufficient to describe the polymer. The substituent group shall be used to capture most modifications of an underlying polymer.

EXAMPLE Methyl cellulose will be described with the methyl group as substituent and not using the elements of modification. Process modifications such as thermal curing could be captured as modifications.Modifications shall be represented as the addition of moieties or molecular fragments to the underlying material when definitive structural modifications occur.

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

The information model for the class modification group is shown in Figure 4.

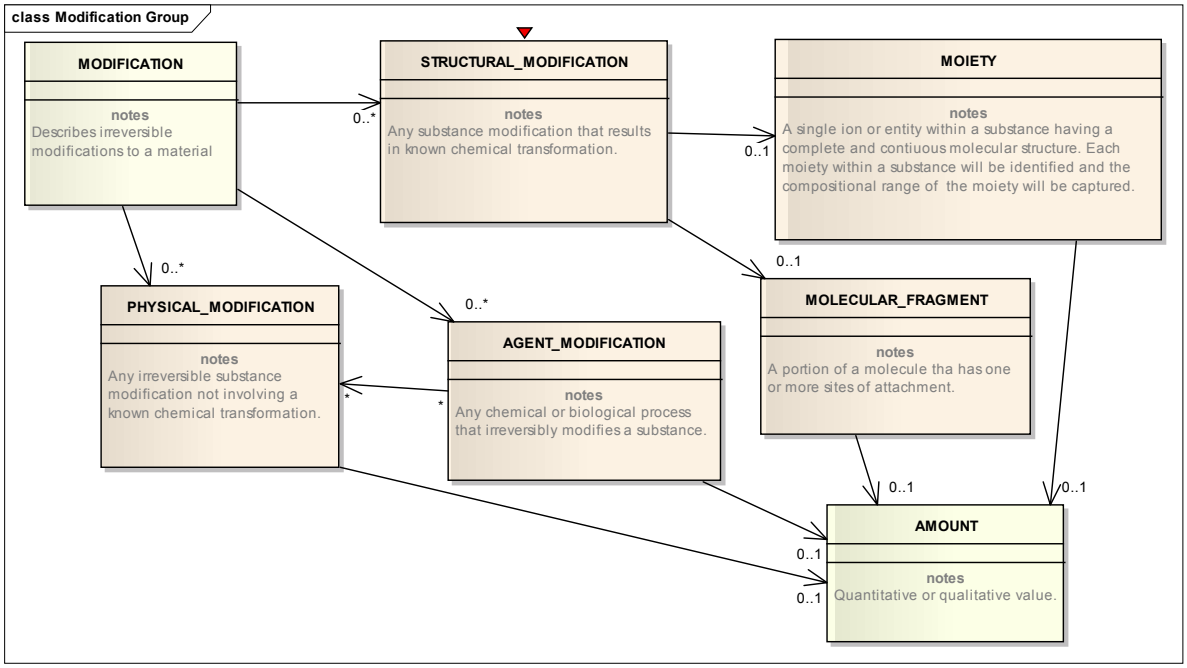


Figure 4 Information model for modifications

4.5.2.3 Reference Information

Additional types of informative reference information may be captured for each type of substance. Such information may include both classification and target information. This standard shall not provide any guidance on the classification of pharmacological effects or the determination of the putative targets for any substance or specified substance. The standard shall allow for the capture of such information if provided or available. This information shall not effect the generation of a new unique identifier.

NOTE Classification systems such as the WHO ATC and the United States Veterans Administration NDF-RT which code classification information to substances are particularly important. Target information is important for monoclonal and polyclonal antibodies and small molecules directed against specific molecular targets.

4.5.2.4 Source Material

Source material is an information class that shall capture 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 class shall be used to define structurally diverse and polymer substances isolated from biological matrices.

Taxonomic and anatomical origins shall be described using controlled vocabulary as required.

The information model for the class source material is shown in Figure 5.

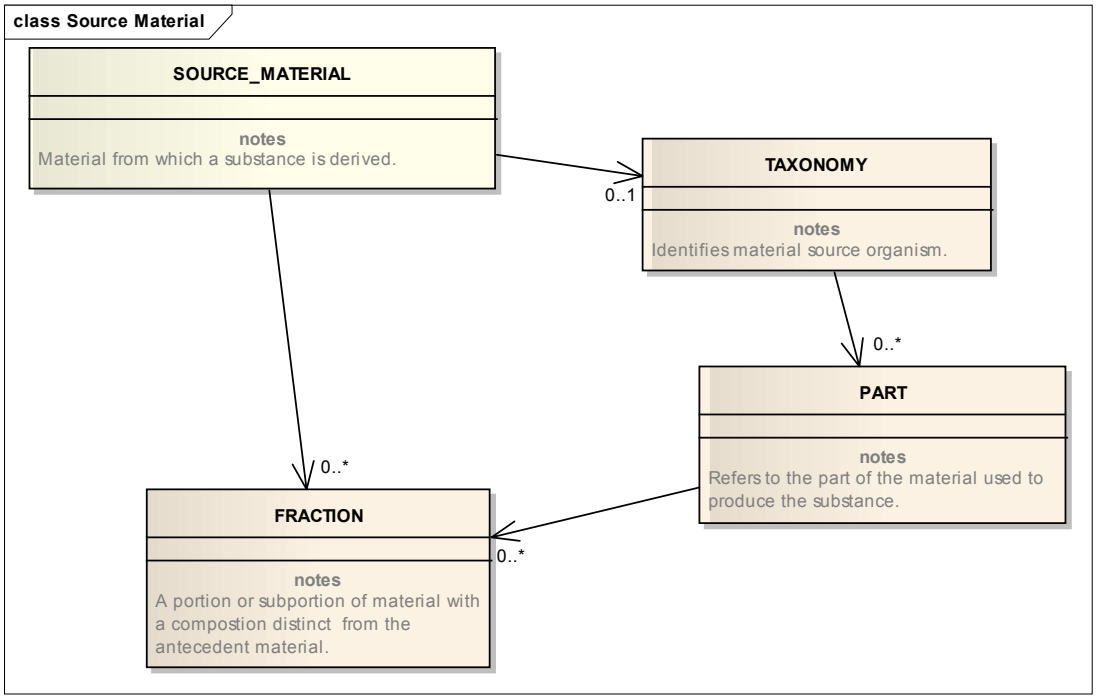


Figure 5 Information model for source material

4.5.2.5 Structure

Structure shall contain a sufficient amount of graphical and textual information to define the underlying atoms and the connectivity between atoms as well as the composition ratio of moieties.

4.5.3 Chemical

Chemical substances shall be defined by a representation of the molecular structure and when necessary stereochemical and related physical characteristics.

Each chemical substance shall be associated with a single structural representation.

Stereochemistry shall be completely defined when known. If not known, positions where stereochemistry is unknown shall be clearly identified.

Underlying the graphical representation of the structure shall be a textual format that indicates the atoms and the connectivity between atoms that represent a molecular structure.

Fixed and variable stoichiometric ratios of moieties within a substance shall be captured. For substances that have moieties with variable stoichiometry the range of composition shall be captured.

Unknown stoichiometry of a given moiety or moieties shall also be clearly identified. Physical properties shall only be used to define single substances that have variable or unknown stoichiometry.

Isotopes shall be described in the structural representation; the specific position or positions of substitution shall be provided if known. Substances shall be defined independent of the specific activity of a given radioisotope.

The information model for the class chemical substance is shown in Figure 6.

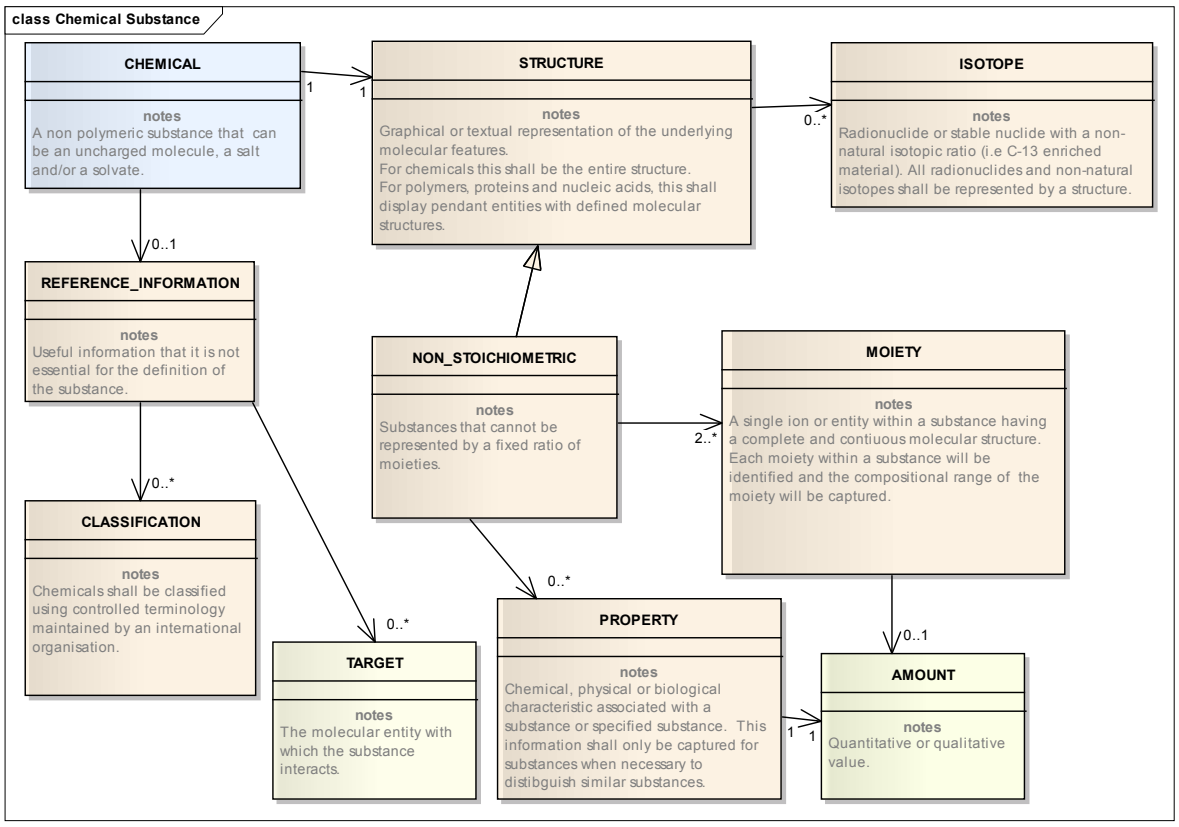


Figure 6 Information model for the chemical substance class

4.5.4 Protein

Proteins that differ in protein sequence, type of glycosylation, disulfide bonds or glycosylation site shall be defined as two separate substances.

Proteins shall consist of one or more protein subunits.

All non-glycosylated proteins shall be defined without regard to the method of synthesis, the cell line or organism from which the protein was produced or isolated from.

Proteins shall be described without regard to microheterogeneity.

Type of glycosylation shall reflect significant differences in overall glycosylation and is determined from the species of the cell or tissue from which the protein was isolated.

A limited set of controlled terminology shall be used to describe the type of glycosylation.

Proteins shall be defined by the final expressed sequence; preproproteins and proproteins shall not be described.

Proteins that are irreversibly modified by either chemical or physical processes shall be defined as different proteins.

The description of modified proteins shall capture the structural change that results from the modification when a definitive structure is known.

Modifications shall be described using either moieties or molecular fragments that are added to the proteins structure or by a description of the modification process if a definitive structural modification does not occur.

The molecular fragment may have a functional role and that role shall be captured using controlled terminology maintained by an international organization.

For specific modifications the site and residue modified shall be defined.

Post-translational modifications shall only be captured if they are essential for activity or present on the predominant forms of the proteins.

In some instances the modification will not result in a definitive structure, in these instances the modification process shall be described in a minimal manner capturing the modifying agent or physical conditions that result in an irreversible change. Purified blood or tissue materials whose putative functionality is attributed to a protein or a limited number of proteins with distinct and known amino acid sequences will be described as a protein.

Non-covalent interactions between proteins or peptide chains shall not be captured with the exception of protein chains that are tightly associated with well-defined stoichiometry.

Non-defining reference information can also be captured, including: type and subtype of protein; ligand, substrate, or target; and type of interaction of the protein; gene from which the protein was derived. Controlled vocabulary shall be used for each of the

Reference information shall be captured using controlled vocabulary as required.

NOTE 1 Monoclonal immunoglobulins are described as proteins. Somatropin, a non-glycosylated protein that can be produced in *E.coli*, yeast or mammalian cells is defined as the same single substance regardless of the cell line it was produced in.

NOTE 2 The current types of glycosylation, include yeast, fungal, plant, insect, avian, mammalian, and human.

NOTE 3 Differences in even a single amino acid would result in two distinct substances i.e. interferon alfa 2a, and interferon alfa 2b will be defined as separate substances since the sequences differ by a single amino acid. Aggregated human serum albumin which formed by irreversible partial physical denaturation would be defined as a separate substance from human serum albumin.

The information model for the class protein substance is shown in Figure 7.

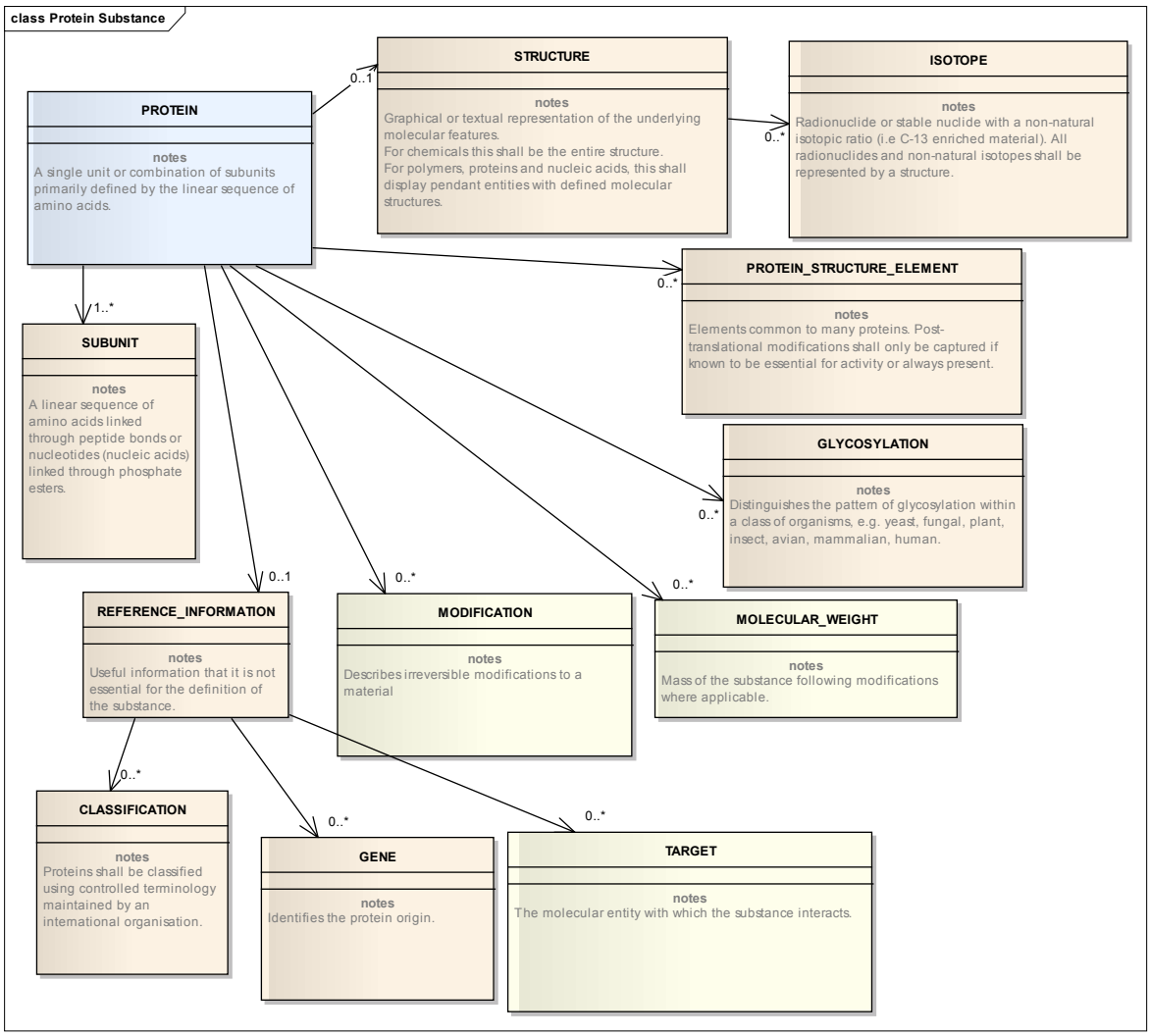


Figure 7 Protein Substances Class

4.5.5 Nucleic acid

The sequence of the nucleic acid, the type (RNA, DNA, mixed) together with any modifications that effect the molecular structure shall be the defining elements for nucleic acids.

Genes, plasmids, and the nucleic acid portion of viral vectors used in gene therapy shall also be described using the nucleic acid schema.

Individual gene elements shall be captured and defined as substances.

Irreversible modifications either physical or chemical that affect the molecular structure shall be described.

For gene therapy, the entire sequence of the transforming/transducing vector shall be used as the defining element. Each gene element that comprises the sequence would also be captured.

NOTE Gene elements would include promoters, enhancers, silencers etc. For nucleic acids used in gene therapy the entire sequence of the transforming/transducing entity would be captured along and each gene element.

The information model for nucleic acid class is shown in Figure 8.

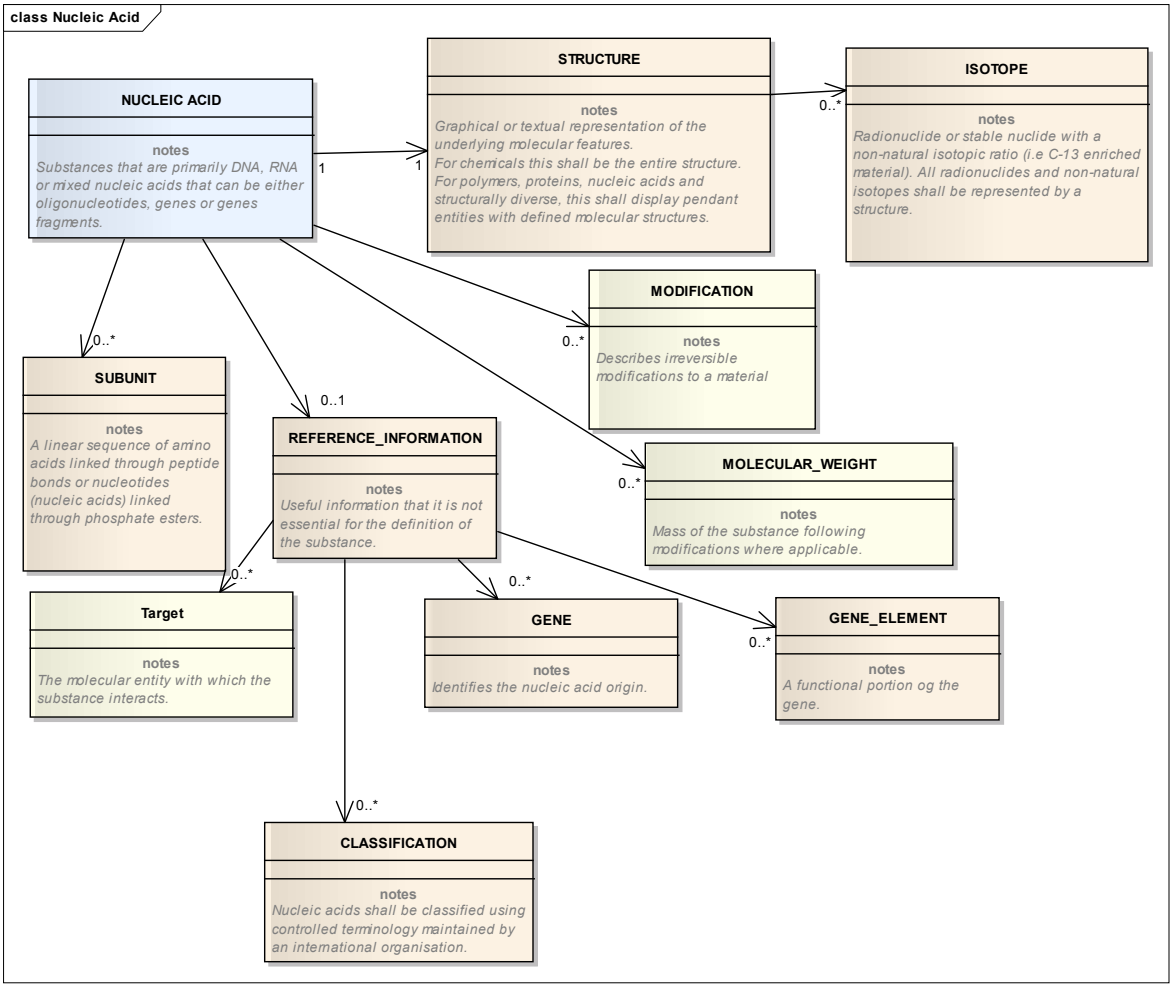


Figure 8 Information model for the nucleic acid substance class

4.5.6 Polymer

Polymers shall refer to material that is inherently heterogeneous and contain structural repeating unit.

Polymers shall be defined using a combination of the molecular structure of the structural repeating units, substituents that are attached to the structural repeating unit, molecular weight, and polydispersity of the polymer substance. Degree of polymerization, monomers used to synthesize synthetic polymers or copolymers, the source material for naturally derived polymers, polymeric end groups, and physical or biological properties shall also be captured when known and needed to distinguish material. Polymers shall be defined to the level of specificity needed to distinguish materials and broad polymeric definitions shall be disfavored.

EXAMPLE Polymers containing polyethylene glycol structural repeating units should always be defined based on either degree of polymerization or molecular weight. A generic polyethylene glycol substance should not be defined as a substance, because of the wide variation in material and safety concerns related to the degree of polymerization of such materials.

Substituents shall be captured as part of the structural repeat unit when the position of substitution is fully occupied. When occupancy of an attachment site on the structural repeat unit, is incomplete, the substituent shall be represented as a molecular fragment and the amount of each substitiuent shall specified. Sites of attachments for both molecular fragments and the structural repeating unit shall be clearly identified.

The polymer type shall be defined by the number of structural repeating units and the connectivity between them. A controlled vocabulary will be developed as required to describe the polymer type.

Physical and biological properties shall only be a defining element if they are necessary to distinguish polymeric substances from one another and are related to the underlying molecular structures of the polymeric ensemble.

NOTE Values for polymer type would include homopolymer, homopolymer branched, copolymer random, copolymer block, and copolymer branched. Polydispersity is usually determined from the ratio of the weight average molecular weight to the number average molecular weight. Properties such as viscosity, sedimentation velocity, enzyme inhibition . Biological properties such a ratio of enzymatic inhibition (Factor IIa/Factor X) for different low molecular weight heparins may also be a distinguishing property.

The information model for polymer substance is shown in Figure 9.

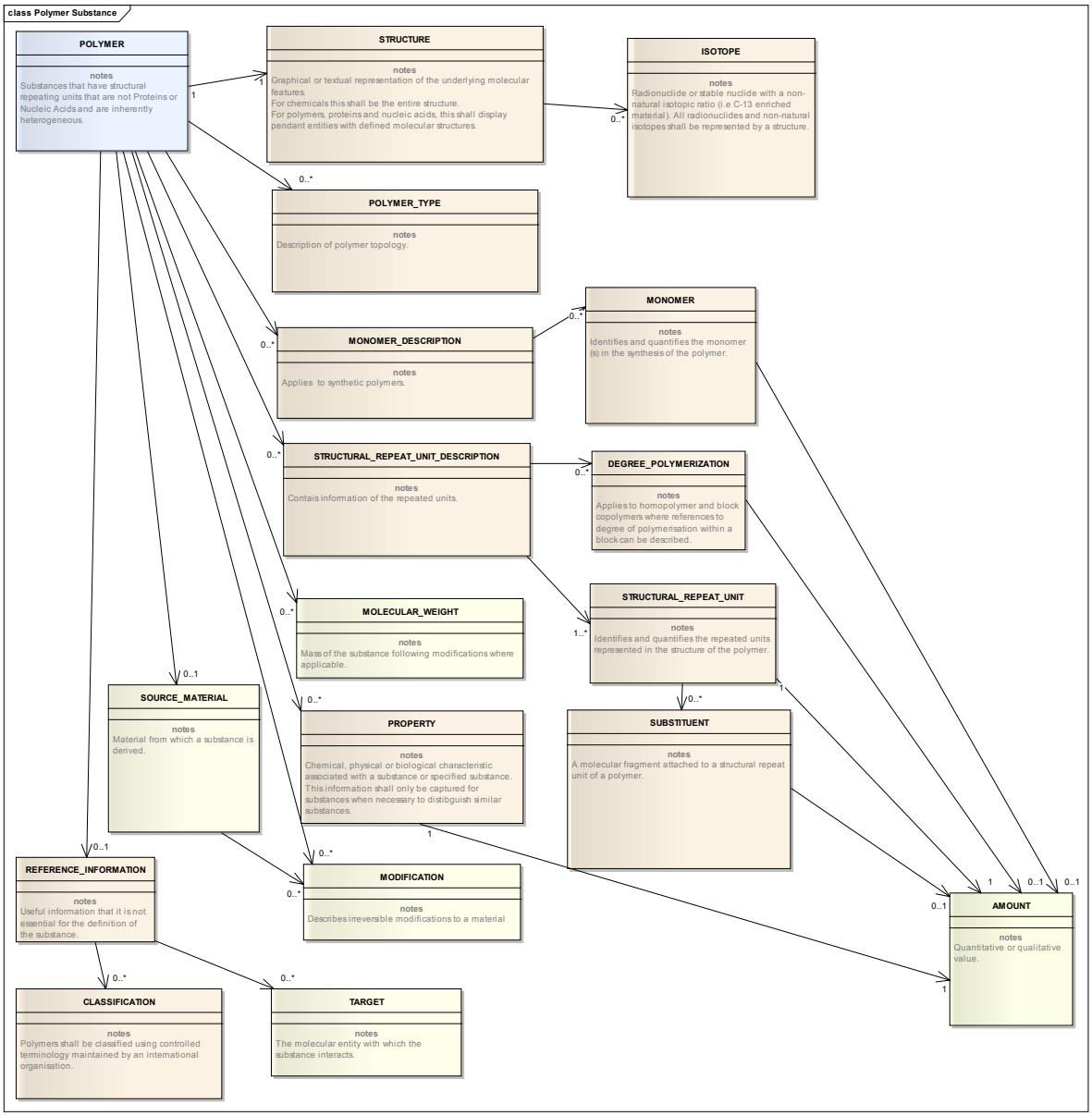


Figure 9 Information model for the polymer substance class

4.5.7 Structurally Diverse

Substances that are not chemicals, polymers, proteins, or nucleic acids shall be described as structurally diverse substances. Structurally diverse substances shall be defined by the source material the substance is

derived from, modifications that result in irreversible changes in the underlying material, and physical or biological properties related to underlying molecular composition of the material.

The majority of structurally diverse substances are derived from a biological organism. They may also be complex natural materials such as coal tar or mineral oil. For organism-based substances, the parent organism from which the source material was derived shall be essential information. Parent organisms shall be defined to at least the species level. Varieties, cultivars, strains or sub-strains of biological material shall be defining if intraspecific differences are distinct and reflect consistent differences in functionality or composition.

NOTE For organism-based substances, the parent organism is essential defining information. Herbals are typically described by parent organism genus, species, and part or parts. If specific parts of a plant are used, identification requires lists of individual parts such as flower, leaf, and stem or an indication of the plant life cycle segment such as flowering top. Because of variability in extraction processes (solvent, temperature, time), time and place of harvest, type of soil and fertilizer, amount of daylight and water, and degree of plant maturity are also not captured at the substance level. A cultivar or variety of a plant may be defined as a different substance if the known differences in constituents or functionality. Other organisms, typically bacteria and viruses require identification of subspecies, variety, strain, serovar type, or to accurately describe them and distinguish them from related substances.

EXAMPLE Broccoli and broccoli extract would be defined as the same substance. Broccoli and cauliflower which differ at the cultivar level are defined as different substances even though they share the same genus and species because there dramatic differences in appearance and constituents. Influenza viruses will be defined at a level that will allow the distinction of various vaccine strains.

Polyclonal immunoglobulins are described as structurally diverse materials and require identification of the immunoglobulin type and targeted antigen. 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.

Commodity oils, juices, and exudates of plants and such are separate substances. Oils and juices shall be described as a fraction, the part necessary to accurately describe the source of the oil or juice will be captured unless the entire organism is used.

EXAMPLE Olive Oil is a fraction of the *Olea Europaea* fruit. Orange juice is *Citrus sinensis* fruit juice. Green tea and green tea extracts will be defined as the leaves of *Camellia sinensis*.

Because of the variable nature of the resultant material and processes (i.e. time, temperature, solvent) used to prepare extracts, tinctures, infusions, and decoctions shall not be defined as separate substances but will map to the parent organism and part from which they were derived.

The information model for the structurally diverse substance class is shown in Figure 10.

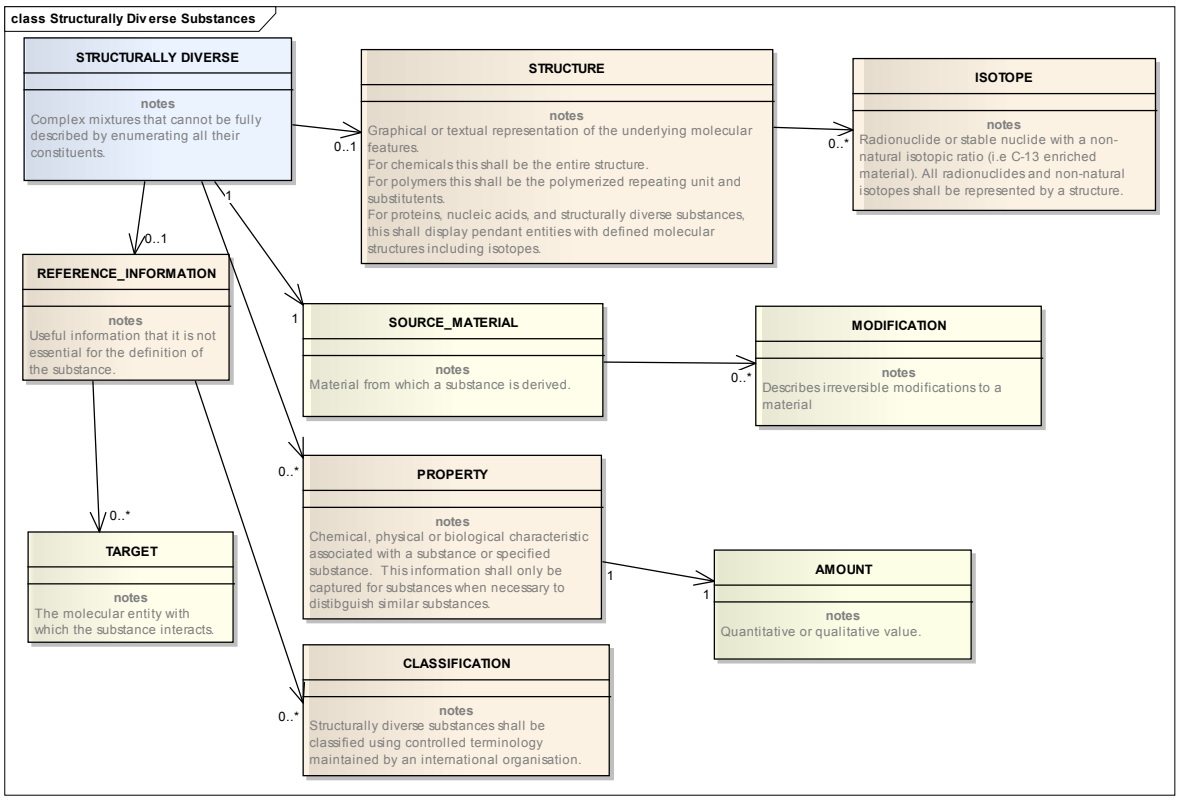


Figure 10 Information model for the structurally diverse substance class

4.6 Defining Specified Substances

4.6.1 General

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. Specified substance provides a general information model that shall be used to further define materials present in medicinal products.

The specified substance shall play two roles in the description of medicinal products. One role is providing elements and a structured format for the capture of additional information on single substances. The second is to describe material of diverse origin that is brought together to form an intermediate product. There shall be only one set of elements used to describe specified substances. It shall not be necessary to specify information in all the elements to generate a specified substance.

The specified substance shall be organized to capture diverse information in consistent manner. This information shall include: purity or grade; manufacturer data including information on the manufacturer and critical processes in manufacturing, analytical data, constituents including the amounts and role of each constituents, specifications (impurites and related substance limits would be captured using constituents), unitage, reference material and analytical methods used for potency determination.

NOTE Information described at the specified is often submitted in regulatory submissions in an unorganized and diffuse manner without a common format.

4.6.2 Analytical Data

Information on assays used to determine potency can be defined as a group of elements. This information shall reference an analytical method and any reference material used in the determination of potency. Individual reference materials shall be defined as specified substances. Different analytical methods or

reference materials used for the determination of potency shall be described, captured and referenced. This standard shall not ascribe that any analytical method or reference material be used for the determination of potency but merely provides a mechanism for the capture of such information.

NOTE Unitage for potency is often dependent on the analytical method and reference material used in the determination. In many instances, unitage can vary across jurisdictions or even among manufacturers within a jurisdiction.

EXAMPLE USP Pancrelipase Units and FIP Pancrelipase Units differ and are not readily convertible because the reference materials are distinct and standardized in a different manner.

4.6.3 Constituent

An element group that shall capture the substances and specified substances related to a given specified substance. Intermediate products shall contain substances and specified substances defined as constituents. The role component will be given to any substance which is an intended constituent of the specified substances. Active marker substances in herbal extracts, impurities, or related substances shall also be constituents and can be linked to the components that make up a specified substance. The amount and role of each constituent shall be captured. Amounts shall either be expressed as definitive amounts (average) limit or ranges of concentration.

NOTE The constituent group provides a great deal of flexibility and utility in that it can link related substances together .

4.6.4 Grade

Grade shall refer to the overall quality or group of specifications a given specified substance is compliant with.

NOTE Pharmacopeial grades or specifications will be referred to when available. A given specified substance could be compliant with specifications from multiple pharmacopeias. Technical grades could also be indicated.

EXAMPLE Pharmacopeial grades or specifications will be referred to when available. The specified substance would distinguish each grade of water from each other. Tap water, USP Sterile Water for Injection, EP Sterile Water for Injection, USP Purified Water, et al. would each be defined as separate specified substances.

4.6.5 Manufacturing

The manufacturing element group shall capture information on the manufacturer and critical manufacturing processes that are necessary to distinguish specified substances. The starting materials, processing material, critical process parameters, equipment used, and the resultant material from the manufacturing process can be captured within this element group.

NOTE The manufacturing group is not intended to capture all the details of manufacturing but only the critical processes that could impact the safety or efficacy of a specified substance used in a medicinal product.

EXAMPLE Cell culture is a critical manufacturing process for material derived using recombinant DNA technology. The cell in which the material is produced is a critical starting material and the growth media in which cells are grown may be a critical processing material that need to be captured.

4.6.6 Physical Form

Physical form shall capture the state in which the specified substance exists. Controlled vocabulary will be used and developed as required.

4.6.7 Property

Physical, chemical or biological properties shall be captured that can be related to the use or function of a specified substances and may be independent of the molecular structure or molecular ensemble that makes up the substances or specified substances within a specified substance. This element group may be used to capture properties related to the supramolecular organization of materials..

4.6.8 Intermediate products

The specified substance shall capture information to define intermediate products. These products shall contain multiple substances and specified substances of diverse origin. The substances and specified substances that make up intermediate products shall be defined as constituents and the amounts and role of each of the constituents will be captured. The role component shall specify that a constituent is an intended substance or specified substances in the formulation of an intermediate product. Intermediate products that differ in the relative amounts of individual constituents would be described as separate specified substances. Many elements in the HL-7 Common Product Model shall be used to capture information on specified substances.

EXAMPLE Aluminum lakes, adjuvated vaccine antigens, isophane insulin and liposomal doxorubicin are intermediate products that would be defined as specified substances.

An information model for specified substances is shown in Figure 11.

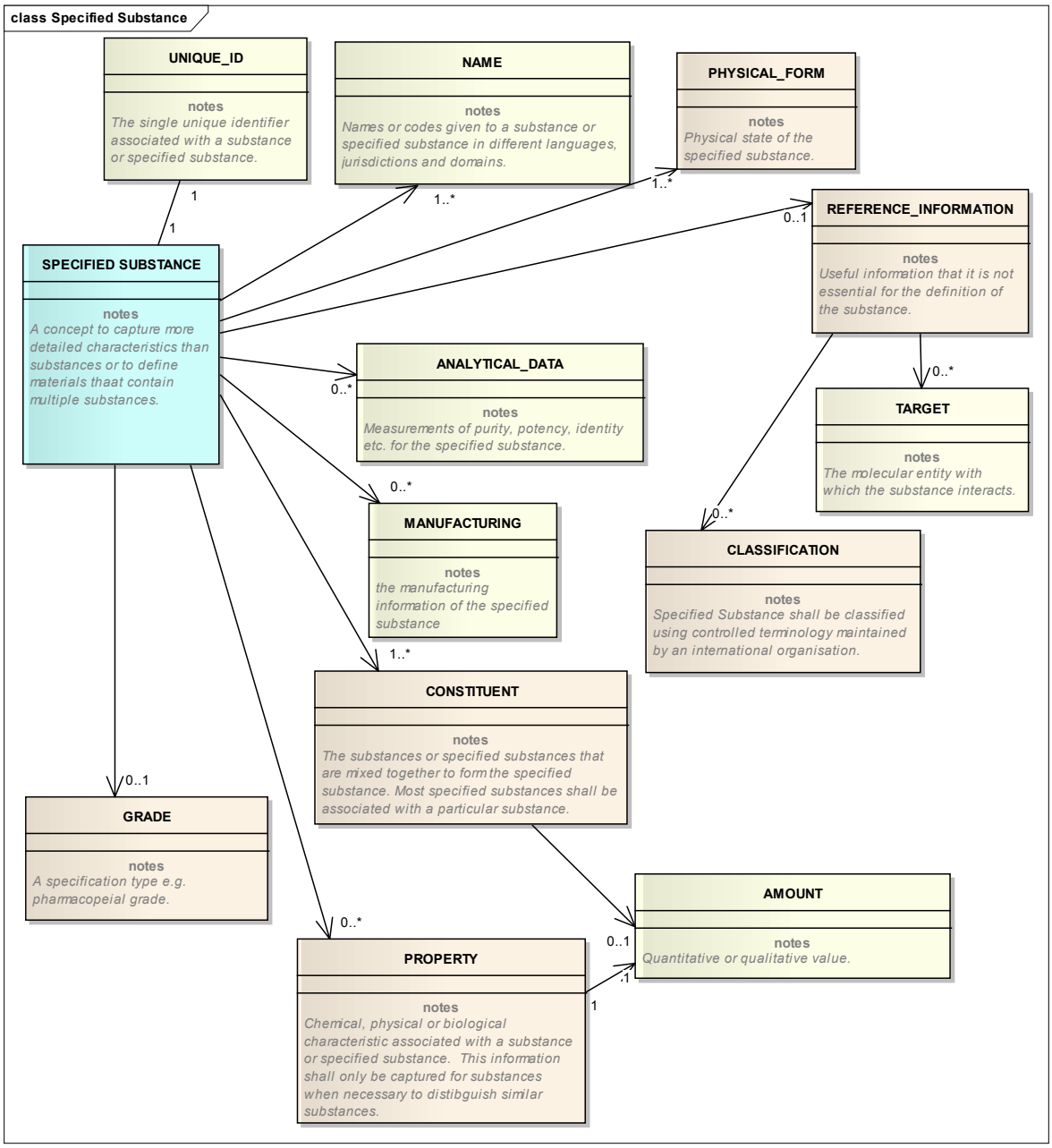


Figure 11 Information model for the specified substance class

Annex A
(informative)

Existing Identifiers and Molecular Structure Representations

A.1 Identifiers

A.1.1 General

Below are descriptions of commonly used identifiers emphasizing the strengths and weaknesses of each identifier for use as a unique identifier for substances in pharmaceutical products. The list is not exhaustive but describes identifiers that are actively being used in data systems.

A.1.2 CAS Registry Numbers

CAS Registry Numbers are numeric identifiers that usually only identify a single substance. Polymers frequently only have one CAS registry number associated with them regardless of difference in molecular mass or other defining elements. 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. There are currently over 33 million organic and inorganic substances and 59 million sequences in the CAS registry system. The primary purpose of the CAS registry system is to link to information in the chemical literature and not necessarily identify or define substances. The CAS registry system is maintained by the Chemical Abstracts Service of the American Chemical Society. Although CAS numbers are widely used they cannot be freely used. CAS has guidelines on the use of CAS registry numbers and has attempted to restrict their use in publicly available databases. The CAS number for formaldehyde is 50-00-0.

A.1.3 InChI and InChIKey

InChI stands for IUPAC International Chemical Identifier. The system was primarily developed at the National Institute of Standards and Technology in the USA. InChI is a linear identifier that deals with chemical representation using a layered approach. InChI is non-proprietary and the software necessary to generate InChI's are provided under an open-source LGPL license. An InChIKey is a fixed length (25 characters) condensed digital representation of the InChI. InChI and InChIKey is really only designed for simple substances that can be defined by a representation of molecular structure and not complex products such as vaccines, blood-derived, botanicals or animal products. The InChI for morphine is InChI=1/C17H19NO3/c1-18-7-6-17-10-3-5-13(20)16(17)21-15-12(19)4-2-9(14(15)17)8-11(10)18/h2-5,10-11,13,16,19-20H,6-8H2,1H3/t10-,11-,13-,16-,17-/m0/s1 and the InChIKey for morphine is BQJCRHHNABKAKU-XKUOQXLYBY.

A.1.4 EC Number

The EC-No. or EC# is a seven digit code that has been allocated by the European Commission for all commercially available substances marketed within the European Union. The seventh digit of the code is a check digit and the code maps to both common and trade names of a given substance. The scope of the EC number is broader that that of InChi in that both simple and complex substances have been assigned EC#'s. The system contains over 100000 substances but is not heavily weighted in the pharmaceutical sector. The codes are also for the most part sequential and were developed from the EINECS (European Inventory of Existing Commercial Chemical Substances), ELINCS (European List of Notified Chemical Substances), and other list of regulated substances. The EC# for formaldehyde is 200-001-8.

A.1.5 UNII

The UNII is a 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, 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. The integrity check on the UNII is stronger than both the EC# and the CAS Registry Number because of the random generation from large number of potential UNIIs and the fact that there are 36 possible check characters compared to 10 with both the EC# and CAS Registry Number. The UNII is freely available for use and there is a mechanism where a manufacturer can petition for the generation of a UNII through the FDA. The system has the capability for both public and restricted access to information, and can be adapted to produce ingredient identifiers.

A.1.6 ASK Number

The ASK Number is a five digit code (and check digit) and is issued and maintained by the German National Competent Authorities (BfArM, PEI, BVL), based on §10 German Drug Law and AMIS-Bezeichnungsverordnung, respectively. The ASK Number is mandatory for applications and correspondence between marketing authorisation holders and competent authorities. The underlying substances database comprises more than 35000 substances which are related to business in the regulatory environment. These are substances of chemical or biological origin, as well as radiopharmaceuticals, homeopathics and anthroposophics. The repository contains mainly active ingredients and excipients, but also gases, packaging materials, chemicals for analysis, impurities, and substances prohibited by law. In addition to the chemical name according to IUPAC a large “collection” of synonyms of international and European sources and all over the lifecycle of a medicinal product are referenced. If applicable, the CAS-Registry Number, molecular formula and molecular mass are available. Related to the different aspects in the daily work of the regulatory authorities, extensive “grouping attributes” have been included for classifying the substances.

A.2 Molecular Structure Representations

A.2.1 General

Representation of the chemical or molecular structure is essential to the development of a controlled vocabulary for simple chemical structures. The system of representation should be both unambiguous and unique. Only one single representation will be allowed for a given structure, and the representation should have enough detail to ensure that unintended ambiguity does not exist. The representation or a form of them should be capable of being stored in a chemical database to facilitate registration and searching. There are other formats that

A.2.2 Molfile

The molfile format was predominantly developed by MDL Information Systems. There are currently two versions in use today V2000 and the extended molfile format V3000. The extended molfile format has enhanced stereochemistry descriptors, that allow relative, unknown, and racemic designations to be associated with each chiral atom. The V2000 format is widely used and interconversion between it and other formats can readily occur. Unlike other representations, the molfile format is not a linear representation but is predominantly tabular. Below is the molfile representation

```
benzene

 ACD/Labs0812062058



  6  6  0  0  0  0  0  0  0  0  1 V2000

    1.9050   -0.7932    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
```

```
  1.9050   -2.1232    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
  0.7531   -0.1282    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
  0.7531   -2.7882    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
 -0.3987   -0.7932    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
 -0.3987   -2.1232    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
 2  1  1  0  0  0  0
 3  1  2  0  0  0  0
 4  2  2  0  0  0  0
 5  3  1  0  0  0  0
 6  4  1  0  0  0  0
 6  5  2  0  0  0  0
M END
$$$$
```

A.2.3 SMILES

Simplified molecular input line entry specification (SMILES) is a specification for a unambiguous linear representation of chemical or molecular structures using ASCI characters. It is predominantly used by Daylight Chemical Information Systems Inc. although an open source version has been recently developed. Canonical smiles is a smiles string that is unique for each structure and can be used to ensure that duplicate structures are not entered into a database. Other linear representation forms for chemical structures include SYBYL line notation (SLN) and the older Wiswesser Line Notation which was the first line notation for the representation of chemical structures. These other formats are not

SMILES representation of Benzene

C1=CC=CC=C1

A.2.4 InChI

Is described above. It is a layered approach to chemical structure representation. There are currently four layers under development

constitutional - expresses pure connectivity of the atoms

stereochemical - includes conventional C-atom sp2 and sp3 stereochemistry

isotopic - enables isotopes to be distinguished

tautomeric - implements simple forms of rapid H-migration isomerization

The InChI for benzene is:

1/C6H6 /c1-2-4-6-5-3-1/h1-6H

Annex B
(informative)

Examples of Substances and Specified Substances

B.1 Substance and Specified Substance Models

B.1.1 Potential Detailed Substance Model in XML

<SUBSTANCE>
<SUBSTANCE_ID>**SUBSTANCE UNIQUE IDENTIFIER**</SUBSTANCE_ID>
<NAME_GROUP>
<NAME>**NAME OF THE SUBSTANCE**</NAME>
<NAME_TYPE>**CV: CODE, BRAND NAME, ESTABLISHED NAME, PRIMARY NAME, OTHER NAME. ESTABLISHED NAME REFERS TO A NON-PROPRIETARY NAME THAT IS ASSIGNED TO A SUBSTANCE BY A SCIENTIFIC ORGANIZATION AND/OR REGULATORY AGENCY (AUTHORITY SOURCE). PRIMARY NAME REFERS TO A NAME IS COMMON USE ASSOCIATED WITH A SUBSTANCE WHERE NO ESTABLISHED NAME IS AVAILABLE. CODES WILL BE ENTERED AS NAMES WHEN THE CODE SYSTEM IS PROPRIETARY. ALL SUBSTANCES WILL HAVE AT LEAST ONE NAME.**</NAME_TYPE>
<LANGUAGE_ISOCODE>**ISO 639-2 ALPHA-3 CODES**</LANGUAGE_ISOCODE>
<ESTABLISHED_NAME_GROUP>
<ESTABLISHED_NAME_TYPE>**CV: USAN, INN, JAN, BAN, USP, EP, INCI, JECFA**</ESTABLISHED_NAME_TYPE>
<ESTABLISHED_NAME_STATUS>**CV: CURRENT, ALTERNATE, SUPERCEDED**</ESTABLISHED_NAME_STATUS>
<ESTABLISHED_NAME_STATUS_CHANGE_DATE>**ISO 8601 DATE FORMAT**</ESTABLISHED_NAME_STATUS_CHANGE_DATE>
<ESTABLISHED_NAME_JURISDICTION_GROUP>
<ESTABLISHED_NAME_JURISDICTION>**EUROPEAN COMMISSION USAGE: ISO 3166-1 ALPHA-2 CODES WITH TWO EXCEPTIONS: EL (NOT GR) IS USED TO REPRESENT GREECE, AND UK (NOT GB) IS USED TO REPRESENT THE UNITED KINGDOM**</ESTABLISHED_NAME_JURISDICTION>
</ESTABLISHED_NAME_JURISDICTION_GROUP>
<ESTABLISHED_NAME_DOMAIN>**MULTIPLE VALUES - CV: FOOD, DRUG, COSMETIC, BIOLOGIC**</ESTABLISHED_NAME_DOMAIN>
</ESTABLISHED_NAME_GROUP>
<REFERENCE_SOURCE_GROUP>
<REFERENCE_SOURCE_CLASS>**CV: PUBLIC, REGULATORY SUBMISSION**</REFERENCE_SOURCE_CLASS>
<REFERENCE_SOURCE_TYPE>**CV: NDA, IND, BLA, ETC.**</REFERENCE_SOURCE_TYPE>
<REFERENCE_SOURCE_CITATION>**LITERATURE OR REGULATORY DOCUMENT SPECIFIC REFERENCE**</REFERENCE_SOURCE_CITATION>
</REFERENCE_SOURCE_GROUP>
</NAME_GROUP>
<CODE_GROUP>
<CODE>**SPECIFIC CODE WITHIN AN AVAILABLE CODE SYSTEM**</CODE>
<CODE_SYSTEM>**CV: CAS, EINECS, HPUS, NSC, HPUS, ASK, FCC, ATC, NDF-RT ETC.**
</CODE_SYSTEM>
<CODE_SYSTEM_ID>**OID FOR THE CODE SYSTEM**</CODE_SYSTEM_ID>
<CODE_STATUS>**CV: CURRENT, ALTERNATE, SUPERCEDED**</CODE_STATUS>

<CODE_STATUS_CHANGE_DATE>**ISO 8601 DATE FORMAT**</CODE_STATUS_CHANGE_DATE>
<CODE_REFERENCE_CITATION>**LITERATURE OR REGULATORY DOCUMENT SPECIFIC REFERENCE**</CODE_REFERENCE_CITATION>
</CODE_GROUP>
<SINGLE_SUBSTANCE>
<STRUCTURE>
<STRUCTURAL_REPRESENTATION_TYPE>**CV: MOL, INCHI, SMILES**</STRUCTURAL_REPRESENTATION_TYPE>
<STRUCTURAL_REPRESENTATION>**REPRESENTATION OF STRUCTURE ACCORDING TO THE STRUCTURAL REPRESENTATION TYPE**</STRUCTURAL_REPRESENTATION>
<STEREOCHEMISTRY>**CV: RACEMIC; MESO; MIXED; UNKNOWN; AXIAL, ABSOLUTE; ACHIRAL (CIS/TRANS OR GEOMETRIC ISOMERISM INDICATED IN STRUCTURE)**</STEREOCHEMISTRY>
<OPTICAL_ACTIVITY>**CV: (+), (-), (+/-) CAN ONLY DEFINING FOR SUBSTANCES WHEN SUBSTANCE IS KNOWN TO BE CHIRAL BUT ABSOLUTE CONFIGURATION IS UNKNOWN**</OPTICAL_ACTIVITY>
<ISOTOPE_GROUP>
<NUCLIDE_ID>**UNIQUE IDENTIFIER FOR EACH NON-NATURAL OR RADIONUCLIDE**</NUCLIDE_ID>
<NUCLIDE_NAME>**EG. C-13 (STABLE ISOTOPE)) , C-14 (RADIOACTIVE ISOTOPE)**</NUCLIDE_NAME>
<SUBSTITUTION_TYPE>**CV: SPECIFIC (SITE OF ATTACHMENT/SUBSTITUTION INDICATED IN STRUCTURE); NON-SPECIFIC (NUCLIDE DISTRIBUTED THROUGHOUT MOLECULE OR SUBSTANCE); UNKNOWN (SITE UNKNOWN); EXTENT OF SUBSTITUTION NOT CAPTURED AT SUBSTANCE LEVEL**</SUBSTITUTION_TYPE>
</ISOTOPE_GROUP>
</STRUCTURE>
<REFERENCE_INFORMATION>
<CLASSIFICATION_GROUP>
<SUBSTANCE_CLASS>
CV: CHEMICAL, PROTEIN, NUCLEIC ACID, POLYMER, STRUCTURALLY DIVERSE, MULTIPLE
<TYPE>
CV: DIFFERENT VALUES FOR EACH SUBSTANCE CLASS GENERAL CLASSIFICATION OF THE SUBSTANCE CV TO BE DEVELOPED
<SUBTYPE>**CV: DIFFERENT VALUES FOR EACH SUBSTANCE CLASS SPECIFIC CLASSIFICATION OF THE SUBSTANCE** </SUBTYPE>
</TYPE>
<SUBSTANCE_CLASS_CODE>**CV: CLASSIFICATION CODE, ATC, NDF_RT ETC.**</SUBSTANCE_CLASS_CODE>
<CLASSIFICATION_SOURCE>**CV: SOURCE OF CLASSIFICATION**</CLASSIFICATION_SOURCE>
</SUBSTANCE_CLASS>
</CLASSIFICATION_GROUP>
<TARGET_GROUP>
<TARGET_GENE_ID>**NUMERIC OR ALPHA NUMERIC ID ASSOCIATED WITH THE HUMAN GENE OF ORIGIN**</TARGET_GENE_ID>
<TARGET_NAME>**TARGET OF A GIVEN SUBSTANCE OR SUBSTRATE UPON WHICH THE SUBSTANCE ACTS**</TARGET_NAME>
<INTERACTION_TYPE>**CV: SUBSTRATE, INHIBITION, BINDING, ACTIVATION, ETC.**</INTERACTION_TYPE>
<TARGET_ORGANISM_TYPE>**CV: HUMAN, BACTERIAL, VIRAL, ETC. THE ORGANISM TYPE FOR WHICH THE ACTIVE SUBSTANCE IS TARGETED** </TARGET_ORGANISM_TYPE>
<TARGET_REFERENCE_SOURCE>**SOURCE WHICH INDICATED THE TARGET OF THE SUBSTANCE (PRIMARY OR REPUTABLE SECONDARY SOURCES WOULD SUFFICE)**</TARGET_REFERENCE_SOURCE>

```
    </TARGET_GROUP>
<GENE_GROUP>
 <GENE_SEQUENCE_ORIGIN>COMMON NAME FOR THE ORGANISM FROM WHICH THE FINAL
    SEQUENCE ORIGINATED. </GENE_SEQUENCE_ORIGIN>
 <GENE_ID>NUMERIC OR ALPHANUMERIC ID ASSOCIATED WITH THE GENE OF
    ORIGIN</GENE_ID>
 <GENE_NAME>COMPLETE NAME GIVEN FOR A GENE</GENE_NAME>
 <GENE_REFERENCE_SOURCE>SOURCE FROM WHICH THE GENE SEQUENCE WAS DERIVED
    (GENBANK)</GENE_REFERENCE_SOURCE>
   </GENE_GROUP>
 <COMMENTS/>
   </REFERENCE_INFORMATION>
<CHEMICAL_SUBSTANCE>
 <STOICHIOMETRIC>YES/NO IF NO, USE NON-STOICHIOMETRIC
    SCHEMA</STOICHIOMETRIC>
<NON_STOICHIOMETRIC>
 <NUMBER_OF_MOIETIES>NON-STOICHIOMETRIC CHEMICAL SUBSTANCES MUST HAVE AT
    LEAST TWO MOIETIES. EACH MOIETY WILL BE REPRESENTED BY A CHEMICAL
    STRUCTURE AND ENCLOSED IN A LABELED BRACKET</NUMBER_OF_MOIETIES>
 <MOIETY_AMOUNT_TYPE>CV: MOLE RATIO, WEIGHT PERCENT</MOIETY_AMOUNT_TYPE>
<MOIETY_GROUP>
 <MOIETY_NAME>NAME OF MOIETY</MOIETY_NAME>
 <MOIETY_ID>A,B,C ETC REFERS TO LABEL ON BRACKET OF THE STRUCTURAL
    REPRESENTATION</MOIETY_ID>
<AMOUNT>
 <AVERAGE>AVERAGE AMOUNT OR SINGLE VALUE</AVERAGE>
 <LOW_LIMIT>LOWER LIMIT IF GIVEN </LOW_LIMIT>
 <HIGH_LIMIT>HIGHER LIMIT IF GIVEN </HIGH_LIMIT>
 <UNIT>CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE</UNIT>
 <NON_NUMERIC_VALUE>CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE
    VALUE </NON_NUMERIC_VALUE>
   </AMOUNT>
   </MOIETY_GROUP>
<PROPERTY_GROUP>
 <PROPERTY_TYPE>CV: PHYSICAL, CHEMICAL, ENZYMATIC,
    IMMUNOLOGICAL</PROPERTY_TYPE>
 <PROPERTY_NAME>CV: VISCOSITY, DENSITY, PH, ENZYMATIC ACTIVITY
    ETC</PROPERTY_NAME>
 <SUBSTANCE_ID>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 POSTIVE CELLS)</SUBSTANCE_ID>
<AMOUNT>
 <AVERAGE>AVERAGE AMOUNT OR SINGLE VALUE</AVERAGE>
 <LOW_LIMIT>LOWER LIMIT IF GIVEN </LOW_LIMIT>
 <HIGH_LIMIT>HIGHER LIMIT IF GIVEN </HIGH_LIMIT>
 <UNIT>CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE</UNIT>
 <NON_NUMERIC_VALUE>CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE
    VALUE </NON_NUMERIC_VALUE>
   </AMOUNT>
   </PROPERTY_GROUP>
   </NON_STOICHIOMETRIC>
 <COMMENTS>INFORMATION NECESSARY TO DISTINGUISH SUBSTANCES OR ESSENTIAL
    TO THE DEFINITION OF A SUBSTANCE SHOULD BE USED SPARINGLY</COMMENTS>
   </CHEMICAL_SUBSTANCE>
```

<PROTEIN>
<SEQUENCE_TYPE>**CV: COMPLETE, PARTIAL PROTEIN DESCRIPTIVE ELEMENTS WILL ONLY BE USED WHEN A CONMPLETE OR PARTIAL AMINO ACID SEQUENCE IS AVAILABLE OR DERIVABLE FROM A NUCLEIC ACID SEQUENCE**</SEQUENCE_TYPE>
<NUMBER_OF_SUBUNITS>**NUMBER OF LINEAR SEQUENCES OF AMINO ACIDS LINKED THROUGH PEPTIDE BONDS**</NUMBER_OF_SUBUNITS>
<SUBUNIT_GROUP>
<SUBUNIT>**1,2,3,.. RELATES BACK TO SUBUNIT 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.**</SUBUNIT>
<LENGTH>**NUMBER OF AMINO ACID RESIDUES LINKED THROUGH PEPTIDE BONDS**</LENGTH>
<SEQUENCE>**SEQUENCES WILL BE DESCRIBED USING THE STANDARD SINGLE LETTER AMINO ACID CODE. TRANSCRIBED PROTEINS WILL ALWAYS BE DESCRIBED USING THE TRANSLATED SEQUENCE. FOR SYNTHETIC PEPTIDES CONTAINING AMINO ACIDS THAT ARE NOT REPRESENTED WITH A SINGLE LETTER CODE AN X WILL BE USED WITHIN THE SEQUENCE.**</SEQUENCE>
<N_TERMINAL_MODIFICATION>**CV: ACETYL, FORMYL, ETC.**</N_TERMINAL_MODIFICATION>
<N_TERMINAL_MODIFICATION_ID>**UNIQUE IDENTIFIER FOR MOLECULAR FRAGMENT MODIFICATION**</N_TERMINAL_MODIFICATION_ID>
<C_TERMINAL_MODIFICATION>**CV: AMIDE, ETHYL ESTER ETC.**</C_TERMINAL_MODIFICATION>
<C_TERMINAL_MODIFICATION_ID>**UNIQUE IDENTIFIER FOR MOLECULAR FRAGMENT MODIFICATION**</C_TERMINAL_MODIFICATION_ID>
</SUBUNIT_GROUP>
<DISULFIDE_LINKAGE>**POSITION OF THE DISULFIDE BONDS IN THE PROTEIN LISTED IN INCREASING ORDER OF SUBUNIT NUMBER AND POSITION WITHIN SUBUNIT**</DISULFIDE_LINKAGE>
<GLYCOSYLATION_GROUP>
<GLYCOSYLATION_TYPE>**CV: HUMAN, MAMMALIAN, AVIAN, FUNGAL, BACTERIAL, VIRAL, PLANT, INSECT ETC. HOST SYSTEM FROM WHICH THE PROTEIN WAS ISOLATED**</GLYCOSYLATION_TYPE>
<N_GLYCOSYLATION>**SITE OF N-GLYCOSYLATION (ASPARAGINE); N-GLYCOSYLATION IS LISTED ACCORDING TO THE PROTEIN SEQUENCE. THE EXTENT AND TYPE OF MODIFICATION AT A GIVEN SITE IS NOT CAPTURED.**</N_GLYCOSYLATION>
<O_GLYCOSYLATION>**SITE OF O-GLYCOSYLATION (SERINE, THREONINE, ETC)**</O_GLYCOSYLATION>
<C_GLYCOSYLATION>**SITE OF C-GLYCOSYLATION (TRYPTOPHAN)**</C_GLYCOSYLATION>
</GLYCOSYLATION_GROUP>
<MODIFICATION_GROUP>
<MODIFICATION_TYPE>**CV: PHYSICAL, STRUCTURAL, AGENT**</MODIFICATION_TYPE>
<RESIDUE_MODIFIED>**CV: 20 AMINO ACIDS PLUS 5 NUCLEOTIDES**</RESIDUE_MODIFIED>
<RESIDUE_SITE>**POSITION OF SPECIFIC MODIFICATIONS (IE. 1_10 REFERS TO THE 10TH RESIDUE ON THE 1ST SUBUNIT)**</RESIDUE_SITE>
<PHYSICAL_MODIFICATION>
<ROLE>**CV: INACTIVATION, ACTIVATION, ETC.**</ROLE>
<NUMBER_OF_PARAMETERS>**NUMBER OF PARAMETER GROUPS NEEDED TO DESCRIBE THE PHYSICAL MODIFICATION**</NUMBER_OF_PARAMETERS>
<PARAMETER_GROUP>
<PARAMETER>**CV: TIME, TEMPERATURE, ETC.**</PARAMETER>
<AMOUNT>
<AVERAGE>**AVERAGE AMOUNT OR SINGLE VALUE**</AVERAGE>
<LOW_LIMIT>**LOWER LIMIT IF GIVEN** </LOW_LIMIT>
<HIGH_LIMIT>**HIGHER LIMIT IF GIVEN** </HIGH_LIMIT>

```
 <UNIT>CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE</UNIT>
 <NON_NUMERIC_VALUE>CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE
     VALUE </NON_NUMERIC_VALUE>
   </AMOUNT>
   </PARAMETER_GROUP>
   </PHYSICAL_MODIFICATION>
<STRUCTURAL_MODIFICATION>
 <STRUCTURAL_MODIFICATION_TYPE>CV: MOIETY, MOLECULAR
     FRAGMENT</STRUCTURAL_MODIFICATION_TYPE>
<MOIETY>
 <ROLE>CV: ANTIGEN, LINKER, CONJUGATE, ETC.</ROLE>
 <MOIETY_ID>UNIQUE IDENTIFIER FOR MODIFYING MOIETY</MOIETY_ID>
 <AMOUNT_TYPE>CV: MOLE RATIO; WEIGHT AVERAGE</AMOUNT_TYPE>
<AMOUNT>
 <AVERAGE>AVERAGE AMOUNT OR SINGLE VALUE</AVERAGE>
 <LOW_LIMIT>LOWER LIMIT IF GIVEN </LOW_LIMIT>
 <HIGH_LIMIT>HIGHER LIMIT IF GIVEN </HIGH_LIMIT>
 <UNIT>CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE</UNIT>
 <NON_NUMERIC_VALUE>CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE
     VALUE </NON_NUMERIC_VALUE>
   </AMOUNT>
   </MOIETY>
<MOLECULAR_FRAGMENT>
 <ROLE>CV: ANTIGEN, LINKER, CONJUGATE, ETC.</ROLE>
 <MOLECULAR_FRAGMENT_ID>UNIQUE IDENTIFIER FOR MOLECULAR
     FRAGMENT</MOLECULAR_FRAGMENT_ID>
 <AMOUNT_TYPE>CV: MOLE RATIO; WEIGHT AVERAGE</AMOUNT_TYPE>
<AMOUNT>
 <AVERAGE>AVERAGE AMOUNT OR SINGLE VALUE</AVERAGE>
 <LOW_LIMIT>LOWER LIMIT IF GIVEN </LOW_LIMIT>
 <HIGH_LIMIT>HIGHER LIMIT IF GIVEN </HIGH_LIMIT>
 <UNIT>CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE</UNIT>
 <NON_NUMERIC_VALUE>CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE
     VALUE </NON_NUMERIC_VALUE>
   </AMOUNT>
   </MOLECULAR_FRAGMENT>
   </STRUCTURAL_MODIFICATION>
<AGENT_MODIFICATION>
 <AGENT_MODIFICATION_TYPE>CV: CHEMICAL, ENZYMATIC, IMMUNOLOGICAL,
     ORGANISMAL</AGENT_MODIFICATION_TYPE>
 <ROLE>CV: INACTIVATION, ACTIVATION, PEGYLATION, PHOSPHORYLATION,
     ETC.</ROLE>
 <MODIFICATION_AGENT>ESTABLISHED OR PRIMARY NAME OF MODIFYING
     AGENT</MODIFICATION_AGENT>
 <MODIFICATION_AGENT_ID>UNIQUE IDENTIFIER FOR MODIFYING
     AGENT</MODIFICATION_AGENT_ID>
 <MODIFICATION_PROCESS>CV: CELL CULTURE, UV IRRADIATION,
     ETC.</MODIFICATION_PROCESS>
<AMOUNT>
 <AVERAGE>AVERAGE AMOUNT OR SINGLE VALUE</AVERAGE>
 <LOW_LIMIT>LOWER LIMIT IF GIVEN </LOW_LIMIT>
 <HIGH_LIMIT>HIGHER LIMIT IF GIVEN </HIGH_LIMIT>
 <UNIT>CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE</UNIT>
```

```
  <NON_NUMERIC_VALUE>CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE
     VALUE </NON_NUMERIC_VALUE>
    </AMOUNT>
    </AGENT_MODIFICATION>
    </MODIFICATION_GROUP>
<MOLECULAR_WEIGHT>
  <MOLECULAR_WEIGHT_TYPE>CV: EXACT, AVERAGE (AKA. NUMBER AVERAGE), WEIGHT
     AVERAGE, ETC.</MOLECULAR_WEIGHT_TYPE>
  <MOLECULAR_WEIGHT_METHOD>CV: SDS_PAGE, CALCULATED, LIGHT SCATTERING
     VISCOSITY, GEL PERMEATION, ETC.</MOLECULAR_WEIGHT_METHOD>
<AMOUNT>
  <AVERAGE>AVERAGE AMOUNT OR SINGLE VALUE</AVERAGE>
  <LOW_LIMIT>LOWER LIMIT IF GIVEN </LOW_LIMIT>
  <HIGH_LIMIT>HIGHER LIMIT IF GIVEN </HIGH_LIMIT>
  <UNIT>CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE</UNIT>
  <NON_NUMERIC_VALUE>CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE
     VALUE </NON_NUMERIC_VALUE>
    </AMOUNT>
    </MOLECULAR_WEIGHT>
  <COMMENTS>INFORMATION NECESSARY TO DISTINGUISH SUBSTANCES OR ESSENTIAL
     TO THE DEFINITION OF A SUBSTANCE SHOULD BE USED SPARINGLY</COMMENTS>
    </PROTEIN>
<NUCLEIC_ACID>
  <SEQUENCE_TYPE>CV: COMPLETE, PARTIAL PROTEIN DESCRIPTIVE ELEMENTS WILL ONLY
     BE USED WHEN A CONMPLETE OR PARTIAL AMINO ACID SEQUENCE IS AVAILABLE OR
     DERIVABLE FROM A NUCLEIC ACID SEQUENCE</SEQUENCE_TYPE>
  <NUMBER_OF_SUBUNITS>NUMBER OF STRANDS THAT UP THE NUCLEIC ACID. IN MOST
     GENE EXAMPLES ONLY THE TRANSCRIBED STRAND WILL BE
     DESCRIBED.</NUMBER_OF_SUBUNITS>
<SUBUNIT_GROUP>
  <SUBUNIT>1,2,3,...NUMBER IN ORDER OF DECREASING SEQUENCE LENGTH</SUBUNIT>
  <LENGTH>NUMBER OF NUCLEOTIDE RESIDUES LINKED THROUGH PHOSPHATE
     ESTERS</LENGTH>
  <SEQUENCE>ACTUAL NUCLEOTIDE SEQUENCE</SEQUENCE>
    </SUBUNIT_GROUP>
<MODIFICATION_GROUP>
  <MODIFICATION_TYPE>CV: PHYSICAL, STRUCTURAL, AGENT</MODIFICATION_TYPE>
  <RESIDUE_MODIFIED>CV: 20 AMINO ACIDS PLUS 5 NUCLEOTIDES</RESIDUE_MODIFIED>
  <RESIDUE_SITE>POSITION OF SPECIFIC MODIFICATIONS (IE. 1_10 REFERS TO THE 10TH
     RESIDUE ON THE 1ST SUBUNIT)</RESIDUE_SITE>
<PHYSICAL_MODIFICATION>
  <ROLE>CV: INACTIVATION, ACTIVATION, ETC.</ROLE>
  <NUMBER_OF_PARAMETERS>NUMBER OF PARAMETER GROUPS NEEDED TO DESCRIBE THE
     PHYSICAL MODIFICATION</NUMBER_OF_PARAMETERS>
<PARAMETER_GROUP>
  <PARAMETER>CV: TIME, TEMPERATURE, ETC.</PARAMETER>
<AMOUNT>
  <AVERAGE>AVERAGE AMOUNT OR SINGLE VALUE</AVERAGE>
  <LOW_LIMIT>LOWER LIMIT IF GIVEN </LOW_LIMIT>
  <HIGH_LIMIT>HIGHER LIMIT IF GIVEN </HIGH_LIMIT>
  <UNIT>CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE</UNIT>
  <NON_NUMERIC_VALUE>CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE
     VALUE </NON_NUMERIC_VALUE>
    </AMOUNT>
```

```
    </PARAMETER_GROUP>
    </PHYSICAL_MODIFICATION>
<STRUCTURAL_MODIFICATION>
 <STRUCTURAL_MODIFICATION_TYPE>CV: MOIETY, MOLECULAR
    FRAGMENT</STRUCTURAL_MODIFICATION_TYPE>
<MOIETY>
 <ROLE>CV: ANTIGEN, LINKER, CONJUGATE, ETC.</ROLE>
 <MOIETY_ID>UNIQUE IDENTIFIER FOR MODIFYING MOIETY</MOIETY_ID>
 <AMOUNT_TYPE>CV: MOLE RATIO; WEIGHT AVERAGE</AMOUNT_TYPE>
<AMOUNT>
 <AVERAGE>AVERAGE AMOUNT OR SINGLE VALUE</AVERAGE>
 <LOW_LIMIT>LOWER LIMIT IF GIVEN </LOW_LIMIT>
 <HIGH_LIMIT>HIGHER LIMIT IF GIVEN </HIGH_LIMIT>
 <UNIT>CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE</UNIT>
 <NON_NUMERIC_VALUE>CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE
    VALUE </NON_NUMERIC_VALUE>
   </AMOUNT>
   </MOIETY>
<MOLECULAR_FRAGMENT>
 <ROLE>CV: ANTIGEN, LINKER, CONJUGATE, ETC.</ROLE>
 <MOLECULAR_FRAGMENT_ID>UNIQUE IDENTIFIER FOR MOLECULAR
    FRAGMENT</MOLECULAR_FRAGMENT_ID>
 <AMOUNT_TYPE>CV: MOLE RATIO; WEIGHT AVERAGE</AMOUNT_TYPE>
<AMOUNT>
 <AVERAGE>AVERAGE AMOUNT OR SINGLE VALUE</AVERAGE>
 <LOW_LIMIT>LOWER LIMIT IF GIVEN </LOW_LIMIT>
 <HIGH_LIMIT>HIGHER LIMIT IF GIVEN </HIGH_LIMIT>
 <UNIT>CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE</UNIT>
 <NON_NUMERIC_VALUE>CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE
    VALUE </NON_NUMERIC_VALUE>
   </AMOUNT>
   </MOLECULAR_FRAGMENT>
   </STRUCTURAL_MODIFICATION>
<AGENT_MODIFICATION>
 <AGENT_MODIFICATION_TYPE>CV: CHEMICAL, ENZYMATIC, IMMUNOLOGICAL,
    ORGANISMAL</AGENT_MODIFICATION_TYPE>
 <ROLE>CV: INACTIVATION, ACTIVATION, PEGYLATION, PHOSPHORYLATION,
    ETC.</ROLE>
 <MODIFICATION_AGENT>ESTABLISHED OR PRIMARY NAME OF MODIFYING
    AGENT</MODIFICATION_AGENT>
 <MODIFICATION_AGENT_ID>UNIQUE IDENTIFIER FOR MODIFYING
    AGENT</MODIFICATION_AGENT_ID>
 <MODIFICATION_PROCESS>CV: CELL CULTURE, UV IRRADIATION,
    ETC.</MODIFICATION_PROCESS>
<AMOUNT>
 <AVERAGE>AVERAGE AMOUNT OR SINGLE VALUE</AVERAGE>
 <LOW_LIMIT>LOWER LIMIT IF GIVEN </LOW_LIMIT>
 <HIGH_LIMIT>HIGHER LIMIT IF GIVEN </HIGH_LIMIT>
 <UNIT>CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE</UNIT>
 <NON_NUMERIC_VALUE>CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE
    VALUE </NON_NUMERIC_VALUE>
   </AMOUNT>
   </AGENT_MODIFICATION>
   </MODIFICATION_GROUP>
```

```
<MOLECULAR_WEIGHT>
 <MOLECULAR_WEIGHT_TYPE>CV: EXACT, AVERAGE (AKA. NUMBER AVERAGE), WEIGHT
    AVERAGE, ETC.</MOLECULAR_WEIGHT_TYPE>
 <MOLECULAR_WEIGHT_METHOD>CV: SDS_PAGE, CALCULATED, LIGHT SCATTERING
    VISCOSITY, GEL PERMEATION, ETC.</MOLECULAR_WEIGHT_METHOD>
<AMOUNT>
 <AVERAGE>AVERAGE AMOUNT OR SINGLE VALUE</AVERAGE>
 <LOW_LIMIT>LOWER LIMIT IF GIVEN </LOW_LIMIT>
 <HIGH_LIMIT>HIGHER LIMIT IF GIVEN </HIGH_LIMIT>
 <UNIT>CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE</UNIT>
 <NON_NUMERIC_VALUE>CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE
    VALUE </NON_NUMERIC_VALUE>
   </AMOUNT>
   </MOLECULAR_WEIGHT>
<REFERENCE_INFORMATION>
<GENE_GROUP>
 <GENE_SEQUENCE_ORIGIN>COMMON NAME FOR THE ORGANISM FROM WHICH THE FINAL
    SEQUENCE ORIGINATED. </GENE_SEQUENCE_ORIGIN>
 <GENE_ID>NUMERIC OR ALPHANUMERIC ID ASSOCIATED WITH THE GENE OF
    ORIGIN</GENE_ID>
 <GENE_NAME>COMPLETE NAME GIVEN FOR A GENE</GENE_NAME>
 <GENE_REFERENCE_SOURCE>SOURCE FROM WHICH THE GENE SEQUENCE WAS DERIVED
    (GENBANK)</GENE_REFERENCE_SOURCE>
   </GENE_GROUP>
<GENE_ELEMENT_GROUP>
 <GENE_ELEMENT_TYPE>CV: ENHANCER, PROMOTER ETC.</GENE_ELEMENT_TYPE>
 <GENE_ELEMENT_ID>UNIQUE IDENTIFIER FOR GENE ELEMENT</GENE_ELEMENT_ID>
 <GENE_ELEMENT_NAME>SPECIFIC GENE ELEMENT NAME</GENE_ELEMENT_NAME>
 <GENE_ELEMENT_REFERENCE_SOURCE>SOURCE FROM WHICH THE GENE ELEMENT WAS
    DERIVED</GENE_ELEMENT_REFERENCE_SOURCE>
   </GENE_ELEMENT_GROUP>
 <COMMENTS/>
   </REFERENCE_INFORMATION>
 <COMMENTS>INFORMATION NECESSARY TO DISTINGUISH SUBSTANCES OR ESSENTIAL
    TO THE DEFINITION OF A SUBSTANCE SHOULD BE USED SPARINGLY</COMMENTS>
   </NUCLEIC_ACID>
<POLYMER>
 <POLYMER_CLASS>CV: HOMOPOLYMER, COPOLYMER</POLYMER_CLASS>
 <POLYMER_GEOMETRY>CV: LINEAR, BRANCHED, CROSSLINKED, NETWORK, DENDRITIC,
    STAR, ETC.</POLYMER_GEOMETRY>
 <COPOLYMER_SEQUENCE_TYPE>CV: RANDOM, ALTERNATING, BLOCK,
    GRAFT</COPOLYMER_SEQUENCE_TYPE>
<MONOMER_DESCRIPTION>
 <NUMBER_OF_MONOMERS>NUMBER OF KINDS OF MONOMER USED TO SYNTHESIZE THE
    POLYMER</NUMBER_OF_MONOMERS>
 <MONOMER_AMOUNT_TYPE>CV: MOLE RATIO, WEIGHT
    PERCENT</MONOMER_AMOUNT_TYPE>
<MONOMER_GROUP>
 <MONOMER_ID>UNIQUE IDENTIFIER FOR MONOMER</MONOMER_ID>
 <MONOMER_NAME>ESTABLISHED OR PRIMARY NAME OF MONOMER</MONOMER_NAME>
<AMOUNT>
 <AVERAGE>AVERAGE AMOUNT OR SINGLE VALUE</AVERAGE>
 <LOW_LIMIT>LOWER LIMIT IF GIVEN </LOW_LIMIT>
 <HIGH_LIMIT>HIGHER LIMIT IF GIVEN </HIGH_LIMIT>
```

```
 <UNIT>CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE</UNIT>
 <NON_NUMERIC_VALUE>CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE
    VALUE </NON_NUMERIC_VALUE>
   </AMOUNT>
   </MONOMER_GROUP>
   </MONOMER_DESCRIPTION>
<STRUCTURAL_REPEAT_UNIT>
 <NUMBER_OF_STRUCTURAL_REPEAT_UNITS>NUMBER OF REPEATED UNITS REPRESENTED
    IN THE STRUCTURE OF THE POLYMER</NUMBER_OF_STRUCTURAL_REPEAT_UNITS>
 <STRUCTURAL_REPEAT_UNIT_AMOUNT_TYPE>CV: MOLE RATIO; WEIGHT
    PERCENT</STRUCTURAL_REPEAT_UNIT_AMOUNT_TYPE>
<STRUCTURAL_REPEAT_UNIT_GROUP>
 <ORIENTATION_OF_POLYMERIZATION>CV: HEAD-TAIL, HEAD-HEAD,
    RANDOM</ORIENTATION_OF_POLYMERIZATION>
 <STRUCTURAL_REPEAT_UNIT>A,B,C,... RELATES BACK TO STRUCTURE IN ORDER OF
    DECREASING MOLECULAR WEIGHT</STRUCTURAL_REPEAT_UNIT>
<AMOUNT>
 <AVERAGE>AVERAGE AMOUNT OR SINGLE VALUE</AVERAGE>
 <LOW_LIMIT>LOWER LIMIT IF GIVEN </LOW_LIMIT>
 <HIGH_LIMIT>HIGHER LIMIT IF GIVEN </HIGH_LIMIT>
 <UNIT>CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE</UNIT>
 <NON_NUMERIC_VALUE>CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE
    VALUE </NON_NUMERIC_VALUE>
   </AMOUNT>
<SUBSTITUENT>
 <NUMBER_OF_SUBSTITUENTS>NUMBER OF SUBSTITUENTS ON STRUCTURAL REPEAT
    UNITS. APPLY WHEN INCOMPLETE OR PARTIAL SITE SUBSTITUTION IS PRESENT
    AND/OR POSITION OR EXTENT OF SUBSTITUTION IS NOT COMPLETELY
    KNOWN</NUMBER_OF_SUBSTITUENTS>
 <SUBSTITUTION_AMOUNT_TYPE>CV: MOLE RATIO; WEIGHT
    PERCENT</SUBSTITUTION_AMOUNT_TYPE>
<SUBSTITUENT_GROUP>
 <SUBSTITUENT>1,2,3,... RELATES BACK TO STRUCTURE IN ORDER OF DECREASING
    MOLECULAR WEIGHT;</SUBSTITUENT>
<AMOUNT>
 <AVERAGE>AVERAGE AMOUNT OR SINGLE VALUE</AVERAGE>
 <LOW_LIMIT>LOWER LIMIT IF GIVEN </LOW_LIMIT>
 <HIGH_LIMIT>HIGHER LIMIT IF GIVEN </HIGH_LIMIT>
 <UNIT>CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE</UNIT>
 <NON_NUMERIC_VALUE>CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE
    VALUE </NON_NUMERIC_VALUE>
   </AMOUNT>
   </SUBSTITUENT_GROUP>
   </SUBSTITUENT>
<DEGREE_POLYMERIZATION>
<AMOUNT>
 <AVERAGE>AVERAGE AMOUNT OR SINGLE VALUE</AVERAGE>
 <LOW_LIMIT>LOWER LIMIT IF GIVEN </LOW_LIMIT>
 <HIGH_LIMIT>HIGHER LIMIT IF GIVEN </HIGH_LIMIT>
 <UNIT>CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE</UNIT>
 <NON_NUMERIC_VALUE>CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE
    VALUE </NON_NUMERIC_VALUE>
   </AMOUNT>
   </DEGREE_POLYMERIZATION>
```

```
    </STRUCTURAL_REPEAT_UNIT_GROUP>
    </STRUCTURAL_REPEAT_UNIT>
<MOLECULAR_WEIGHT>
 <MOLECULAR_WEIGHT_TYPE>CV: EXACT, AVERAGE (AKA. NUMBER AVERAGE), WEIGHT
    AVERAGE, ETC.</MOLECULAR_WEIGHT_TYPE>
 <MOLECULAR_WEIGHT_METHOD>CV: SDS_PAGE, CALCULATED, LIGHT SCATTERING
    VISCOSITY, GEL PERMEATION, ETC.</MOLECULAR_WEIGHT_METHOD>
<AMOUNT>
 <AVERAGE>TYPICALLY GIVEN IN A MONOGRAPH OR PRODUCT DESCRIPTION OF A
    SUBSTANCE. IF LIMITS ARE ONLY GIVEN THE ARITHMETIC MEAN WOULD BE THE
    AVERAGE. USE FOR EXACT VALUES AND/OR VALUES WITHOUT STATED
    LIMITS.</AVERAGE>
 <LOW_LIMIT>TYPICALLY GIVEN IN A MONOGRAPH OR PRODUCT DESCRIPTION OF A
    SUBSTANCE</LOW_LIMIT>
 <UNIT>IF RELATIVE AMOUNTS ARE NOT EXPRESSED AS A RATIO</UNIT>
 <HIGH_LIMIT>TYPICALLY GIVEN IN A MONOGRAPH OR PRODUCT DESCRIPTION OF A
    SUBSTANCE</HIGH_LIMIT>
 <NON_NUMERIC_VALUE>CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE
    VALUE </NON_NUMERIC_VALUE>
   </AMOUNT>
   </MOLECULAR_WEIGHT>
<PROPERTY_GROUP>
 <PROPERTY_TYPE>CV: PHYSICAL, CHEMICAL, ENZYMATIC,
    IMMUNOLOGICAL</PROPERTY_TYPE>
 <PROPERTY>CV: VISCOSITY, DENSITY, PH, ENZYMATIC ACTIVITY, PRESENCE OR
    ABSENCE OF ANTIGEN ETC.</PROPERTY>
 <SUBSTANCE_ID>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 POSTIVE CELLS) </SUBSTANCE_ID>
<AMOUNT>
 <AVERAGE>AVERAGE AMOUNT OR SINGLE VALUE</AVERAGE>
 <LOW_LIMIT>LOWER LIMIT IF GIVEN </LOW_LIMIT>
 <HIGH_LIMIT>HIGHER LIMIT IF GIVEN </HIGH_LIMIT>
 <UNIT>CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE</UNIT>
 <NON_NUMERIC_VALUE>CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE
    VALUE </NON_NUMERIC_VALUE>
   </AMOUNT>
   </PROPERTY_GROUP>
<SOURCE_MATERIAL>
 <SOURCE_MATERIAL_CLASS>CV: ORGANISM, MINERAL</SOURCE_MATERIAL_CLASS>
 <SOURCE_MATERIAL_TYPE>CV: BACTERIUM, HUMAN, MAMMAL, FUNGUS, VIRUS, PLANT,
    VIRUS, ETC. DESCRIBES THE GENERAL CLASSIFICATION OF THE SOURCE MATERIAL.
    PRECISION IS BASED ON RELATEDNESS TO HUMANS. FOR VACCINES THIS IS THE
    CLASS OF INFECTIOUS AGENT.</SOURCE_MATERIAL_TYPE>
 <SOURCE_MATERIAL_STATE>CV: LIVE, ACTIVATED, INACTIVATED, ATTENUATED,
    CONJUGATED, LIVE(ATTENUATED). FOR INACTIVATED VACCINES, THE INACTIVATION
    METHOD AND AGENT IS TO BE SPECIFIED IN THE MODIFICATION
    GROUP</SOURCE_MATERIAL_STATE>
 <ORGANISM_NAME_ID>IDENTIFIER ASSOCIATED WITH THE SOURCE MATERIAL PARENT
    ORGANISM</ORGANISM_NAME_ID>
 <ORGANISM_NAME>ORGANISM ACCEPTED LATIN NAME</ORGANISM_NAME>
 <DEVELOPMENTAL_STAGE>CV: FETUS, INFANT, JUVENILE, ADULT, SENESCENT, LEAFING,
    PRE-FLOWERING, FLOWERING, FRUITING, ETC.</DEVELOPMENTAL_STAGE>
<PART_GROUP>
 <PART>PORTION OF AN ORGANISM WITH A DEFINABLE ANATOMICAL LOCATION</PART>
```

<PART_LOCATION>**ANATOMICAL LOCATION**</PART_LOCATION>
</PART_GROUP>
<FRACTION>
<MATERIAL_TYPE>**CV: CELL, PROTEIN, POLYSACCHARIDE, NUCLEIC ACID, EXPRESSED OIL, ESSENTIAL OIL**</MATERIAL_TYPE>
<FRACTION>**FRACTIONS OF FRACTIONS ARE ALLOWED**</FRACTION>
</FRACTION>
<MODIFICATION_GROUP>
<MODIFICATION_TYPE>**CV: PHYSICAL, STRUCTURAL, AGENT**</MODIFICATION_TYPE>
<RESIDUE_MODIFIED>**CV: 20 AMINO ACIDS PLUS 5 NUCLEOTIDES**</RESIDUE_MODIFIED>
<RESIDUE_SITE>**POSITION OF SPECIFIC MODIFICATIONS (IE. 1_10 REFERS TO THE 10TH RESIDUE ON THE 1ST SUBUNIT)**</RESIDUE_SITE>
<PHYSICAL_MODIFICATION>
<ROLE>**CV: INACTIVATION, ACTIVATION, ETC.**</ROLE>
<NUMBER_OF_PARAMETERS>**NUMBER OF PARAMETER GROUPS NEEDED TO DESCRIBE THE PHYSICAL MODIFICATION**</NUMBER_OF_PARAMETERS>
<PARAMETER_GROUP>
<PARAMETER>**CV: TIME, TEMPERATURE, ETC.**</PARAMETER>
<AMOUNT>
<AVERAGE>**AVERAGE AMOUNT OR SINGLE VALUE**</AVERAGE>
<LOW_LIMIT>**LOWER LIMIT IF GIVEN** </LOW_LIMIT>
<HIGH_LIMIT>**HIGHER LIMIT IF GIVEN** </HIGH_LIMIT>
<UNIT>**CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE**</UNIT>
<NON_NUMERIC_VALUE>**CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE VALUE** </NON_NUMERIC_VALUE>
</AMOUNT>
</PARAMETER_GROUP>
</PHYSICAL_MODIFICATION>
<STRUCTURAL_MODIFICATION>
<STRUCTURAL_MODIFICATION_TYPE>**CV: MOIETY, MOLECULAR FRAGMENT**</STRUCTURAL_MODIFICATION_TYPE>
<MOIETY>
<ROLE>**CV: ANTIGEN, LINKER, CONJUGATE, ETC.**</ROLE>
<MOIETY_ID>**UNIQUE IDENTIFIER FOR MODIFYING MOIETY**</MOIETY_ID>
<AMOUNT_TYPE>**CV: MOLE RATIO; WEIGHT AVERAGE**</AMOUNT_TYPE>
<AMOUNT>
<AVERAGE>**AVERAGE AMOUNT OR SINGLE VALUE**</AVERAGE>
<LOW_LIMIT>**LOWER LIMIT IF GIVEN** </LOW_LIMIT>
<HIGH_LIMIT>**HIGHER LIMIT IF GIVEN** </HIGH_LIMIT>
<UNIT>**CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE**</UNIT>
<NON_NUMERIC_VALUE>**CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE VALUE** </NON_NUMERIC_VALUE>
</AMOUNT>
</MOIETY>
<MOLECULAR_FRAGMENT>
<ROLE>**CV: ANTIGEN, LINKER, CONJUGATE, ETC.**</ROLE>
<MOLECULAR_FRAGMENT_ID>**UNIQUE IDENTIFIER FOR MOLECULAR FRAGMENT**</MOLECULAR_FRAGMENT_ID>
<AMOUNT_TYPE>**CV: MOLE RATIO; WEIGHT AVERAGE**</AMOUNT_TYPE>
<AMOUNT>
<AVERAGE>**AVERAGE AMOUNT OR SINGLE VALUE**</AVERAGE>
<LOW_LIMIT>**LOWER LIMIT IF GIVEN** </LOW_LIMIT>
<HIGH_LIMIT>**HIGHER LIMIT IF GIVEN** </HIGH_LIMIT>
<UNIT>**CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE**</UNIT>

<NON_NUMERIC_VALUE>**CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE VALUE** </NON_NUMERIC_VALUE>
</AMOUNT>
</MOLECULAR_FRAGMENT>
</STRUCTURAL_MODIFICATION>
<AGENT_MODIFICATION>
<AGENT_MODIFICATION_TYPE>**CV: CHEMICAL, ENZYMATIC, IMMUNOLOGICAL, ORGANISMAL**</AGENT_MODIFICATION_TYPE>
<ROLE>**CV: INACTIVATION, ACTIVATION, PEGYLATION, PHOSPHORYLATION, ETC.**</ROLE>
<MODIFICATION_AGENT>**ESTABLISHED OR PRIMARY NAME OF MODIFYING AGENT**</MODIFICATION_AGENT>
<MODIFICATION_AGENT_ID>**UNIQUE IDENTIFIER FOR MODIFYING AGENT**</MODIFICATION_AGENT_ID>
<MODIFICATION_PROCESS>**CV: CELL CULTURE, UV IRRADIATION, ETC.**</MODIFICATION_PROCESS>
<AMOUNT>
<AVERAGE>**AVERAGE AMOUNT OR SINGLE VALUE**</AVERAGE>
<LOW_LIMIT>**LOWER LIMIT IF GIVEN** </LOW_LIMIT>
<HIGH_LIMIT>**HIGHER LIMIT IF GIVEN** </HIGH_LIMIT>
<UNIT>**CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE**</UNIT>
<NON_NUMERIC_VALUE>**CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE VALUE** </NON_NUMERIC_VALUE>
</AMOUNT>
</AGENT_MODIFICATION>
</MODIFICATION_GROUP>
</SOURCE_MATERIAL>
<COMMENTS>**INFORMATION NECESSARY TO DISTINGUISH SUBSTANCES OR ESSENTIAL TO THE DEFINITION OF A SUBSTANCE SHOULD BE USED SPARINGLY**</COMMENTS>
</POLYMER>
<STRUCTURALLY_DIVERSE>
<SOURCE_MATERIAL>
<SOURCE_MATERIAL_CLASS>**CV: ORGANISM, MINERAL**</SOURCE_MATERIAL_CLASS>
<SOURCE_MATERIAL_TYPE>**CV: BACTERIUM, HUMAN, MAMMAL, FUNGUS, VIRUS, PLANT, VIRUS, ETC. DESCRIBES THE GENERAL CLASSIFICATION OF THE SOURCE MATERIAL. PRECISION IS BASED ON RELATEDNESS TO HUMANS. FOR VACCINES THIS IS THE CLASS OF INFECTIOUS AGENT.**</SOURCE_MATERIAL_TYPE>
<SOURCE_MATERIAL_STATE>**CV: LIVE, ACTIVATED, INACTIVATED, ATTENUATED, CONJUGATED, LIVE(ATTENUATED). FOR INACTIVATED VACCINES, THE INACTIVATION METHOD AND AGENT IS TO BE SPECIFIED IN THE MODIFICATION GROUP**</SOURCE_MATERIAL_STATE>
<ORGANISM_NAME_ID>**IDENTIFIER ASSOCIATED WITH THE SOURCE MATERIAL PARENT ORGANISM**</ORGANISM_NAME_ID>
<ORGANISM_NAME>**ORGANISM ACCEPTED LATIN NAME**</ORGANISM_NAME>
<DEVELOPMENTAL_STAGE>**CV: FETUS, INFANT, JUVENILE, ADULT, SENESCENT, LEAFING, PRE-FLOWERING, FLOWERING, FRUITING, ETC.**</DEVELOPMENTAL_STAGE>
<PART_GROUP>
<PART>**PORTION OF AN ORGANISM WITH A DEFINABLE ANATOMICAL LOCATION**</PART>
<PART_LOCATION>**ANATOMICAL LOCATION**</PART_LOCATION>
</PART_GROUP>
<FRACTION>
<MATERIAL_TYPE>**CV: CELL, PROTEIN, POLYSACCHARIDE, NUCLEIC ACID, EXPRESSED OIL, ESSENTIAL OIL**</MATERIAL_TYPE>
<FRACTION>**FRACTIONS OF FRACTIONS ARE ALLOWED**</FRACTION>

```
    </FRACTION>
<MODIFICATION_GROUP>
 <MODIFICATION_TYPE>CV: PHYSICAL, STRUCTURAL, AGENT</MODIFICATION_TYPE>
 <RESIDUE_MODIFIED>CV: 20 AMINO ACIDS PLUS 5 NUCLEOTIDES</RESIDUE_MODIFIED>
 <RESIDUE_SITE>POSITION OF SPECIFIC MODIFICATIONS (IE. 1_10 REFERS TO THE 10TH
    RESIDUE ON THE 1ST SUBUNIT)</RESIDUE_SITE>
<PHYSICAL_MODIFICATION>
 <ROLE>CV: INACTIVATION, ACTIVATION, ETC.</ROLE>
 <NUMBER_OF_PARAMETERS>NUMBER OF PARAMETER GROUPS NEEDED TO DESCRIBE THE
    PHYSICAL MODIFICATION</NUMBER_OF_PARAMETERS>
<PARAMETER_GROUP>
 <PARAMETER>CV: TIME, TEMPERATURE, ETC.</PARAMETER>
<AMOUNT>
 <AVERAGE>AVERAGE AMOUNT OR SINGLE VALUE</AVERAGE>
 <LOW_LIMIT>LOWER LIMIT IF GIVEN </LOW_LIMIT>
 <HIGH_LIMIT>HIGHER LIMIT IF GIVEN </HIGH_LIMIT>
 <UNIT>CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE</UNIT>
 <NON_NUMERIC_VALUE>CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE
    VALUE </NON_NUMERIC_VALUE>
   </AMOUNT>
   </PARAMETER_GROUP>
   </PHYSICAL_MODIFICATION>
<STRUCTURAL_MODIFICATION>
 <STRUCTURAL_MODIFICATION_TYPE>CV: MOIETY, MOLECULAR
    FRAGMENT</STRUCTURAL_MODIFICATION_TYPE>
<MOIETY>
 <ROLE>CV: ANTIGEN, LINKER, CONJUGATE, ETC.</ROLE>
 <MOIETY_ID>UNIQUE IDENTIFIER FOR MODIFYING MOIETY</MOIETY_ID>
 <AMOUNT_TYPE>CV: MOLE RATIO; WEIGHT AVERAGE</AMOUNT_TYPE>
<AMOUNT>
 <AVERAGE>AVERAGE AMOUNT OR SINGLE VALUE</AVERAGE>
 <LOW_LIMIT>LOWER LIMIT IF GIVEN </LOW_LIMIT>
 <HIGH_LIMIT>HIGHER LIMIT IF GIVEN </HIGH_LIMIT>
 <UNIT>CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE</UNIT>
 <NON_NUMERIC_VALUE>CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE
    VALUE </NON_NUMERIC_VALUE>
   </AMOUNT>
   </MOIETY>
<MOLECULAR_FRAGMENT>
 <ROLE>CV: ANTIGEN, LINKER, CONJUGATE, ETC.</ROLE>
 <MOLECULAR_FRAGMENT_ID>UNIQUE IDENTIFIER FOR MOLECULAR
    FRAGMENT</MOLECULAR_FRAGMENT_ID>
 <AMOUNT_TYPE>CV: MOLE RATIO; WEIGHT AVERAGE</AMOUNT_TYPE>
<AMOUNT>
 <AVERAGE>AVERAGE AMOUNT OR SINGLE VALUE</AVERAGE>
 <LOW_LIMIT>LOWER LIMIT IF GIVEN </LOW_LIMIT>
 <HIGH_LIMIT>HIGHER LIMIT IF GIVEN </HIGH_LIMIT>
 <UNIT>CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE</UNIT>
 <NON_NUMERIC_VALUE>CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE
    VALUE </NON_NUMERIC_VALUE>
   </AMOUNT>
   </MOLECULAR_FRAGMENT>
   </STRUCTURAL_MODIFICATION>
<AGENT_MODIFICATION>
```

```
 <AGENT_MODIFICATION_TYPE>CV: CHEMICAL, ENZYMATIC, IMMUNOLOGICAL,
    ORGANISMAL</AGENT_MODIFICATION_TYPE>
 <ROLE>CV: INACTIVATION, ACTIVATION, PEGYLATION, PHOSPHORYLATION,
    ETC.</ROLE>
 <MODIFICATION_AGENT>ESTABLISHED OR PRIMARY NAME OF MODIFYING
    AGENT</MODIFICATION_AGENT>
 <MODIFICATION_AGENT_ID>UNIQUE IDENTIFIER FOR MODIFYING
    AGENT</MODIFICATION_AGENT_ID>
 <MODIFICATION_PROCESS>CV: CELL CULTURE, UV IRRADIATION,
    ETC.</MODIFICATION_PROCESS>
<AMOUNT>
 <AVERAGE>AVERAGE AMOUNT OR SINGLE VALUE</AVERAGE>
 <LOW_LIMIT>LOWER LIMIT IF GIVEN </LOW_LIMIT>
 <HIGH_LIMIT>HIGHER LIMIT IF GIVEN </HIGH_LIMIT>
 <UNIT>CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE</UNIT>
 <NON_NUMERIC_VALUE>CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE
    VALUE </NON_NUMERIC_VALUE>
   </AMOUNT>
   </AGENT_MODIFICATION>
   </MODIFICATION_GROUP>
   </SOURCE_MATERIAL>
<PROPERTY_GROUP>
 <PROPERTY_TYPE>CV: PHYSICAL, CHEMICAL, ENZYMATIC,
    IMMUNOLOGICAL</PROPERTY_TYPE>
 <PROPERTY>CV: VISCOSITY, DENSITY, PH, ENZYMATIC ACTIVITY, PRESENCE OR
    ABSENCE OF ANTIGEN ETC.</PROPERTY>
 <SUBSTANCE_ID>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 POSTIVE CELLS) </SUBSTANCE_ID>
<AMOUNT>
 <AVERAGE>AVERAGE AMOUNT OR SINGLE VALUE</AVERAGE>
 <LOW_LIMIT>LOWER LIMIT IF GIVEN </LOW_LIMIT>
 <HIGH_LIMIT>HIGHER LIMIT IF GIVEN </HIGH_LIMIT>
 <UNIT>CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE</UNIT>
 <NON_NUMERIC_VALUE>CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE
    VALUE </NON_NUMERIC_VALUE>
   </AMOUNT>
   </PROPERTY_GROUP>
 <COMMENTS>INFORMATION NECESSARY TO DISTINGUISH SUBSTANCES OR ESSENTIAL
    TO THE DEFINITION OF A SUBSTANCE SHOULD BE USED SPARINGLY</COMMENTS>
   </STRUCTURALLY_DIVERSE>
   </SINGLE_SUBSTANCE>
<MIXTURE_SUBSTANCE>
 <MIXTURE_TYPE>CV: ONE OF, ALL OF, ANY OF</MIXTURE_TYPE>
<SOURCE_MATERIAL>
 <SOURCE_MATERIAL_CLASS>CV: ORGANISM, MINERAL</SOURCE_MATERIAL_CLASS>
 <SOURCE_MATERIAL_TYPE>CV: BACTERIUM, HUMAN, MAMMAL, FUNGUS, VIRUS, PLANT,
    VIRUS, ETC. DESCRIBES THE GENERAL CLASSIFICATION OF THE SOURCE MATERIAL.
    PRECISION IS BASED ON RELATEDNESS TO HUMANS. FOR VACCINES THIS IS THE
    CLASS OF INFECTIOUS AGENT.</SOURCE_MATERIAL_TYPE>
 <SOURCE_MATERIAL_STATE>CV: LIVE, ACTIVATED, INACTIVATED, ATTENUATED,
    CONJUGATED, LIVE(ATTENUATED). FOR INACTIVATED VACCINES, THE INACTIVATION
    METHOD AND AGENT IS TO BE SPECIFIED IN THE MODIFICATION
    GROUP</SOURCE_MATERIAL_STATE>
```

<ORGANISM_NAME_ID>**IDENTIFIER ASSOCIATED WITH THE SOURCE MATERIAL PARENT ORGANISM**</ORGANISM_NAME_ID>
<ORGANISM_NAME>**ORGANISM ACCEPTED LATIN NAME**</ORGANISM_NAME>
<DEVELOPMENTAL_STAGE>**CV: FETUS, INFANT, JUVENILE, ADULT, SENESCENT, LEAFING, PRE-FLOWERING, FLOWERING, FRUITING, ETC.**</DEVELOPMENTAL_STAGE>
<PART_GROUP>
<PART>**PORTION OF AN ORGANISM WITH A DEFINABLE ANATOMICAL LOCATION**</PART>
<PART_LOCATION>**ANATOMICAL LOCATION**</PART_LOCATION>
</PART_GROUP>
<FRACTION>
<MATERIAL_TYPE>**CV: CELL, PROTEIN, POLYSACCHARIDE, NUCLEIC ACID, EXPRESSED OIL, ESSENTIAL OIL**</MATERIAL_TYPE>
<FRACTION>**FRACTIONS OF FRACTIONS ARE ALLOWED**</FRACTION>
</FRACTION>
<MODIFICATION_GROUP>
<MODIFICATION_TYPE>**CV: PHYSICAL, STRUCTURAL, AGENT**</MODIFICATION_TYPE>
<RESIDUE_MODIFIED>**CV: 20 AMINO ACIDS PLUS 5 NUCLEOTIDES**</RESIDUE_MODIFIED>
<RESIDUE_SITE>**POSITION OF SPECIFIC MODIFICATIONS (IE. 1_10 REFERS TO THE 10TH RESIDUE ON THE 1ST SUBUNIT)**</RESIDUE_SITE>
<PHYSICAL_MODIFICATION>
<ROLE>**CV: INACTIVATION, ACTIVATION, ETC.**</ROLE>
<NUMBER_OF_PARAMETERS>**NUMBER OF PARAMETER GROUPS NEEDED TO DESCRIBE THE PHYSICAL MODIFICATION**</NUMBER_OF_PARAMETERS>
<PARAMETER_GROUP>
<PARAMETER>**CV: TIME, TEMPERATURE, ETC.**</PARAMETER>
<AMOUNT>
<AVERAGE>**AVERAGE AMOUNT OR SINGLE VALUE**</AVERAGE>
<LOW_LIMIT>**LOWER LIMIT IF GIVEN** </LOW_LIMIT>
<HIGH_LIMIT>**HIGHER LIMIT IF GIVEN** </HIGH_LIMIT>
<UNIT>**CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE**</UNIT>
<NON_NUMERIC_VALUE>**CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE VALUE** </NON_NUMERIC_VALUE>
</AMOUNT>
</PARAMETER_GROUP>
</PHYSICAL_MODIFICATION>
<STRUCTURAL_MODIFICATION>
<STRUCTURAL_MODIFICATION_TYPE>**CV: MOIETY, MOLECULAR FRAGMENT**</STRUCTURAL_MODIFICATION_TYPE>
<MOIETY>
<ROLE>**CV: ANTIGEN, LINKER, CONJUGATE, ETC.**</ROLE>
<MOIETY_ID>**UNIQUE IDENTIFIER FOR MODIFYING MOIETY**</MOIETY_ID>
<AMOUNT_TYPE>**CV: MOLE RATIO; WEIGHT AVERAGE**</AMOUNT_TYPE>
<AMOUNT>
<AVERAGE>**AVERAGE AMOUNT OR SINGLE VALUE**</AVERAGE>
<LOW_LIMIT>**LOWER LIMIT IF GIVEN** </LOW_LIMIT>
<HIGH_LIMIT>**HIGHER LIMIT IF GIVEN** </HIGH_LIMIT>
<UNIT>**CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE**</UNIT>
<NON_NUMERIC_VALUE>**CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE VALUE** </NON_NUMERIC_VALUE>
</AMOUNT>
</MOIETY>
<MOLECULAR_FRAGMENT>
<ROLE>**CV: ANTIGEN, LINKER, CONJUGATE, ETC.**</ROLE>

```
 <MOLECULAR_FRAGMENT_ID>UNIQUE IDENTIFIER FOR MOLECULAR
    FRAGMENT</MOLECULAR_FRAGMENT_ID>
 <AMOUNT_TYPE>CV: MOLE RATIO; WEIGHT AVERAGE</AMOUNT_TYPE>
<AMOUNT>
 <AVERAGE>AVERAGE AMOUNT OR SINGLE VALUE</AVERAGE>
 <LOW_LIMIT>LOWER LIMIT IF GIVEN </LOW_LIMIT>
 <HIGH_LIMIT>HIGHER LIMIT IF GIVEN </HIGH_LIMIT>
 <UNIT>CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE</UNIT>
 <NON_NUMERIC_VALUE>CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE
    VALUE </NON_NUMERIC_VALUE>
   </AMOUNT>
   </MOLECULAR_FRAGMENT>
   </STRUCTURAL_MODIFICATION>
<AGENT_MODIFICATION>
 <AGENT_MODIFICATION_TYPE>CV: CHEMICAL, ENZYMATIC, IMMUNOLOGICAL,
    ORGANISMAL</AGENT_MODIFICATION_TYPE>
 <ROLE>CV: INACTIVATION, ACTIVATION, PEGYLATION, PHOSPHORYLATION,
    ETC.</ROLE>
 <MODIFICATION_AGENT>ESTABLISHED OR PRIMARY NAME OF MODIFYING
    AGENT</MODIFICATION_AGENT>
 <MODIFICATION_AGENT_ID>UNIQUE IDENTIFIER FOR MODIFYING
    AGENT</MODIFICATION_AGENT_ID>
 <MODIFICATION_PROCESS>CV: CELL CULTURE, UV IRRADIATION,
    ETC.</MODIFICATION_PROCESS>
<AMOUNT>
 <AVERAGE>AVERAGE AMOUNT OR SINGLE VALUE</AVERAGE>
 <LOW_LIMIT>LOWER LIMIT IF GIVEN </LOW_LIMIT>
 <HIGH_LIMIT>HIGHER LIMIT IF GIVEN </HIGH_LIMIT>
 <UNIT>CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE</UNIT>
 <NON_NUMERIC_VALUE>CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE
    VALUE </NON_NUMERIC_VALUE>
   </AMOUNT>
   </AGENT_MODIFICATION>
   </MODIFICATION_GROUP>
   </SOURCE_MATERIAL>
<MIXTURE_CONSTITUENT_GROUP>
 <CONSTITUENT_NAME>ESTABLISHED OR PRIMARY SUBSTANCE
    NAME</CONSTITUENT_NAME>
 <CONSTITUENT_ID>UNIQUE IDENTIFIER FOR CONSTITUENT
    SUBSTANCE</CONSTITUENT_ID>
 <CONSTITUENT_REQUIREMENT>CV: ALWAYS PRESENT, MAY BE
    PRESENT</CONSTITUENT_REQUIREMENT>
   </MIXTURE_CONSTITUENT_GROUP>
   </MIXTURE_SUBSTANCE>
   </SUBSTANCE>
```

B.1.2 POTENTIAL DETAILED SPECIFIED SUBSTANCE MODEL IN XML

```
<SPECIFIED_SUBSTANCE>
 <SPECIFIED_SUBSTANCE_ID>SPECIFIED SUBSTANCE UNIQUE
    IDENTIFIER</SPECIFIED_SUBSTANCE_ID>
<NAME_GROUP>
 <NAME>NAME OF THE SUBSTANCE</NAME>
```

<NAME_TYPE>**CV: CODE, BRAND NAME, ESTABLISHED NAME, PRIMARY NAME, OTHER NAME. ESTABLISHED NAME REFERS TO A NON-PROPRIETARY NAME THAT IS ASSIGNED TO A SUBSTANCE BY A SCIENTIFIC ORGANIZATION AND/OR REGULATORY AGENCY (AUTHORITY SOURCE). PRIMARY NAME REFERS TO A NAME IS COMMON USE ASSOCIATED WITH A SUBSTANCE WHERE NO ESTABLISHED NAME IS AVAILABLE. CODES WILL BE ENTERED AS NAMES WHEN THE CODE SYSTEM IS PROPRIETARY. ALL SUBSTANCES WILL HAVE AT LEAST ONE NAME.**</NAME_TYPE>
<LANGUAGE_ISOCODE>**ISO 639-2 ALPHA-3 CODES**</LANGUAGE_ISOCODE>
<ESTABLISHED_NAME_GROUP>
<ESTABLISHED_NAME_TYPE>**CV: USAN, INN, JAN, BAN, USP, EP, INCI, JECFA**</ESTABLISHED_NAME_TYPE>
<ESTABLISHED_NAME_STATUS>**CV: CURRENT, ALTERNATE, SUPERCEDED**</ESTABLISHED_NAME_STATUS>
<ESTABLISHED_NAME_STATUS_CHANGE_DATE>**ISO 8601 DATE FORMAT**</ESTABLISHED_NAME_STATUS_CHANGE_DATE>
<ESTABLISHED_NAME_JURISDICTION_GROUP>
<ESTABLISHED_NAME_JURISDICTION>**EUROPEAN COMMISSION USAGE: ISO 3166-1 ALPHA-2 CODES WITH TWO EXCEPTIONS: EL (NOT GR) IS USED TO REPRESENT GREECE, AND UK (NOT GB) IS USED TO REPRESENT THE UNITED KINGDOM**</ESTABLISHED_NAME_JURISDICTION>
</ESTABLISHED_NAME_JURISDICTION_GROUP>
<ESTABLISHED_NAME_DOMAIN>**MULTIPLE VALUES - CV: FOOD, DRUG, COSMETIC, BIOLOGIC**</ESTABLISHED_NAME_DOMAIN>
</ESTABLISHED_NAME_GROUP>
<REFERENCE_SOURCE_GROUP>
<REFERENCE_SOURCE_CLASS>**CV: PUBLIC, REGULATORY SUBMISSION**</REFERENCE_SOURCE_CLASS>
<REFERENCE_SOURCE_TYPE>**CV: NDA, IND, BLA, ETC.**</REFERENCE_SOURCE_TYPE>
<REFERENCE_SOURCE_CITATION>**LITERATURE OR REGULATORY DOCUMENT SPECIFIC REFERENCE**</REFERENCE_SOURCE_CITATION>
</REFERENCE_SOURCE_GROUP>
</NAME_GROUP>
<REFERENCE_INFORMATION>
<CLASSIFICATION_GROUP>
<SUBSTANCE_CLASS>
CV: CHEMICAL, PROTEIN, NUCLEIC ACID, POLYMER, STRUCTURALLY DIVERSE, MULTIPLE
<TYPE>
CV: DIFFERENT VALUES FOR EACH SUBSTANCE CLASS GENERAL CLASSIFICATION OF THE SUBSTANCE CV TO BE DEVELOPED
<SUBTYPE>**CV: DIFFERENT VALUES FOR EACH SUBSTANCE CLASS SPECIFIC CLASSIFICATION OF THE SUBSTANCE** </SUBTYPE>
</TYPE>
<SUBSTANCE_CLASS_CODE>**CV: CLASSIFICATION CODE, ATC, NDF_RT ETC.**</SUBSTANCE_CLASS_CODE>
<CLASSIFICATION_SOURCE>**CV: SOURCE OF CLASSIFICATION**</CLASSIFICATION_SOURCE>
</SUBSTANCE_CLASS>
</CLASSIFICATION_GROUP>
<TARGET_GROUP>
<TARGET_GENE_ID>**NUMERIC OR ALPHA NUMERIC ID ASSOCIATED WITH THE HUMAN GENE OF ORIGIN**</TARGET_GENE_ID>
<TARGET_NAME>**TARGET OF A GIVEN SUBSTANCE OR SUBSTRATE UPON WHICH THE SUBSTANCE ACTS**</TARGET_NAME>
<INTERACTION_TYPE>**CV: SUBSTRATE, INHIBITION, BINDING, ACTIVATION, ETC.**</INTERACTION_TYPE>

<TARGET_ORGANISM_TYPE>**CV: HUMAN, BACTERIAL, VIRAL, ETC. THE ORGANISM TYPE FOR WHICH THE ACTIVE SUBSTANCE IS TARGETED** </TARGET_ORGANISM_TYPE>
<TARGET_REFERENCE_SOURCE>**SOURCE WHICH INDICATED THE TARGET OF THE SUBSTANCE (PRIMARY OR REPUTABLE SECONDARY SOURCES WOULD SUFFICE)**</TARGET_REFERENCE_SOURCE>
</TARGET_GROUP>
<GENE_GROUP>
<GENE_SEQUENCE_ORIGIN>**COMMON NAME FOR THE ORGANISM FROM WHICH THE FINAL SEQUENCE ORIGINATED.** </GENE_SEQUENCE_ORIGIN>
<GENE_ID>**NUMERIC OR ALPHANUMERIC ID ASSOCIATED WITH THE GENE OF ORIGIN**</GENE_ID>
<GENE_NAME>**COMPLETE NAME GIVEN FOR A GENE**</GENE_NAME>
<GENE_REFERENCE_SOURCE>**SOURCE FROM WHICH THE GENE SEQUENCE WAS DERIVED (GENBANK)**</GENE_REFERENCE_SOURCE>
</GENE_GROUP>
<COMMENTS/>
</REFERENCE_INFORMATION>
<PHYSICAL_FORM_GROUP>
<PHYSICAL_STATE>**CV: SOLID, LIQUID, GAS, EMULSION**</PHYSICAL_STATE>
<PHYSICAL_FORM_TYPE>**CV: CRYSTALINE, AMORPHOUS, TINCTURE, DRY, EXTRACT**</PHYSICAL_FORM_TYPE>
<AMOUNT_TYPE>**CV: CONCENTRATION; MASS; RATIO TYPE**</AMOUNT_TYPE>
<AMOUNT>
<AVERAGE>**AVERAGE AMOUNT OR SINGLE VALUE**</AVERAGE>
<LOW_LIMIT>**LOWER LIMIT IF GIVEN** </LOW_LIMIT>
<HIGH_LIMIT>**HIGHER LIMIT IF GIVEN** </HIGH_LIMIT>
<UNIT>**CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE**</UNIT>
<NON_NUMERIC_VALUE>**CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE VALUE** </NON_NUMERIC_VALUE>
</AMOUNT>
</PHYSICAL_FORM_GROUP>
<GRADE_GROUP>
<GRADE_TYPE>**CV: QUANTIFIED, STANDARDISED, USP, EP, ETC. PHARMACOPOEIAL SPECIFICATION TYPE OR OTHER SPECIFICATION TYPE**</GRADE_TYPE>
<GRADE_REFERENCE_SOURCE>**RELEVANT REFERENCE SOURCE OR MONOGRAPH (USP, EP, ACS, TECHNICAL)**</GRADE_REFERENCE_SOURCE>
</GRADE_GROUP>
<ANALYTICAL_DATA>
<UNITS_OF_MEASURE>
<UNIT_TYPE>**TYPE OF UNIT USED TO QUANTIFY THE SPECIFIED SUBSTANCE (CV, MASS, CONCENTRATION, ACTIVITY ETC.)** </UNIT_TYPE>
<UNIT_SOURCE>**CV: UCUM, USP, EP, WHO**</UNIT_SOURCE>
<REFERENCE_STANDARD>**USED WHEN THE AMOUNT OF MATERIAL IS DETERMINED RELATIVE TO A REFERENCE STANDARD**</REFERENCE_STANDARD>
<REFERENCE_STANDARD_ID>**SPECIFIED SUBSTANCE ID FOR REFERENCE MATERIAL**</REFERENCE_STANDARD_ID>
<REFERENCE_STANDARD_SOURCE>**SOURCE OF THE REFERENCE MATERIAL**</REFERENCE_STANDARD_SOURCE>
</UNITS_OF_MEASURE>
<ANALYTICAL_DATA_GROUP>
<ANALYTICAL_DATA_TYPE>**CV: PURITY, POTENCY, IDENTITY, LIMIT, ETC.**</ANALYTICAL_DATA_TYPE>
<ANALYTICAL_METHOD_GROUP>

<ANALYTICAL_METHOD_TYPE>**CV: SPECTROSCOPIC, CHROMATOGRAPHIC, MICROBIOLOGICAL** </ANALYTICAL_METHOD_TYPE>
<ANALYTICAL_METHOD>**CV: NMR, LC-MS, ELISA ETC**</ANALYTICAL_METHOD>
<ANALYTICAL_METHOD_DETAILS>**EXAMPLE: HPLC, L1, 4.6X250 MM, ISOCRATIC, 80:20 WATER:ACETONITRILE, 1 ML/MIN; BRIEF FREE TEXT DESCRIPTION**</ANALYTICAL_METHOD_DETAILS>
<ANALYTICAL_METHOD_REFERENCE_SOURCE>**E.G. EP, JP, USP, COMPANY, ETC**</ANALYTICAL_METHOD_REFERENCE_SOURCE>
<ANALYTICAL_METHOD_REFERENCE_DATA>**REFERENCE TO VALIDATION OR QUALIFICATION REPORT**</ANALYTICAL_METHOD_REFERENCE_DATA>
<ANALYSIS_SPECIFICATION_REFERENCE>
<SPECIFICATION_TYPE>**CV: PURITY; POTENCY; LIMIT**</SPECIFICATION_TYPE>
<SPECIFICATION_SOURCE>**CV; USP; EP; COMPANY**</SPECIFICATION_SOURCE>
<S_SUBSTANCE_NAME>**USUALLY NAME OF SUBSTANCE UPON WHICH SPECIFICATION IS SET; SPECIFIED SUBSTANCE WOULD BE USED IN RARE INSTANCES** </S_SUBSTANCE_NAME>
<S_SUBSTANCE_ID>**ID OF STANDARDS**</S_SUBSTANCE_ID>
<SPECIFICATION_AMOUNT_TYPE>**CV: PERCENT; MASS**</SPECIFICATION_AMOUNT_TYPE>
<AMOUNT>
<AVERAGE>**AVERAGE AMOUNT OR SINGLE VALUE**</AVERAGE>
<LOW_LIMIT>**LOWER LIMIT IF GIVEN** </LOW_LIMIT>
<HIGH_LIMIT>**HIGHER LIMIT IF GIVEN** </HIGH_LIMIT>
<UNIT>**CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE**</UNIT>
<NON_NUMERIC_VALUE>**CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE VALUE** </NON_NUMERIC_VALUE>
</AMOUNT>
</ANALYSIS_SPECIFICATION_REFERENCE>
</ANALYTICAL_METHOD_GROUP>
</ANALYTICAL_DATA_GROUP>
</ANALYTICAL_DATA>
<NUMBER_OF_MANUFACTURERS>**MANUFACTURERS RESPONSIBLE FOR PERFORMING CRITICAL PROCESSES**</NUMBER_OF_MANUFACTURERS>
<MANUFACTURING_GROUP>
<MANUFACTURER_NAME>**DISPLAY NAME OF THE ORGANISATION THAT PRODUCED THE RESULTANT SPECIFIED SUBSTANCE**</MANUFACTURER_NAME>
<MANUFACTURER_ID>**CODE USED BY REGULATORS TO TRACK MANUFACTURERS (E.G. DUNS NUMBER)**</MANUFACTURER_ID>
<MANUFACTURING_TYPE>**SIC CODE**</MANUFACTURING_TYPE>
<PRODUCTION_METHOD_TYPE>**CV: SYNTHETIC, EXTRACTIVE, RECOMBINANT, HYBRIDOMA, SEMI-SYNTHETIC**</PRODUCTION_METHOD_TYPE>
<NUMBER_OF_CRITICAL_PROCESSES>**NUMBER OF CRITICAL PROCESS STEPS ACCOMPLISHED BY THE MANUFACTURER**</NUMBER_OF_CRITICAL_PROCESSES>
<CRITICAL_PROCESS_GROUP>
<CRITICAL_PROCESS_ID>**CODE USED TO TRACK CRITICAL PROCESS STEPS (NECESSARY ELEMENTS: SPECIFIED SUBSTANCE, MANUFACTURER, PROCESS STEP)**</CRITICAL_PROCESS_ID>
<CRITICAL_PROCESS_SITE>**CODE USED BY REGULATORS TO TRACK MANUFACTURERS (E.G. DUNS NUMBER)**</CRITICAL_PROCESS_SITE>
<STARTING_MATERIAL_GROUP>
<NUMBER_OF_STARTING_MATERIALS>**NUMBER OF STARTING MATERIALS USED BY THE MANUFACTURER IN THE CRITICAL PROCESS STEP**</NUMBER_OF_STARTING_MATERIALS>
<AMOUNT_TYPE>**WEIGHT, CONCENTRATION ETC.**</AMOUNT_TYPE>
<SUBSTANCE_GROUP>

<SUBSTANCE_NAME>**ESTABLISHED OR PRIMARY NAME FOR THE SUBSTANCE OR SPECIFIED SUBSTANCE**</SUBSTANCE_NAME>
<SUBSTANCE_ID>**SUBSTANCE OR SPECIFIED SUBSTANCE UNIQUE IDENTIFIER**</SUBSTANCE_ID>
<SUBSTANCE_ROLE>**PURPOSE OF THE SUBSTANCE OR SPECIFIED SUBSTANCE IN THE CRITICAL PROCESS STEP**</SUBSTANCE_ROLE>
<SPECIFICATION_TYPE>**SPECIFICATION OF STARTING MATERIAL USED (GRADE)**</SPECIFICATION_TYPE>
<AMOUNT>
<AVERAGE>**AVERAGE AMOUNT OR SINGLE VALUE**</AVERAGE>
<LOW_LIMIT>**LOWER LIMIT IF GIVEN** </LOW_LIMIT>
<HIGH_LIMIT>**HIGHER LIMIT IF GIVEN** </HIGH_LIMIT>
<UNIT>**CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE**</UNIT>
<NON_NUMERIC_VALUE>**CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE VALUE** </NON_NUMERIC_VALUE>
</AMOUNT>
</SUBSTANCE_GROUP>
</STARTING_MATERIAL_GROUP>
<PROCESSING_MATERIAL_GROUP>
<NUMBER_OF_PROCESSING_MATERIALS>**NUMBER OF MATERIALS IN DIRECT CONTACT WITH STARTING MATERIAL BUT NOT PRESENT IN THE PRODUCT**</NUMBER_OF_PROCESSING_MATERIALS>
<MATERIAL_TYPE>**EXTRACTION SOLVENT, SOLVENT, ETC. GENERAL ROLE OF THE MATERIAL IN THE PROCESS**</MATERIAL_TYPE>
<PROCESSING_SUBSTANCE_GROUP>
<NUMBER_OF_PROCESSING_SUBSTANCES>**NUMBER OF SUBSTANCE OR SPECIFIED SUBSTANCES USED AS PROCESSING MATERIALS IN THE CRITICAL PROCESS STEP** </NUMBER_OF_PROCESSING_SUBSTANCES>
<AMOUNT_TYPE>**WEIGHT, CONCENTRATION, MOLE FRACTION, ETC.**</AMOUNT_TYPE>
<SUBSTANCE_GROUP>
<SUBSTANCE_NAME>**ESTABLISHED OR PRIMARY NAME FOR THE SUBSTANCE OR SPECIFIED SUBSTANCE**</SUBSTANCE_NAME>
<SUBSTANCE_ID>**SUBSTANCE OR SPECIFIED SUBSTANCE UNIQUE IDENTIFIER**</SUBSTANCE_ID>
<SUBSTANCE_ROLE>**PURPOSE OF THE SUBSTANCE OR SPECIFIED SUBSTANCE IN THE CRITICAL PROCESS STEP**</SUBSTANCE_ROLE>
<SPECIFICATION_TYPE>**SPECIFICATION OF STARTING MATERIAL USED (GRADE)**</SPECIFICATION_TYPE>
<AMOUNT>
<AVERAGE>**AVERAGE AMOUNT OR SINGLE VALUE**</AVERAGE>
<LOW_LIMIT>**LOWER LIMIT IF GIVEN** </LOW_LIMIT>
<HIGH_LIMIT>**HIGHER LIMIT IF GIVEN** </HIGH_LIMIT>
<UNIT>**CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE**</UNIT>
<NON_NUMERIC_VALUE>**CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE VALUE** </NON_NUMERIC_VALUE>
</AMOUNT>
</SUBSTANCE_GROUP>
</PROCESSING_SUBSTANCE_GROUP>
</PROCESSING_MATERIAL_GROUP>
<RESULTANT_MATERIAL_GROUP>
<RESULTANT_MATERIAL_NUMBER>**NUMBER OF MATERIALS PRODUCED IN THE CRITICAL PROCESS STEP**</RESULTANT_MATERIAL_NUMBER>
<MATERIAL_TYPE>**CV: API, INTERMEDIATE PRODUCT, FORMULATION, ETC.**</MATERIAL_TYPE>

<RESULTANT_SUBSTANCE_GROUP>
<NUMBER_OF_RESULTANT_SPECIFIED_SUBSTANCES>**NUMBER OF SUBSTANCES OR SPECIFIED SUBSTANCES IN THE RESULTANT MATERIAL**</NUMBER_OF_RESULTANT_SPECIFIED_SUBSTANCES>
<AMOUNT_TYPE>**WEIGHT, CONCENTRATION, MOLE FRACTION, ETC.**</AMOUNT_TYPE>
<SUBSTANCE_GROUP>
<SUBSTANCE_NAME>**ESTABLISHED OR PRIMARY NAME FOR THE SUBSTANCE OR SPECIFIED SUBSTANCE**</SUBSTANCE_NAME>
<SUBSTANCE_ID>**SUBSTANCE OR SPECIFIED SUBSTANCE UNIQUE IDENTIFIER**</SUBSTANCE_ID>
<SUBSTANCE_ROLE>**PURPOSE OF THE SUBSTANCE OR SPECIFIED SUBSTANCE IN THE CRITICAL PROCESS STEP**</SUBSTANCE_ROLE>
<SPECIFICATION_TYPE>**SPECIFICATION OF STARTING MATERIAL USED (GRADE)**</SPECIFICATION_TYPE>
<AMOUNT>
<AVERAGE>**AVERAGE AMOUNT OR SINGLE VALUE**</AVERAGE>
<LOW_LIMIT>**LOWER LIMIT IF GIVEN** </LOW_LIMIT>
<HIGH_LIMIT>**HIGHER LIMIT IF GIVEN** </HIGH_LIMIT>
<UNIT>**CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE**</UNIT>
<NON_NUMERIC_VALUE>**CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE VALUE** </NON_NUMERIC_VALUE>
</AMOUNT>
</SUBSTANCE_GROUP>
</RESULTANT_SUBSTANCE_GROUP>
</RESULTANT_MATERIAL_GROUP>
<PROCESSING_EQUPMENT_GROUP>
<EQUIPMENT_NUMBER>**NUMBER OF CRITICAL PIECES OF EQUIPMENT USED IN THE CRITICAL PROCESS STEP**</EQUIPMENT_NUMBER>
<EQUIPMENT_GROUP>
<EQUIPMENT_NAME>**CV: PUMP, BIOREACTOR LYOPHILIZER, ETC. GENERAL NAME FOR EQUIPMENT** </EQUIPMENT_NAME>
<EQUIPMENT_ID>**MODEL OR PRODUCT NUMBER FROM MANUFACTURER**</EQUIPMENT_ID>
<MANUFACTURER_NAME>**PROCESS EQUIPMENT MANUFACTURER**</MANUFACTURER_NAME>
<MANUFACTURER_ID>**CODE USED BY REGULATORS TO TRACK MANUFACTURERS (E.G. DUNS NUMBER)**</MANUFACTURER_ID>
</EQUIPMENT_GROUP>
</PROCESSING_EQUPMENT_GROUP>
<PROCESSING_PARAMETER_GROUP>
<NUMBER_OF_PARAMETERS>**NUMBER OF PARAMETER GROUPS NEEDED TO DESCRIBE THE CRITICAL PROCESS STEP**</NUMBER_OF_PARAMETERS>
<PARAMETER_GROUP>
<PARAMETER>**CV: TIME, TEMPERATURE, ETC.**</PARAMETER>
<AMOUNT>
<AVERAGE>**AVERAGE AMOUNT OR SINGLE VALUE**</AVERAGE>
<LOW_LIMIT>**LOWER LIMIT IF GIVEN** </LOW_LIMIT>
<HIGH_LIMIT>**HIGHER LIMIT IF GIVEN** </HIGH_LIMIT>
<UNIT>**CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE**</UNIT>
<NON_NUMERIC_VALUE>**CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE VALUE** </NON_NUMERIC_VALUE>
</AMOUNT>
</PARAMETER_GROUP>
</PROCESSING_PARAMETER_GROUP>
</CRITICAL_PROCESS_GROUP>
</MANUFACTURING_GROUP>

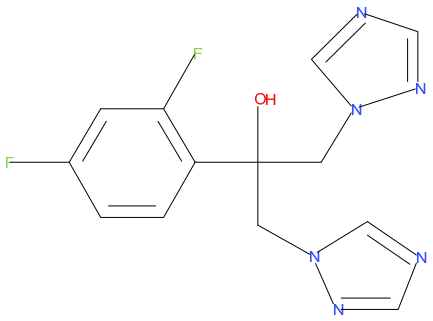
```
<CONSTITUENT_GROUP>
 <SUBSTANCE_ID>SUBSTANCE UNIQUE IDENTIFIER</SUBSTANCE_ID>
 <SUBSTANCE_NAME>ESTABLISHED OR PRIMARY NAME OF THE CONSTITUENT
    SUBSTANCE</SUBSTANCE_NAME>
 <SUBSTANCE_ROLE>ROLE OF THE SUBSTANCE IN THE SPECIFIED SUBSTANCE CV: ACTIVE
    MARKER, COMPONENT, IMPURITY, DEGRADENT, METABOLITE, ETC.
    </SUBSTANCE_ROLE>
 <AMOUNT_TYPE>CV: MOLE RATIO, WEIGHT PERCENT</AMOUNT_TYPE>
<AMOUNT>
 <AVERAGE>AVERAGE AMOUNT OR SINGLE VALUE</AVERAGE>
 <LOW_LIMIT>LOWER LIMIT IF GIVEN </LOW_LIMIT>
 <HIGH_LIMIT>HIGHER LIMIT IF GIVEN </HIGH_LIMIT>
 <UNIT>CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE</UNIT>
 <NON_NUMERIC_VALUE>CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE
    VALUE </NON_NUMERIC_VALUE>
   </AMOUNT>
   </CONSTITUENT_GROUP>
<PROPERTY_GROUP>
 <PROPERTY>CV: VISCOSITY, DENSITY, PH, ENZYMATIC ACTIVITY, ANTIGEN PRESENT (,
    ETC.</PROPERTY>
 <SUBSTANCE_ID>SUBSTANCE UNIQUE IDENTIFIER FOR ANTIGEN</SUBSTANCE_ID>
<AMOUNT>
 <AVERAGE>AVERAGE AMOUNT OR SINGLE VALUE</AVERAGE>
 <LOW_LIMIT>LOWER LIMIT IF GIVEN </LOW_LIMIT>
 <HIGH_LIMIT>HIGHER LIMIT IF GIVEN </HIGH_LIMIT>
 <UNIT>CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE</UNIT>
 <NON_NUMERIC_VALUE>CV: YES, NO, POSITIVE, NEGATIVE, COMPLETE,
    PARTIAL</NON_NUMERIC_VALUE>
   </AMOUNT>
   </PROPERTY_GROUP>
   </SPECIFIED_SUBSTANCE>
```

B.2 Substance Examples

B.2.1 Simple Chemical

FLUCONAZOLE

Graphical Structural Representation




```
<SUBSTANCE>
 <SUBSTANCE_ID/>
<NAME_GROUP>
 <NAME>FLUCONAZOLE</NAME>
 <NAME_TYPE>ESTABLISHED</NAME_TYPE>
 <LANGUAGE_ISOCODE>ENG</LANGUAGE_ISOCODE>
<ESTABLISHED_NAME_GROUP>
 <ESTABLISHED_NAME_TYPE>USAN</ESTABLISHED_NAME_TYPE>
 <ESTABLISHED_NAME_STATUS>CURRENT</ESTABLISHED_NAME_STATUS>
 <ESTABLISHED_NAME_STATUS_CHANGE_DATE/>
<ESTABLISHED_NAME_JURISDICTION_GROUP>
 <ESTABLISHED_NAME_JURISDICTION>USA</ESTABLISHED_NAME_JURISDICTION>
   </ESTABLISHED_NAME_JURISDICTION_GROUP>
 <ESTABLISHED_NAME_DOMAIN>DRUG</ESTABLISHED_NAME_DOMAIN>
   </ESTABLISHED_NAME_GROUP>
<ESTABLISHED_NAME_GROUP>
 <ESTABLISHED_NAME_TYPE>JAN</ESTABLISHED_NAME_TYPE>
 <ESTABLISHED_NAME_STATUS>CURRENT</ESTABLISHED_NAME_STATUS>
 <ESTABLISHED_NAME_STATUS_CHANGE_DATE/>
<ESTABLISHED_NAME_JURISDICTION_GROUP>
 <ESTABLISHED_NAME_JURISDICTION>JAPAN</ESTABLISHED_NAME_JURISDICTION>
   </ESTABLISHED_NAME_JURISDICTION_GROUP>
 <ESTABLISHED_NAME_DOMAIN>DRUG</ESTABLISHED_NAME_DOMAIN>
   </ESTABLISHED_NAME_GROUP>
<ESTABLISHED_NAME_GROUP>
 <ESTABLISHED_NAME_TYPE>BAN</ESTABLISHED_NAME_TYPE>
 <ESTABLISHED_NAME_STATUS>CURRENT</ESTABLISHED_NAME_STATUS>
 <ESTABLISHED_NAME_STATUS_CHANGE_DATE/>
<ESTABLISHED_NAME_JURISDICTION_GROUP>
 <ESTABLISHED_NAME_JURISDICTION>GREAT BRITAIN</ESTABLISHED_NAME_JURISDICTION>
   </ESTABLISHED_NAME_JURISDICTION_GROUP>
 <ESTABLISHED_NAME_DOMAIN>DRUG</ESTABLISHED_NAME_DOMAIN>
   </ESTABLISHED_NAME_GROUP>
<ESTABLISHED_NAME_GROUP>
 <ESTABLISHED_NAME_TYPE>INN</ESTABLISHED_NAME_TYPE>
 <ESTABLISHED_NAME_STATUS>CURRENT</ESTABLISHED_NAME_STATUS>
 <ESTABLISHED_NAME_STATUS_CHANGE_DATE/>
<ESTABLISHED_NAME_JURISDICTION_GROUP>
 <ESTABLISHED_NAME_JURISDICTION/>
   </ESTABLISHED_NAME_JURISDICTION_GROUP>
 <ESTABLISHED_NAME_DOMAIN>DRUG</ESTABLISHED_NAME_DOMAIN>
```

```
    </ESTABLISHED_NAME_GROUP>
<REFERENCE_SOURCE_GROUP>
 <REFERENCE_SOURCE_CLASS>PUBLIC</REFERENCE_SOURCE_CLASS>
 <REFERENCE_SOURCE_TYPE/>
 <REFERENCE_SOURCE_CITATION>USP DICTIONARY 2009</REFERENCE_SOURCE_CITATION>
    </REFERENCE_SOURCE_GROUP>
    </NAME_GROUP>
<NAME_GROUP>
 <NAME>2,4-DIFLUORO-1',1'-BIS(1H-1,2,4-TRIAZOL-1-YLMETHYL)BENZYL
    ALCOHOL</NAME>
 <NAME_TYPE>CHEMICAL</NAME_TYPE>
 <LANGUAGE_ISOCODE>ENG</LANGUAGE_ISOCODE>
<ESTABLISHED_NAME_GROUP>
 <ESTABLISHED_NAME_TYPE/>
 <ESTABLISHED_NAME_STATUS/>
 <ESTABLISHED_NAME_STATUS_CHANGE_DATE/>
<ESTABLISHED_NAME_JURISDICTION_GROUP>
 <ESTABLISHED_NAME_JURISDICTION>USA</ESTABLISHED_NAME_JURISDICTION>
    </ESTABLISHED_NAME_JURISDICTION_GROUP>
 <ESTABLISHED_NAME_DOMAIN/>
    </ESTABLISHED_NAME_GROUP>
<REFERENCE_SOURCE_GROUP>
 <REFERENCE_SOURCE_CLASS>PUBLIC</REFERENCE_SOURCE_CLASS>
 <REFERENCE_SOURCE_TYPE/>
 <REFERENCE_SOURCE_CITATION>USP DICTIONARY 2009</REFERENCE_SOURCE_CITATION>
    </REFERENCE_SOURCE_GROUP>
    </NAME_GROUP>
<CODE_GROUP>
 <CODE>86386-73-4</CODE>
 <CODE_SYSTEM>CAS</CODE_SYSTEM>
 <CODE_SYSTEM_ID/>
 <CODE_STATUS>CURRENT</CODE_STATUS>
 <CODE_STATUS_CHANGE_DATE/>
 <CODE_REFERENCE_CITATION>USP DICTIONARY 2009</CODE_REFERENCE_CITATION>
    </CODE_GROUP>
<CODE_GROUP>
 <CODE>123631-92-5</CODE>
 <CODE_SYSTEM>CAS</CODE_SYSTEM>
 <CODE_SYSTEM_ID/>
 <CODE_STATUS>DELETED</CODE_STATUS>
 <CODE_STATUS_CHANGE_DATE/>
 <CODE_REFERENCE_CITATION>SCIFINDER</CODE_REFERENCE_CITATION>
    </CODE_GROUP>
<CODE_GROUP>
 <CODE>D01AC15</CODE>
 <CODE_SYSTEM>ATC</CODE_SYSTEM>
 <CODE_SYSTEM_ID/>
 <CODE_STATUS>CURRENT</CODE_STATUS>
 <CODE_STATUS_CHANGE_DATE/>
 <CODE_REFERENCE_CITATION>WHO ATC;
    http://www.whocc.no/atc_ddd_index/?code=D01AC15</CODE_REFERENCE_CITATION
    >
   </CODE_GROUP>
<CODE_GROUP>
```

<CODE>**J02AC01**</CODE>
<CODE_SYSTEM>**ATC**</CODE_SYSTEM>
<CODE_SYSTEM_ID/>
<CODE_STATUS>**CURRENT**</CODE_STATUS>
<CODE_STATUS_CHANGE_DATE/>
<CODE_REFERENCE_CITATION>**WHO ATC; http://www.whocc.no/atc_ddd_index/?code=J02AC01**</CODE_REFERENCE_CITATION>
</CODE_GROUP>
<SINGLE_SUBSTANCE>
<STRUCTURE>
<STRUCTURAL_REPRESENTATION_TYPE>**INCHI**</STRUCTURAL_REPRESENTATION_TYPE>
<STRUCTURAL_REPRESENTATION>**1S/C13H12F2N6O/c14-10-1-2-11(12(15)3-10)13(22,4-20-8-16-6-18-20)5-21-9-17-7-19-21/h1-3,6-9,22H,4-5H2**</STRUCTURAL_REPRESENTATION>
<STEREOCHEMISTRY/>
<OPTICAL_ACTIVITY/>
<ISOTOPE_GROUP>
<NUCLIDE_ID/>
<NUCLIDE_NAME/>
<SUBSTITUTION_TYPE/>
</ISOTOPE_GROUP>
</STRUCTURE>
<REFERENCE_INFORMATION>
<CLASSIFICATION_GROUP>
<SUBSTANCE_CLASS>
CHEMICAL
<TYPE>
ANTI-FUNGAL
<SUBTYPE>**TRIAZOLE**</SUBTYPE>
</TYPE>
</SUBSTANCE_CLASS>
</CLASSIFICATION_GROUP>
<TARGET_GROUP>
<TARGET_GENE_ID/>
<TARGET_NAME/>
<INTERACTION_TYPE/>
<TARGET_ORGANISM_TYPE/>
<TARGET_REFERENCE_SOURCE/>
</TARGET_GROUP>
<COMMENTS/>
</REFERENCE_INFORMATION>
<CHEMICAL_SUBSTANCE>
<STOICHIOMETRIC>**YES**</STOICHIOMETRIC>
<NON_STOICHIOMETRIC>
<NUMBER_OF_MOIETIES/>
<MOIETY_AMOUNT_TYPE/>
<MOIETY_GROUP>
<MOIETY/>
<AMOUNT>
<AVERAGE/>
<LOW_LIMIT/>
<HIGH_LIMIT/>
<UNIT/>
<NON_NUMERIC_VALUE/>

```
    </AMOUNT>
    </MOIETY_GROUP>
  <MOIETY/>
<PROPERTY_GROUP>
  <PROPERTY_TYPE/>
  <PROPERTY/>
  <SUBSTANCE_ID/>
<AMOUNT>
  <AVERAGE/>
  <LOW_LIMIT/>
  <HIGH_LIMIT/>
  <UNIT/>
  <NON_NUMERIC_VALUE/>
    </AMOUNT>
    </PROPERTY_GROUP>
    </NON_STOICHIOMETRIC>
    <COMMENTS/>
   </CHEMICAL_SUBSTANCE>
    </SINGLE_SUBSTANCE>
    </SUBSTANCE>
```

B.2.2 Non-stoichiometric chemical

Aluminum Zirconium Tetrachlorohydrex Gly

Defined based on USP Monograph and USP Dictionary

USP Dictionary

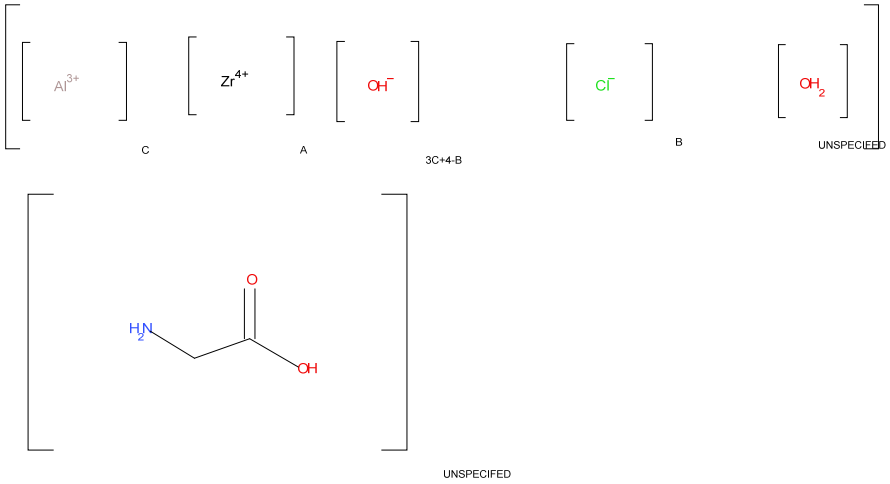
A coordination complex of aluminum zirconium tetrachlorohydrate [AlyZr(OH)3y+4-xClx.nH2O] and glycine in which some of the water molecules normally coordinated to the metals have been displaced by the glycine. (1) Glycine aluminum-zirconium complex; (2) Aluminum zirconium glycine tetrachloro hydrate complex.

USP/NF

Aluminum Zirconium Tetrachlorohydrex Gly is a derivative of Aluminum Zirconium Tetrachlorohydrate in which some of the water molecules have been displaced by glycine, calcium glycinate, magnesium glycinate, potassium glycinate, sodium glycinate, or zinc glycinate. It encompasses a range of aluminum-to-zirconium atomic ratios between 2.0:1 and 5.99:1 and a range of (aluminum plus zirconium)-to-chloride atomic ratios between 1.5:1 and 0.9:1. It contains not less than 90.0 percent and not more than 110.0 percent of the labeled amount of anhydrous aluminum zirconium tetrachlorohydrate

Moiety descriptions will only be used for moieties with a defined ratio range and that are independent. In this case A, B and C. All ratios are relative to zirconium (A) which is assigned a value of 1. Ratio of glycine and water not given.

Graphical Structural Representation



<SUBSTANCE>
<SUBSTANCE_ID>**8O386558JE**</SUBSTANCE_ID>
<NAME_GROUP>
<NAME>**ALUMINUM ZIRCONIUM TETRACHLOROHYDREX GLY**</NAME>
<NAME_TYPE>**ESTABLISHED**</NAME_TYPE>
<LANGUAGE_ISOCODE>**ENG**</LANGUAGE_ISOCODE>
<ESTABLISHED_NAME_GROUP>
<ESTABLISHED_NAME_TYPE>**USAN**</ESTABLISHED_NAME_TYPE>
<ESTABLISHED_NAME_STATUS>**CURRENT**</ESTABLISHED_NAME_STATUS>
<ESTABLISHED_NAME_STATUS_CHANGE_DATE/>
<ESTABLISHED_NAME_JURISDICTION_GROUP>
<ESTABLISHED_NAME_JURISDICTION>**USA**</ESTABLISHED_NAME_JURISDICTION>
</ESTABLISHED_NAME_JURISDICTION_GROUP>
<ESTABLISHED_NAME_DOMAIN>**DRUG**</ESTABLISHED_NAME_DOMAIN>
</ESTABLISHED_NAME_GROUP>
<REFERENCE_SOURCE_GROUP>
<REFERENCE_SOURCE_CLASS>**PUBLIC**</REFERENCE_SOURCE_CLASS>
<REFERENCE_SOURCE_TYPE/>
<REFERENCE_SOURCE_CITATION>**USP DICTIONARY 2009**</REFERENCE_SOURCE_CITATION>
</REFERENCE_SOURCE_GROUP>
</NAME_GROUP>
<CODE_GROUP>
<CODE/>
<CODE_SYSTEM/>
<CODE_SYSTEM_ID/>
<CODE_STATUS/>
<CODE_STATUS_CHANGE_DATE/>
<CODE_REFERENCE_CITATION/>
</CODE_GROUP>
<SINGLE_SUBSTANCE>
<STRUCTURE>
<STRUCTURAL_REPRESENTATION_TYPE>**INCHI**</STRUCTURAL_REPRESENTATION_TYPE>
<STRUCTURAL_REPRESENTATION>**1S/C2H5NO2.AL.CLH.2H2O.ZR/C3-1-2(4)5;;;;;/H1,3H2,(H,4,5);;1H;2*1H2;/Q;+3;;;;+4/P-2**</STRUCTURAL_REPRESENTATION>
<STEREOCHEMISTRY/>
<OPTICAL_ACTIVITY/>

<ISOTOPE_GROUP>
<NUCLIDE_ID/>
<NUCLIDE_NAME/>
<SUBSTITUTION_TYPE/>
</ISOTOPE_GROUP>
</STRUCTURE>
<REFERENCE_INFORMATION>
<CLASSIFICATION_GROUP>
<SUBSTANCE_CLASS>
CHEMICAL
<TYPE>
<SUBTYPE/>
</TYPE>
</SUBSTANCE_CLASS>
</CLASSIFICATION_GROUP>
<TARGET_GROUP>
<TARGET_GENE_ID/>
<TARGET_NAME/>
<INTERACTION_TYPE/>
<TARGET_ORGANISM_TYPE/>
<TARGET_REFERENCE_SOURCE/>
</TARGET_GROUP>
<COMMENTS/>
</REFERENCE_INFORMATION>
<CHEMICAL_SUBSTANCE>
<STOICHIOMETRIC>**YES**</STOICHIOMETRIC>
<NON_STOICHIOMETRIC>
<NUMBER_OF_MOIETIES>**6**</NUMBER_OF_MOIETIES>
<MOIETY_AMOUNT_TYPE>**MOLE RATIO**</MOIETY_AMOUNT_TYPE>
<MOIETY_GROUP>
<MOIETY_NAME>**ZIRCONIUM ION (4+)**</MOIETY_NAME>
<MOIETY_ID>**A**</MOIETY_ID>
<AMOUNT>
<AVERAGE>**1**</AVERAGE>
<LOW_LIMIT/>
<HIGH_LIMIT/>
<UNIT/>
<NON_NUMERIC_VALUE/>
</AMOUNT>
</MOIETY_GROUP>
<MOIETY_GROUP>
<MOIETY_NAME>**CHLORIDE ANION**</MOIETY_NAME>
<MOIETY_ID>**B**</MOIETY_ID>
<AMOUNT>
<AVERAGE>**4**</AVERAGE>
<LOW_LIMIT>**2**</LOW_LIMIT>
<HIGH_LIMIT>**7.78**</HIGH_LIMIT>
<UNIT>**RATIO TO ZIRCONIUM**</UNIT>
<NON_NUMERIC_VALUE/>
</AMOUNT>
</MOIETY_GROUP>
<MOIETY_GROUP>
<MOIETY_NAME>**ALUMINUM ION**</MOIETY_NAME>
<MOIETY_ID>**C**</MOIETY_ID>

<AMOUNT>
<AVERAGE>**4**</AVERAGE>
<LOW_LIMIT>**2**</LOW_LIMIT>
<HIGH_LIMIT>**5.99**</HIGH_LIMIT>
<UNIT>**RATIO TO ZIRCONIUM**</UNIT>
<NON_NUMERIC_VALUE/>
</AMOUNT>
</MOIETY_GROUP>
<MOIETY_GROUP>
<MOIETY_NAME>**HYDROXIDE ANION**</MOIETY_NAME>
<MOIETY_ID>**3C+4-B**</MOIETY_ID>
<AMOUNT>
<AVERAGE>**12**</AVERAGE>
<LOW_LIMIT/>
<HIGH_LIMIT/>
<UNIT>
<UNIT>**RATIO TO ZIRCONIUM**</UNIT>
</UNIT>
<NON_NUMERIC_VALUE/>
</AMOUNT>
</MOIETY_GROUP>
<MOIETY_GROUP>
<MOIETY_NAME>**GLYCINE**</MOIETY_NAME>
<MOIETY_ID>**UNDEFINED**</MOIETY_ID>
<AMOUNT>
<AVERAGE/>
<LOW_LIMIT/>
<HIGH_LIMIT/>
<UNIT/>
<NON_NUMERIC_VALUE>**UNDEFINED**</NON_NUMERIC_VALUE>
</AMOUNT>
</MOIETY_GROUP>
<MOIETY_GROUP>
<MOIETY_NAME>**WATER**</MOIETY_NAME>
<MOIETY_ID/>
<AMOUNT>
<AVERAGE/>
<LOW_LIMIT/>
<HIGH_LIMIT/>
<UNIT/>
<NON_NUMERIC_VALUE>**UNDEFINED**</NON_NUMERIC_VALUE>
</AMOUNT>
</MOIETY_GROUP>
<PROPERTY_GROUP>
<PROPERTY_TYPE/>
<PROPERTY/>
<SUBSTANCE_ID/>
<AMOUNT>
<AVERAGE/>
<LOW_LIMIT/>
<HIGH_LIMIT/>
<UNIT/>
<NON_NUMERIC_VALUE/>

```
    </AMOUNT>
    </PROPERTY_GROUP>
    <COMMENTS/>
   </NON_STOICHIOMETRIC>
    </CHEMICAL_SUBSTANCE>
    </SINGLE_SUBSTANCE>
    </SUBSTANCE>
```

B.2.3 Peptide Example

Calcitonin Salmon is a peptide hormone that is either chemically synthesized or manufactured using recombinant DNA technology.

```
<SUBSTANCE>
 <SUBSTANCE_ID/>
<NAME_GROUP>
 <NAME>CALCITONIN SALMON</NAME>
 <NAME_TYPE>ESTABLISHED NAME</NAME_TYPE>
 <LANGUAGE_ISOCODE>ENG</LANGUAGE_ISOCODE>
<ESTABLISHED_NAME_GROUP>
 <ESTABLISHED_NAME_TYPE>USP</ESTABLISHED_NAME_TYPE>
 <ESTABLISHED_NAME_STATUS>CURRENT</ESTABLISHED_NAME_STATUS>
 <ESTABLISHED_NAME_STATUS_CHANGE_DATE/>
<ESTABLISHED_NAME_JURISDICTION_GROUP>
 <ESTABLISHED_NAME_JURISDICTION/>
   </ESTABLISHED_NAME_JURISDICTION_GROUP>
 <ESTABLISHED_NAME_DOMAIN>DRUG</ESTABLISHED_NAME_DOMAIN>
   </ESTABLISHED_NAME_GROUP>
<REFERENCE_SOURCE_GROUP>
 <REFERENCE_SOURCE_CLASS>USP </REFERENCE_SOURCE_CLASS>
 <REFERENCE_SOURCE_TYPE/>
 <REFERENCE_SOURCE_CITATION>USP 32</REFERENCE_SOURCE_CITATION>
   </REFERENCE_SOURCE_GROUP>
   </NAME_GROUP>
<CODE_GROUP>
 <CODE>47931-85-1</CODE>
 <CODE_SYSTEM>CAS</CODE_SYSTEM>
 <CODE_SYSTEM_ID/>
 <CODE_STATUS>CURRENT</CODE_STATUS>
 <CODE_STATUS_CHANGE_DATE/>
 <CODE_REFERENCE_CITATION>USP DICTIONARY 2009</CODE_REFERENCE_CITATION>
   </CODE_GROUP>
<CODE_GROUP>
 <CODE>H05BA01</CODE>
 <CODE_SYSTEM>ATC</CODE_SYSTEM>
 <CODE_SYSTEM_ID/>
 <CODE_STATUS>CURRENT</CODE_STATUS>
 <CODE_STATUS_CHANGE_DATE/>
 <CODE_REFERENCE_CITATION>WHO ATC:
   http://www.whocc.no/atc_ddd_index/?code=H05BA01</CODE_REFERENCE_CITATION
   >
   </CODE_GROUP>
<SINGLE_SUBSTANCE>
```

<STRUCTURE>
<STRUCTURAL_REPRESENTATION_TYPE/>
<STRUCTURAL_REPRESENTATION/>
<STEREOCHEMISTRY/>
<OPTICAL_ACTIVITY/>
<ISOTOPE_GROUP>
<NUCLIDE_ID/>
<NUCLIDE_NAME/>
<SUBSTITUTION_TYPE/>
</ISOTOPE_GROUP>
</STRUCTURE>
<REFERENCE_INFORMATION>
<CLASSIFICATION_GROUP>
<SUBSTANCE_CLASS>
PROTEIN
<TYPE>
HORMONE
<SUBTYPE>**PEPTIDE**</SUBTYPE>
</TYPE>
</SUBSTANCE_CLASS>
</CLASSIFICATION_GROUP>
<TARGET_GROUP>
<TARGET_GENE_ID>**799**</TARGET_GENE_ID>
<TARGET_NAME>**CALCR CALCITONIN RECEPTOR (HOMO SAPIENS)** </TARGET_NAME>
<INTERACTION_TYPE>**AGONIST**</INTERACTION_TYPE>
<TARGET_ORGANISM_TYPE>**HUMAN**</TARGET_ORGANISM_TYPE>
<TARGET_REFERENCE_SOURCE>**GENBANK**</TARGET_REFERENCE_SOURCE>
</TARGET_GROUP>
<GENE_GROUP>
<GENE_SEQUENCE_ORIGIN>**SALMON**</GENE_SEQUENCE_ORIGIN>
<GENE_ID>**100196557**</GENE_ID>
<GENE_NAME>**calc1 Calcitonin-1 (Salmo salar)** </GENE_NAME>
<GENE_REFERENCE_SOURCE>**GENE**</GENE_REFERENCE_SOURCE>
</GENE_GROUP>
<COMMENTS/>
</REFERENCE_INFORMATION>
<PROTEIN>
<SEQUENCE_TYPE>**COMPLETE**</SEQUENCE_TYPE>
<NUMBER_OF_SUBUNITS>**1**</NUMBER_OF_SUBUNITS>
<SUBUNIT_GROUP>
<SUBUNIT>**1**</SUBUNIT>
<LENGTH>**32**</LENGTH>
<SEQUENCE>**CSNLSTCVLGKLSQELHKLQTYPRTNTGSGTP**</SEQUENCE>
<N_TERMINAL_MODIFICATION/>
<N_TERMINAL_MODIFICATION_ID/>
<C_TERMINAL_MODIFICATION>**AMIDE**</C_TERMINAL_MODIFICATION>
<C_TERMINAL_MODIFICATION_ID/>
</SUBUNIT_GROUP>
<DISULFIDE_LINKAGE>**1_1-1_7**</DISULFIDE_LINKAGE>
<GLYCOSYLATION_GROUP>
<GLYCOSYLATION_TYPE/>
<N_GLYCOSYLATION/>
<O_GLYCOSYLATION/>
<C_GLYCOSYLATION/>

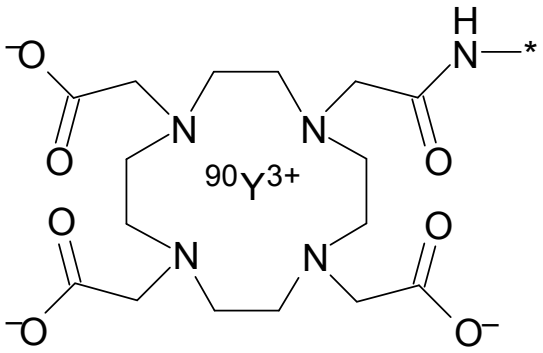
```
    </GLYCOSYLATION_GROUP>
<MODIFICATION_GROUP>
 <MODIFICATION_TYPE/>
 <RESIDUE_MODIFIED/>
 <RESIDUE_SITE/>
<PHYSICAL_MODIFICATION>
 <ROLE/>
 <NUMBER_OF_PARAMETERS/>
<PARAMETER_GROUP>
 <PARAMETER/>
<AMOUNT>
 <AVERAGE/>
 <LOW_LIMIT/>
 <HIGH_LIMIT/>
 <UNIT/>
 <NON_NUMERIC_VALUE/>
    </AMOUNT>
    </PARAMETER_GROUP>
    </PHYSICAL_MODIFICATION>
<STRUCTURAL_MODIFICATION>
 <STRUCTURAL_MODIFICATION_TYPE/>
<MOIETY>
 <ROLE/>
 <MOIETY_ID/>
 <AMOUNT_TYPE/>
<AMOUNT>
 <AVERAGE/>
 <LOW_LIMIT/>
 <HIGH_LIMIT/>
 <UNIT/>
 <NON_NUMERIC_VALUE/>
    </AMOUNT>
    </MOIETY>
<MOLECULAR_FRAGMENT>
 <ROLE/>
 <MOLECULAR_FRAGMENT_ID/>
 <AMOUNT_TYPE/>
<AMOUNT>
 <AVERAGE/>
 <LOW_LIMIT/>
 <HIGH_LIMIT/>
 <UNIT/>
 <NON_NUMERIC_VALUE/>
    </AMOUNT>
    </MOLECULAR_FRAGMENT>
    </STRUCTURAL_MODIFICATION>
<AGENT_MODIFICATION>
 <AGENT_MODIFICATION_TYPE/>
 <ROLE/>
 <MODIFICATION_AGENT/>
 <MODIFICATION_AGENT_ID/>
 <MODIFICATION_PROCESS/>
<AMOUNT>
 <AVERAGE/>
```

```
 <LOW_LIMIT/>
 <HIGH_LIMIT/>
 <UNIT/>
 <NON_NUMERIC_VALUE/>
   </AMOUNT>
   </AGENT_MODIFICATION>
   </MODIFICATION_GROUP>
<MOLECULAR_WEIGHT>
 <MOLECULAR_WEIGHT_TYPE>CALCULATED</MOLECULAR_WEIGHT_TYPE>
 <MOLECULAR_WEIGHT_METHOD/>
<AMOUNT>
 <AVERAGE>3432</AVERAGE>
 <LOW_LIMIT/>
 <HIGH_LIMIT/>
 <UNIT/>
 <NON_NUMERIC_VALUE/>
   </AMOUNT>
   </MOLECULAR_WEIGHT>
 <COMMENTS/>
   </PROTEIN>
   </SINGLE_SUBSTANCE>
   </SUBSTANCE>
```

B.2.4 Radiolabelled Monoclonal Antibody

YTTRIUM Y-90 CLIVATUZUMAB TETRAXETAN is a monoclonal antibody with a covalently attached chelating agent.

GRAPHICAL STRUCTURAL REPRESENTATION




```
<SUBSTANCE>
 <SUBSTANCE_ID>2L271110ED</SUBSTANCE_ID>
<NAME_GROUP>
 <NAME>YTTRIUM Y-90 CLIVATUZUMAB TETRAXETAN</NAME>
 <NAME_TYPE>ESTABLISHED</NAME_TYPE>
 <LANGUAGE_ISOCODE>ENG</LANGUAGE_ISOCODE>
<ESTABLISHED_NAME_GROUP>
 <ESTABLISHED_NAME_TYPE>USAN</ESTABLISHED_NAME_TYPE>
```

<ESTABLISHED_NAME_STATUS>**CURRENT**</ESTABLISHED_NAME_STATUS>
<ESTABLISHED_NAME_STATUS_CHANGE_DATE/>
<ESTABLISHED_NAME_JURISDICTION_GROUP>
<ESTABLISHED_NAME_JURISDICTION>**USA**</ESTABLISHED_NAME_JURISDICTION>
</ESTABLISHED_NAME_JURISDICTION_GROUP>
<ESTABLISHED_NAME_DOMAIN>**DRUG**</ESTABLISHED_NAME_DOMAIN>
</ESTABLISHED_NAME_GROUP>
<REFERENCE_SOURCE_GROUP>
<REFERENCE_SOURCE_CLASS>**PUBLIC**</REFERENCE_SOURCE_CLASS>
<REFERENCE_SOURCE_TYPE/>
<REFERENCE_SOURCE_CITATION>**USP DICTIONARY 2009**</REFERENCE_SOURCE_CITATION>
</REFERENCE_SOURCE_GROUP>
</NAME_GROUP>
<CODE_GROUP>
<CODE>**943976-23-6**</CODE>
<CODE_SYSTEM>**CAS**</CODE_SYSTEM>
<CODE_SYSTEM_ID/>
<CODE_STATUS>**CURRENT**</CODE_STATUS>
<CODE_STATUS_CHANGE_DATE/>
<CODE_REFERENCE_CITATION>**USP DICTIONARY 2009**</CODE_REFERENCE_CITATION>
</CODE_GROUP>
<SINGLE_SUBSTANCE>
<STRUCTURE>
<STRUCTURAL_REPRESENTATION_TYPE>**MOL**</STRUCTURAL_REPRESENTATION_TYPE>
<STRUCTURAL_REPRESENTATION>**30 29 0 0 0 0 0 0 0 0999 V2000 9.9563 -7.3055 0.0000 Y 1 1 0 0 0 0 0 0 0 0 0 0 15.0355 -4.8847 0.0000 * 0 0 0 0 0 0 0 0 0 0 0 0 13.3609 -8.0134 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0 13.8867 -9.9869 0.0000 O 0 5 0 0 0 0 0 0 0 0 0 6.4178 -6.8678 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0 5.8872 -4.8955 0.0000 O 0 5 0 0 0 0 0 0 0 0 0 0 6.7218 -5.7285 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 13.0541 -9.1519 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 13.3408 -6.8634 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0 13.8599 -4.8881 0.0000 N 0 0 0 0 0 0 0 0 0 0 0 0 13.0301 -5.7260 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 5.9099 -9.9441 0.0000 O 0 5 0 0 0 0 0 0 0 0 0 0 6.4492 -7.9743 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0 6.7482 -9.1149 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 7.8605 -5.4221 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 11.8897 -5.4263 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 11.9147 -9.4555 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 7.8855 -9.4263 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 7.6897 -8.0305 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 7.6897 -6.8513 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 8.7018 -6.2618 0.0000 N 0 0 0 0 0 0 0 0 0 0 0 0 9.2908 -5.2506 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 10.4700 -5.2524 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 11.0577 -6.2664 0.0000 N 0 0 0 0 0 0 0 0 0 0 0 0 12.0761 -6.8427 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 12.0891 -8.0218 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 8.7257 -8.5952 0.0000 N 0 0 0 0 0 0 0 0 0 0 0 0 11.0839 -8.6223 0.0000 N 0 0 0 0 0 0 0 0 0 0 0 0 10.4848 -9.6275 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 9.3057 -9.6139 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 10 2 1 0 0 0 0 8 3 2 0 0 0 0 25 24 1 0 0 0 0 8 4 1 0 0 0 0 27 18 1 0 0 0 0 7 5 2 0 0 0 0 26 28 1 0 0 0 0 7 6 1 0 0 0 0 19 27 1 0 0 0 0 15 7 1 0 0 0 0 20 21 1 0 0 0 0 17 8 1 0 0 0 0 30 27 1 0 0 0 0 11 9 2 0 0 0 0 30 29 1 0 0 0 0 11 10 1 0 0 0 0 20 19 1 0 0 0 0 16 11 1 0 0 0 0 22 21 1 0 0 0 0 14 12 1 0 0 0 0 23 24 1 0 0 0 0 14 13 2 0 0 0 0 18 14 1 0 0 0 0 26 25 1 0 0 0 0 21 15 1 0 0 0 0 29 28 1 0 0 0 0 24 16 1 0 0 0 0 23 22 1 0 0 0 0 28 17 1 0 0 0 0 M CHG 4 1 3 4 -1 6 -1 12 -1 M ISO 1 1 90 M END** </STRUCTURAL_REPRESENTATION>
<STEREOCHEMISTRY/>
<OPTICAL_ACTIVITY/>
<ISOTOPE_GROUP>
<NUCLIDE_NAME>**YTTRIUM-90**</NUCLIDE_NAME>
<NUCLIDE_ID>**1K8M7UR6O1**</NUCLIDE_ID>
<SUBSTITUTION_TYPE>**CHELATED**</SUBSTITUTION_TYPE>
</ISOTOPE_GROUP>

```
    </STRUCTURE>
<REFERENCE_INFORMATION>
<CLASSIFICATION_GROUP>
<SUBSTANCE_CLASS>
    PROTEIN
<TYPE>
    MONOCLONAL ANTIBODY, RADIOLABELLED
 <SUBTYPE>IGG1, HUMANIZED</SUBTYPE>
    </TYPE>
    </SUBSTANCE_CLASS>
    </CLASSIFICATION_GROUP>
<TARGET_GROUP>
 <TARGET_GENE_ID>4582</TARGET_GENE_ID>
 <TARGET_NAME>MUC1 MUCIN 1, CELL SURFACE ASSOCIATED (HOMO
    SAPIENS)</TARGET_NAME>
 <INTERACTION_TYPE>BINDING</INTERACTION_TYPE>
 <TARGET_ORGANISM_TYPE>HUMAN</TARGET_ORGANISM_TYPE>
 <TARGET_REFERENCE_SOURCE>GENBANK</TARGET_REFERENCE_SOURCE>
    </TARGET_GROUP>
<GENE_GROUP>
 <GENE_SEQUENCE_ORIGIN>HUMANIZED</GENE_SEQUENCE_ORIGIN>
 <GENE_ID/>
 <GENE_NAME/>
 <GENE_REFERENCE_SOURCE/>
    </GENE_GROUP>
 <COMMENTS/>
    </REFERENCE_INFORMATION>
<PROTEIN>
 <SEQUENCE_TYPE>COMPLETE</SEQUENCE_TYPE>
 <NUMBER_OF_SUBUNITS>4</NUMBER_OF_SUBUNITS>
<SUBUNIT_GROUP>
 <SUBUNIT>1</SUBUNIT>
 <LENGTH>449</LENGTH>
 <SEQUENCE>QVQLQQSGAEVKKFGASVKVSCEASGYTFPSYVLHWVKQAPGQGLEWIGYINPYND
    GTQYNKKFKGKATLTRDTSINTAYMELSRLRSDDTAVYYCARGFGGSYGFAYWGQGTLVIVSSA
    STKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSS
    VVTVPSSSLGTQTYICNVNHKPSNTKVDKRVEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKD
    TLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPRESQYNSTYRVVSVLTVLHQDW
    LNGKEYKCKVSNEALPAPIEKTISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAV
    EWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVNHEALHNHYTQKSLSL
    SPGK</SEQUENCE>
 <N_TERMINAL_MODIFICATION/>
 <N_TERMINAL_MODIFICATION_ID/>
 <C_TERMINAL_MODIFICATION/>
 <C_TERMINAL_MODIFICATION_ID/>
    </SUBUNIT_GROUP>
<SUBUNIT_GROUP>
 <SUBUNIT>2</SUBUNIT>
 <LENGTH>449</LENGTH>
 <SEQUENCE>QVQLQQSGAEVKKFGASVKVSCEASGYTFPSYVLHWVKQAPGQGLEWIGYINPYND
    GTQYNKKFKGKATLTRDTSINTAYMELSRLRSDDTAVYYCARGFGGSYGFAYWGQGTLVIVSSA
    STKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSS
    VVTVPSSSLGTQTYICNVNHKPSNTKVDKRVEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKD
    TLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPRESQYNSTYRVVSVLTVLHQDW
```

LNGKEYKCKVSNEALPAPIEKTISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAV EWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVNHEALHNHYTQKSLSL SPGK</SEQUENCE>
<N_TERMINAL_MODIFICATION/>
<N_TERMINAL_MODIFICATION_ID/>
<C_TERMINAL_MODIFICATION/>
<C_TERMINAL_MODIFICATION_ID/>
</SUBUNIT_GROUP>
<SUBUNIT_GROUP>
<SUBUNIT>**3**</SUBUNIT>
<LENGTH>**215**</LENGTH>
<SEQUENCE>**DIQLTQSPSSLSASVGDRVTMTCSASSSVSSSYLYWYQQKPGKAPKLWIYSTSNLAS GVPARFSGSGSGTDFTLTISSLQPEDSASYFCHQWNRYPYTFGGGTRLEIKRTVAAPSVFIFPPSD EQLKSGTASVVCLLNNFYEAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSPRKADYE KHKVYACEVTHQGLSSPVTKSFNRGEC**</SEQUENCE>
<N_TERMINAL_MODIFICATION/>
<N_TERMINAL_MODIFICATION_ID/>
<C_TERMINAL_MODIFICATION/>
<C_TERMINAL_MODIFICATION_ID/>
</SUBUNIT_GROUP>
<SUBUNIT_GROUP>
<SUBUNIT>**4**</SUBUNIT>
<LENGTH>**215**</LENGTH>
<SEQUENCE>**DIQLTQSPSSLSASVGDRVTMTCSASSSVSSSYLYWYQQKPGKAPKLWIYSTSNLAS GVPARFSGSGSGTDFTLTISSLQPEDSASYFCHQWNRYPYTFGGGTRLEIKRTVAAPSVFIFPPSD EQLKSGTASVVCLLNNFYEAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSPRKADYE KHKVYACEVTHQGLSSPVTKSFNRGEC**</SEQUENCE>
<N_TERMINAL_MODIFICATION/>
<N_TERMINAL_MODIFICATION_ID/>
<C_TERMINAL_MODIFICATION/>
<C_TERMINAL_MODIFICATION_ID/>
</SUBUNIT_GROUP>
<DISULFIDE_LINKAGE>**1_22-1_96;1_146-1_202;1_222-3_215;1_228-2_228;1_231-2_231;1_263-1_323;1_369-1_427;2_22-2_96;2_146-2_202;2_222-4_215;2_263-2_323;2_369-2_427;3_23-3_89;3_135-3_195;4_23-4_89;4_135-4_195**</DISULFIDE_LINKAGE>
<GLYCOSYLATION_GROUP>
<GLYCOSYLATION_TYPE>**MAMMALIAN**</GLYCOSYLATION_TYPE>
<N_GLYCOSYLATION>**1_299; 2_299**</N_GLYCOSYLATION>
<O_GLYCOSYLATION/>
<C_GLYCOSYLATION/>
</GLYCOSYLATION_GROUP>
<MODIFICATION_GROUP>
<MODIFICATION_TYPE>**CHEMICAL**</MODIFICATION_TYPE>
<RESIDUE_MODIFIED>**LYSINE**</RESIDUE_MODIFIED>
<RESIDUE_SITE/>
<PHYSICAL_MODIFICATION>
<ROLE/>
<NUMBER_OF_PARAMETERS/>
<PARAMETER_GROUP>
<PARAMETER/>
<AMOUNT>
<AVERAGE/>
<LOW_LIMIT/>

```
 <HIGH_LIMIT/>
 <UNIT/>
 <NON_NUMERIC_VALUE/>
   </AMOUNT>
   </PARAMETER_GROUP>
   </PHYSICAL_MODIFICATION>
<STRUCTURAL_MODIFICATION>
 <STRUCTURAL_MODIFICATION_TYPE>MOLECULAR
    FRAGMENT</STRUCTURAL_MODIFICATION_TYPE>
<MOIETY>
 <ROLE/>
 <MOIETY_ID/>
 <AMOUNT_TYPE/>
<AMOUNT>
 <AVERAGE/>
 <LOW_LIMIT/>
 <HIGH_LIMIT/>
 <UNIT/>
 <NON_NUMERIC_VALUE/>
   </AMOUNT>
   </MOIETY>
<MOLECULAR_FRAGMENT>
 <ROLE>CHELATOR</ROLE>
 <MOLECULAR_FRAGMENT_ID>1HTE449DGZ</MOLECULAR_FRAGMENT_ID>
 <AMOUNT_TYPE>MOLE RATIO</AMOUNT_TYPE>
<AMOUNT>
 <AVERAGE/>
 <LOW_LIMIT>2</LOW_LIMIT>
 <HIGH_LIMIT>5</HIGH_LIMIT>
 <UNIT/>
 <NON_NUMERIC_VALUE/>
   </AMOUNT>
   </MOLECULAR_FRAGMENT>
   </STRUCTURAL_MODIFICATION>
<AGENT_MODIFICATION>
 <AGENT_MODIFICATION_TYPE/>
 <ROLE/>
 <MODIFICATION_AGENT/>
 <MODIFICATION_AGENT_ID/>
 <MODIFICATION_PROCESS/>
<AMOUNT>
 <AVERAGE>AVERAGE AMOUNT OR SINGLE VALUE</AVERAGE>
 <LOW_LIMIT>LOWER LIMIT IF GIVEN </LOW_LIMIT>
 <HIGH_LIMIT>HIGHER LIMIT IF GIVEN </HIGH_LIMIT>
 <UNIT>CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE</UNIT>
 <NON_NUMERIC_VALUE>CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE
    VALUE </NON_NUMERIC_VALUE>
   </AMOUNT>
   </AGENT_MODIFICATION>
   </MODIFICATION_GROUP>
<MODIFICATION_GROUP>
 <MODIFICATION_TYPE>RADIOLIGAND</MODIFICATION_TYPE>
 <RESIDUE_MODIFIED/>
```

```
 <RESIDUE_SITE/>
<PHYSICAL_MODIFICATION>
 <ROLE/>
 <NUMBER_OF_PARAMETERS/>
<PARAMETER_GROUP>
 <PARAMETER/>
<AMOUNT>
 <AVERAGE/>
 <LOW_LIMIT/>
 <HIGH_LIMIT/>
 <UNIT/>
 <NON_NUMERIC_VALUE/>
   </AMOUNT>
   </PARAMETER_GROUP>
   </PHYSICAL_MODIFICATION>
<STRUCTURAL_MODIFICATION>
 <STRUCTURAL_MODIFICATION_TYPE>MOIETY</STRUCTURAL_MODIFICATION_TYPE>
<MOIETY>
 <ROLE>CHELATED RADIONUCLIDE</ROLE>
 <MOIETY_ID>433ME2ATHW</MOIETY_ID>
 <AMOUNT_TYPE>MOLE RATIO</AMOUNT_TYPE>
<AMOUNT>
 <AVERAGE/>
 <LOW_LIMIT>2</LOW_LIMIT>
 <HIGH_LIMIT>5</HIGH_LIMIT>
 <UNIT/>
 <NON_NUMERIC_VALUE/>
   </AMOUNT>
   </MOIETY>
<MOLECULAR_FRAGMENT>
 <ROLE>CHELATOR</ROLE>
 <MOLECULAR_FRAGMENT_ID>1HTE449DGZ</MOLECULAR_FRAGMENT_ID>
 <AMOUNT_TYPE>MOLE RATIO</AMOUNT_TYPE>
<AMOUNT>
 <AVERAGE/>
 <LOW_LIMIT>2</LOW_LIMIT>
 <HIGH_LIMIT>5</HIGH_LIMIT>
 <UNIT/>
 <NON_NUMERIC_VALUE/>
   </AMOUNT>
   </MOLECULAR_FRAGMENT>
   </STRUCTURAL_MODIFICATION>
<AGENT_MODIFICATION>
 <AGENT_MODIFICATION_TYPE/>
 <ROLE/>
 <MODIFICATION_AGENT/>
 <MODIFICATION_AGENT_ID/>
 <MODIFICATION_PROCESS/>
<AMOUNT>
 <AVERAGE>AVERAGE AMOUNT OR SINGLE VALUE</AVERAGE>
 <LOW_LIMIT>LOWER LIMIT IF GIVEN </LOW_LIMIT>
 <HIGH_LIMIT>HIGHER LIMIT IF GIVEN </HIGH_LIMIT>
 <UNIT>CONCENTRATION OR MASS IF NOT EXPRESSED AS A RATIO; RATIO TYPE</UNIT>
```

<NON_NUMERIC_VALUE>**CV: YES, NO, COMPLETE, PARTIAL REFERS TO A QUALITATIVE VALUE** </NON_NUMERIC_VALUE>
</AMOUNT>
</AGENT_MODIFICATION>
</MODIFICATION_GROUP>
<MOLECULAR_WEIGHT>
<MOLECULAR_WEIGHT_TYPE>**ESTIMATED**</MOLECULAR_WEIGHT_TYPE>
<MOLECULAR_WEIGHT_METHOD>**SDS-PAGE**</MOLECULAR_WEIGHT_METHOD>
<AMOUNT>
<AVERAGE>**145.7**</AVERAGE>
<LOW_LIMIT/>
<HIGH_LIMIT/>
<UNIT/>
<NON_NUMERIC_VALUE/>
</AMOUNT>
</MOLECULAR_WEIGHT>
<COMMENTS/>
</PROTEIN>
</SINGLE_SUBSTANCE>
</SUBSTANCE>

B.2.5 Polyclonal immunoglobulins

BLACK WIDOW SPIDER ANTIVENIN (EQUINE)

This antivenin is a polyclonal serum made in horses containing multiple isotypes, IgG, IgM, IgA etc. Horses are immunized with the toxin and we refer to this as the targeted antigen.

<SUBSTANCE>
<SUBSTANCE_ID>**5VJA14972G**</SUBSTANCE_ID>
<NAME_GROUP>
<NAME>**ANTIVENIN (LATRODECTUS MACTANS) (EQUINE)** </NAME>
<NAME_TYPE>**PROPER NAME**</NAME_TYPE>
<LANGUAGE_ISOCODE>**ENG**</LANGUAGE_ISOCODE>
<ESTABLISHED_NAME_GROUP>
<ESTABLISHED_NAME_TYPE>**PROPER**</ESTABLISHED_NAME_TYPE>
<ESTABLISHED_NAME_STATUS>**CURRENT**</ESTABLISHED_NAME_STATUS>
<ESTABLISHED_NAME_STATUS_CHANGE_DATE/>
<ESTABLISHED_NAME_JURISDICTION_GROUP>
<ESTABLISHED_NAME_JURISDICTION>**US**</ESTABLISHED_NAME_JURISDICTION>
</ESTABLISHED_NAME_JURISDICTION_GROUP>
<ESTABLISHED_NAME_DOMAIN>**BIOLOGIC**</ESTABLISHED_NAME_DOMAIN>
</ESTABLISHED_NAME_GROUP>
<REFERENCE_SOURCE_GROUP>
<REFERENCE_SOURCE_CLASS>**PUBLIC**</REFERENCE_SOURCE_CLASS>
<REFERENCE_SOURCE_TYPE/>
<REFERENCE_SOURCE_CITATION>**PACKAGE INSERT**</REFERENCE_SOURCE_CITATION>
</REFERENCE_SOURCE_GROUP>
</NAME_GROUP>
<NAME_GROUP>
<NAME>**BLACK WIDOW SPIDER ANTIVENIN (EQUINE)**</NAME>

<NAME_TYPE>**COMMON**</NAME_TYPE>
<LANGUAGE_ISOCODE>**ENG**</LANGUAGE_ISOCODE>
<ESTABLISHED_NAME_GROUP>
<ESTABLISHED_NAME_TYPE/>
<ESTABLISHED_NAME_STATUS/>
<ESTABLISHED_NAME_STATUS_CHANGE_DATE/>
<ESTABLISHED_NAME_JURISDICTION_GROUP>
<ESTABLISHED_NAME_JURISDICTION>**US**</ESTABLISHED_NAME_JURISDICTION>
</ESTABLISHED_NAME_JURISDICTION_GROUP>
<ESTABLISHED_NAME_DOMAIN/>
</ESTABLISHED_NAME_GROUP>
<REFERENCE_SOURCE_GROUP>
<REFERENCE_SOURCE_CLASS>**PUBLIC**</REFERENCE_SOURCE_CLASS>
<REFERENCE_SOURCE_TYPE/>
<REFERENCE_SOURCE_CITATION>**PACKAGE INSERT**</REFERENCE_SOURCE_CITATION>
</REFERENCE_SOURCE_GROUP>
</NAME_GROUP>
<CODE_GROUP>
<CODE/>
<CODE_SYSTEM/>
<CODE_SYSTEM_ID/>
<CODE_STATUS/>
<CODE_STATUS_CHANGE_DATE/>
<CODE_REFERENCE_CITATION/>
</CODE_GROUP>
<SINGLE_SUBSTANCE>
<STRUCTURE>
<STRUCTURAL_REPRESENTATION_TYPE/>
<STRUCTURAL_REPRESENTATION/>
<STEREOCHEMISTRY/>
<OPTICAL_ACTIVITY/>
<ISOTOPE_GROUP>
<NUCLIDE_ID/>
<NUCLIDE_NAME/>
<SUBSTITUTION_TYPE/>
</ISOTOPE_GROUP>
</STRUCTURE>
<REFERENCE_INFORMATION>
<CLASSIFICATION_GROUP>
<SUBSTANCE_CLASS>
PROTEIN
<TYPE>
IMMUNOSERA
<SUBTYPE/>
</TYPE>
</SUBSTANCE_CLASS>
</CLASSIFICATION_GROUP>
<TARGET_GROUP>
<TARGET_GENE_ID/>
<TARGET_NAME/>
<INTERACTION_TYPE/>
<TARGET_ORGANISM_TYPE/>
<TARGET_REFERENCE_SOURCE/>
</TARGET_GROUP>

```
  <COMMENTS/>
    </REFERENCE_INFORMATION>
<STRUCTURALLY_DIVERSE>
<SOURCE_MATERIAL>
  <SOURCE_MATERIAL_CLASS>ORGANISM</SOURCE_MATERIAL_CLASS>
  <SOURCE_MATERIAL_TYPE>MAMMAL</SOURCE_MATERIAL_TYPE>
  <SOURCE_MATERIAL_STATE/>
  <ORGANISM_NAME_ID>XXXXXXXXX</ORGANISM_NAME_ID>
  <ORGANISM_NAME>EQUUS CABALLUS</ORGANISM_NAME>
  <DEVELOPMENTAL_STAGE/>
<PART_GROUP>
  <PART>BLOOD</PART>
  <PART_LOCATION/>
    </PART_GROUP>
<FRACTION>
  <MATERIAL_TYPE/>
  <FRACTION>SERA</FRACTION>
    </FRACTION>
<FRACTION>
  <MATERIAL_TYPE>PROTEIN</MATERIAL_TYPE>
  <FRACTION>IMMUNOGLOBULINS</FRACTION>
    </FRACTION>
<MODIFICATION_GROUP>
  <MODIFICATION_TYPE>BIOLOGICAL</MODIFICATION_TYPE>
  <RESIDUE_MODIFIED/>
  <RESIDUE_SITE/>
<PHYSICAL_MODIFICATION>
  <ROLE/>
  <NUMBER_OF_PARAMETERS/>
<PARAMETER_GROUP>
  <PARAMETER/>
<AMOUNT>
  <AVERAGE/>
  <LOW_LIMIT/>
  <HIGH_LIMIT/>
  <UNIT/>
  <NON_NUMERIC_VALUE/>
    </AMOUNT>
    </PARAMETER_GROUP>
    </PHYSICAL_MODIFICATION>
<STRUCTURAL_MODIFICATION>
  <STRUCTURAL_MODIFICATION_TYPE/>
<MOIETY>
  <ROLE/>
  <MOIETY_ID/>
  <AMOUNT_TYPE/>
<AMOUNT>
  <AVERAGE/>
  <LOW_LIMIT/>
  <HIGH_LIMIT/>
  <UNIT/>
  <NON_NUMERIC_VALUE/>
    </AMOUNT>
    </MOIETY>
```

```
<MOLECULAR_FRAGMENT>
 <ROLE/>
 <MOLECULAR_FRAGMENT_ID/>
 <AMOUNT_TYPE/>
<AMOUNT>
 <AVERAGE/>
 <LOW_LIMIT/>
 <HIGH_LIMIT/>
 <UNIT/>
 <NON_NUMERIC_VALUE/>
   </AMOUNT>
   </MOLECULAR_FRAGMENT>
   </STRUCTURAL_MODIFICATION>
<AGENT_MODIFICATION>
 <AGENT_MODIFICATION_TYPE>IMMUNOLOGICAL</AGENT_MODIFICATION_TYPE>
 <ROLE>TARGETED ANTIGEN</ROLE>
 <MODIFICATION_AGENT>LATRODECTUS MACTANS VENOM</MODIFICATION_AGENT>
 <MODIFICATION_AGENT_ID>TI41393577</MODIFICATION_AGENT_ID>
<AMOUNT>
 <AVERAGE/>
 <LOW_LIMIT/>
 <HIGH_LIMIT/>
 <UNIT/>
 <NON_NUMERIC_VALUE/>
   </AMOUNT>
 <MODIFICATION_PROCESS/>
   </AGENT_MODIFICATION>
   </MODIFICATION_GROUP>
   </SOURCE_MATERIAL>
<PROPERTY_GROUP>
 <PROPERTY_TYPE/>
 <PROPERTY/>
 <SUBSTANCE_ID/>
<AMOUNT>
 <AVERAGE/>
 <LOW_LIMIT/>
 <HIGH_LIMIT/>
 <UNIT/>
 <NON_NUMERIC_VALUE/>
   </AMOUNT>
   </PROPERTY_GROUP>
  <COMMENTS/>
   </STRUCTURALLY_DIVERSE>
   </SINGLE_SUBSTANCE>
   </SUBSTANCE>
```

B.2.6 Advanced Cellular Therapy Substance

The following is an example of an element by element data model in XML for an advanced therapy substance that USAN has assigned the name SIPULEUCEL-T. It describes autologous PMBC that have been

stimulated ex-vivo with antigen fused to a growth factor. The data model is best viewed in an xml viewer. This represents a possible implementation of the standard.

<SUBSTANCE>
<SUBSTANCE_ID>**8Q622VDR18**</SUBSTANCE_ID>
<NAME_GROUP>
<NAME>**SIPULEUCEL-T**</NAME>
<NAME_TYPE>**ESTABLISHED NAME**</NAME_TYPE>
<LANGUAGE_ISOCODE>**ENG**</LANGUAGE_ISOCODE>
<ESTABLISHED_NAME_GROUP>
<ESTABLISHED_NAME_TYPE>**USAN**</ESTABLISHED_NAME_TYPE>
<ESTABLISHED_NAME_STATUS>**CURRENT**</ESTABLISHED_NAME_STATUS>
<ESTABLISHED_NAME_STATUS_CHANGE_DATE/>
<ESTABLISHED_NAME_DOMAIN>**BIOLOGIC**</ESTABLISHED_NAME_DOMAIN>
<ESTABLISHED_NAME_JURISDICTION_GROUP>
<JURISDICTION>**USA**</JURISDICTION>
</ESTABLISHED_NAME_JURISDICTION_GROUP>
</ESTABLISHED_NAME_GROUP>
<REFERENCE_SOURCE_GROUP>
<REFERENCE_SOURCE_TYPE>**PUBLIC**</REFERENCE_SOURCE_TYPE>
<REFERENCE_SOURCE_CODE>**USAN**</REFERENCE_SOURCE_CODE>
<REFERENCE_SOURCE_CITATION>**USP DICTIONARY 2009**</REFERENCE_SOURCE_CITATION>
</REFERENCE_SOURCE_GROUP>
</NAME_GROUP>
<NAME_GROUP>
<NAME>**APC8015**</NAME>
<NAME_TYPE>**COMPANY CODE**</NAME_TYPE>
<LANGUAGE_ISOCODE>**ENG**</LANGUAGE_ISOCODE>
<ESTABLISHED_NAME_GROUP>
<ESTABLISHED_NAME_TYPE/>
<ESTABLISHED_NAME_STATUS/>
<ESTABLISHED_NAME_STATUS_CHANGE_DATE/>
<ESTABLISHED_NAME_DOMAIN/>
<ESTABLISHED_NAME_JURISDICTION_GROUP>
<JURISDICTION/>
</ESTABLISHED_NAME_JURISDICTION_GROUP>
</ESTABLISHED_NAME_GROUP>
<REFERENCE_SOURCE_GROUP>
<REFERENCE_SOURCE_TYPE>**PUBLIC**</REFERENCE_SOURCE_TYPE>
<REFERENCE_SOURCE_CODE>**USAN**</REFERENCE_SOURCE_CODE>
<REFERENCE_SOURCE_CITATION>**USP DICTIONARY 2009**</REFERENCE_SOURCE_CITATION>
</REFERENCE_SOURCE_GROUP>
</NAME_GROUP>
<CODE_GROUP>
<CODE>**917381-47-6**</CODE>
<CODE_SYSTEM>**CAS**</CODE_SYSTEM>
<CODE_SYSTEM_ID/>
<CODE_REFERENCE_CITATION>**USP DICTIONARY 2009**</CODE_REFERENCE_CITATION>
<CODE_STATUS>**CURRENT**</CODE_STATUS>
<CODE_STATUS_CHANGE_DATE/>
</CODE_GROUP>
<SINGLE_SUBSTANCE>
<STRUCTURE>
<STRUCTURAL_REPRESENTATION/>

<STEREOCHEMISTRY/>
<OPTICAL_ACTIVITY/>
<ISOTOPE_GROUP>
<NUCLIDE_ID/>
<NUCLIDE_NAME/>
<SUBSTITUTION_TYPE/>
</ISOTOPE_GROUP>
</STRUCTURE>
<REFERENCE_INFORMATION>
<CLASSIFICATION_GROUP>
<SUBSTANCE_CLASS>
THERAPEUTIC VACCINE
<TYPE>
AUTOLOGOUS CELLULAR
<SUBTYPE>**ANTIGEN STIMULATED**</SUBTYPE>
</TYPE>
</SUBSTANCE_CLASS>
</CLASSIFICATION_GROUP>
<TARGET_GROUP>
<TARGET_ORGANISM>**HUMAN**</TARGET_ORGANISM>
<TARGET_GENE_ID>**55**</TARGET_GENE_ID>
<TARGET_NAME>**ACPP ACID PHOSPHATASE, PROSTATE (HOMO SAPIENS)**
</TARGET_NAME>
<TARGET_REFERENCE_SOURCE>**EXPERT REV. ANTICANCER THER. 6:1163-1167 (2006)**</TARGET_REFERENCE_SOURCE>
</TARGET_GROUP>
<TARGET_GROUP>
<TARGET_ORGANISM>**HUMAN**</TARGET_ORGANISM>
<TARGET_GENE_ID/>
<TARGET_NAME>**HUMAN PROSTATE TUMOR CELLS**</TARGET_NAME>
<TARGET_REFERENCE_SOURCE>**EXPERT REV. ANTICANCER THER. 6:1163-1167 (2006)**</TARGET_REFERENCE_SOURCE>
</TARGET_GROUP>
<COMMENTS/>
</REFERENCE_INFORMATION>
<STRUCTURALLY_DIVERSE>
<SOURCE_MATERIAL>
<SOURCE_MATERIAL_CLASS>**ORGANISM**</SOURCE_MATERIAL_CLASS>
<SOURCE_MATERIAL_TYPE>**HUMAN AUTOLOGOUS**</SOURCE_MATERIAL_TYPE>
<SOURCE_MATERIAL_STATE>**LIVE**</SOURCE_MATERIAL_STATE>
<ORGANISM_NAME_ID>**ITIS TSN 180092**</ORGANISM_NAME_ID>
<ORGANISM_NAME>**HOMO SAPIENS**</ORGANISM_NAME>
<PART>
BLOOD
<PART_LOCATION/>
</PART>
<DEVELOPMENTAL_STAGE/>
<FRACTION_GROUP>
<MATERIAL_TYPE>**CELL**</MATERIAL_TYPE>
<FRACTION>**PERIPHERAL BLOOD MONONUCLEAR**</FRACTION>
</FRACTION_GROUP>
<MODIFICATION_GROUP>
<MODIFICATION_TYPE>**AGENT**</MODIFICATION_TYPE>
<RESIDUE_MODIFIED/>

<RESIDUE_SITE/>
<PHYSICAL_MODIFICATION>
<ROLE/>
<NUMBER_OF_PARAMETER/>
<PARAMETER_GROUP>
<PARAMETER/>
<AMOUNT>
<AVERAGE/>
<LOW_LIMIT/>
<HIGH_LIMIT/>
<UNIT/>
<NON_NUMERIC_VALUE/>
</AMOUNT>
</PARAMETER_GROUP>
</PHYSICAL_MODIFICATION>
<STRUCTURAL_MODIFICATION>
<STRUCTURAL_MODIFICATION_TYPE/>
<MOIETY>
<ROLE/>
<MOIETY_ID/>
<AMOUNT>
<AVERAGE/>
<LOW_LIMIT/>
<HIGH_LIMIT/>
<UNIT/>
<NON_NUMERIC_VALUE/>
</AMOUNT>
</MOIETY>
<MOLECULAR_FRAGMENT>
<ROLE/>
<MOLECULAR_FRAGMENT_ID/>
<AMOUNT>
<AVERAGE/>
<LOW_LIMIT/>
<HIGH_LIMIT/>
<UNIT/>
<NON_NUMERIC_VALUE/>
</AMOUNT>
</MOLECULAR_FRAGMENT>
</STRUCTURAL_MODIFICATION>
<AGENT_MODIFICATION>
<AGENT_MODIFICATION_TYPE>**IMMUNOLOGICAL**</AGENT_MODIFICATION_TYPE>
<ROLE>**ANTIGEN; IMMUNOSTIMULANT**</ROLE>
<MODIFICATION_AGENT>**PROSTATIC ACID PHOSPHATASE (PAP)-GRANULOCYTE-MACROPHAGE COLONY-STIMULATING FACTOR (GM-CSF) FUSION**</MODIFICATION_AGENT>
<MODIFICATION_AGENT_ID>**N5E5Q8249O**</MODIFICATION_AGENT_ID>
<AMOUNT>
<AVERAGE/>
<LOW_LIMIT/>
<HIGH_LIMIT/>
<UNIT/>
<NON_NUMERIC_VALUE/>
</AMOUNT>

```
 <ROLE>ANTIGEN LOADING</ROLE>
 <MODIFICATION_PROCESS>CELL CULTURE</MODIFICATION_PROCESS>
 <MODIFICATION_PROCESS_ID>A</MODIFICATION_PROCESS_ID>
   </AGENT_MODIFICATION>
   </MODIFICATION_GROUP>
   </SOURCE_MATERIAL>
<PROPERTY_GROUP>
 <PROPERTY_TYPE/>
 <PROPERTY/>
 <SUBSTANCE_ID/>
<AMOUNT>
 <AVERAGE/>
 <LOW_LIMIT/>
 <HIGH_LIMIT/>
 <UNIT/>
 <NON_NUMERIC_VALUE/>
   </AMOUNT>
   </PROPERTY_GROUP>
   </STRUCTURALLY_DIVERSE>
   </SINGLE_SUBSTANCE>
   </SUBSTANCE>
```

Bibliography

[1] International Conference on Harmonization (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use Draft Consensus Guideline – Data Elements and Standards for Drug Dictionaries – M5 Revision 4, 2 February 2007. http:/www.fda.gov/Cder/guidance/6886dft.pdf

[2] Resources related to FDA's Substance Registration system http://www.fda.gov/oc/datacouncil/SRS.htm

[3] For information on NCI Thesaurus (NCIt) and other NCI Enterprise Vocabulary Services (EVS) resources and documentation: http://nciterms.nci.nih.gov/NCIBrowser/Dictionary.do

[4] The Integrated Taxonomic Information System http://www.itis.gov/

[5] The Food and Drugs Act and Regulations and related Health Canada Guidelines: http://www.hc-sc.gc.ca/hpfb-dgpsa/tpd-dpt/index_e.html

[6] The Natural Health Product Regulations and related Health Canada Guidelines: http://www.hc-sc.gc.ca/hpfb-dgpsa/nhpd-dpsn/index_e.html

[7] Directive 2001/83/EC of the European Parliament and of the Council of 6 November 2001 on the Community code relating to medicinal products for human use as amended by Directive 2002/98/EC, Directive 2004/24/EC and Directive 2004/27/EC http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol-1/consol_2004/human_code.pdf

[8] "Guideline on the categorisation on New applications versus Variations", The Rules governing Medicinal Products in the European Union, Notice to Applicants, Volume 2C

[9] The Rules governing Medicinal Products in the European Union, Notice to Applicants, Volume 2A and Volume 2C http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/homev2.htm

[10] A Guideline on Summary of Product Characteristics, December 1999, (Doc. Ref. Notice to Applicants, Final – revision 0) http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol-2/c/spcguidrev1-oct2005.pdf

[11] Directive 2001/20/EC of the European Parliament and of the Council of 4 April 2001 on the approximation of the laws, regulations and administrative provisions of the Member States relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol-1/dir_2001_20/dir_2001_20_en.pdf

[12] Commission Directive 2003/63/EC of 25 June 2003 amending Directive 2001/83/EC of the European Parliament and of the Council on the Community code relating to medicinal products for human use http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol-1/dir_2003_63/dir_2003_63_en.pdf

[13] Regulation (EC) No 726/2004 of the European Parliament and of the Council of 31 March 2004 laying down Community procedures for the authorisation and supervision of medicinal products for human and veterinary use and establishing a European Medicines Agency http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol-1/reg_2004_726/reg_2004_726_en.pdf

[14] Regulation (EC) No 1394/2007 of the European Parliament and of the Council of 13 November 2007 on advanced therapy medicinal products and amending Directive 2001/83/EC and Regulation (EC) No 726/2004 http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol-1/reg_2007_1394/reg_2007_1394_en.pdf

[15] Commission Directive 2003/94/EC of 8 October 2003 laying down the principles and guidelines of good manufacturing practice in respect of medicinal products for human use and investigational medicinal products for human use.

[16] http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol-1/dir_2003_94/dir_2003_94_en.pdf

[17] Guidelines on Pharmacovigilance for Medicinal Products for Human Use, Volume 9A of the Rules Governing Medicinal Products in the European Union

[18] http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol-9/pdf/vol9_2007-07_upd07.pdf

[19] Guideline on declaration of botanical substances and botanical preparations in botanical medicinal products / traditional

[20] Botanical medicinal products in the SPC, 26 July 2007 (Doc. Ref. EMEA/HMPC/CHMP/CVMP/287539/2005)

[21] http://www.emea.europa.eu/pdfs/human/hmpc/28753905en2.pdf

[22] Guideline on adjuvants in vaccines for human use, 20 January 2005 (Doc. Ref. EMEA/CHMP/VEG/134716/2004)

[23] http://www.emea.europa.eu/pdfs/human/vwp/13471604en.pdf

[24] Guideline on Similar Biological Medicinal Products, 30 October 2005 (Doc. Ref. CHMP/437/04)

[25] http://www.emea.europa.eu/pdfs/human/biosimilar/043704en.pdf

[26] Guideline on Pharmaceutical Aspects of the Product Information for Human Vaccines, 26 November 2003 (Doc. Ref. EMEA/CPMP/BWP/2758/02) http://www.emea.europa.eu/pdfs/human/bwp/275802en.pdf

[27] Guideline on the acceptability of names for human medicinal products processed through the centralised procedure, 11 December 2007 (Doc. Ref. CPMP/328/98, Revision 5)

[28] http://www.emea.europa.eu/pdfs/human/regaffair/032898en.pdf

[29] Guideline on the Chemistry of new Active Substances, 17 December 2003 (Doc. Ref. CPMP/QWP/130/96 Rev. 1) http://www.emea.europa.eu/pdfs/human/qwp/013096en.pdf

[30] Standard Terms: Dosage Forms, Routes of Administration and Containers, EDQM, Fifth Edition, December 2004, Version 5.0.0 http://st.pheur.org/entry.htm

[31] EudraVigilance Medicinal Product Dictionary (EVMPD) Version 2.0 Technical Specifications, 9 November 2004 (Doc. Ref. EMEA/140190/2004) http://eudravigilance.emea.europa.eu/human/docs/EVMPD%20Technical%20Specifications.pdf

[32] EudraVigilance Medicinal Product Dictionary (EVMPD) Version 2.0 Message and Acknowledgement Specifications, 8 December 2004 (Doc. Ref. EMEA/178966/2004) http://eudravigilance.emea.europa.eu/human/docs/EVMPD%20Message%20and%20%20Acknowledgement%20Description.pdf

[33] EudraVigilance (EV) Access simple Database Version 2.0 8 November 2004 (Doc. Ref: EMEA/140327/2004) http://eudravigilance.emea.europa.eu/human/docs/EV%20Access%20Simple%20Database%202.0%20-%20Tables%20Documentation.pdf

[34] EudraVigilance (EV) Access Simple Database Version 2.0 Forms Documentation, 31 January 2005, (Doc. Ref: EMEA/35416/2005) http://eudravigilance.emea.europa.eu/human/docs/EVAccessSimple%20Database_v20_Formsdocumentation.pdf

[35] EudraVigilance (EV) Access Simple Database Version 2.0 Step by Step Guide 8 December 2004, (Doc. Ref: EMEA/191986/2004) http://eudravigilance.emea.eu.int/human/docs/EV%20Access%20Simple%20Database%202.0%20-%20Step%20by%20Step%20Guide.pdF

[36] Product Information Management http://pim.emea.europa.eu/

[37] European Pharmacopeia (Ph. Eur.): http://online.pheur.org/entry.htm

[38] Ministry of Health and Welfare, PSB/SD Notification No.37 (29th March 1999) (All documents are available only in Japanese, Please note documents' names are only tentative translation) Electronic package insert information accompanied by the Utility system of Information provision on drug safety https://skw.info.pmda.go.jp/notice/nt990329.pdf

[39] Appendix: Electronic format of package insert information on ethical drugs (SGML/DTD Ver2.0) Appendix1 DTD Ver.2.0 https://skw.info.pmda.go.jp/notice/nt990329dtd.zip

[40] Appendix2 Overview of SGML creation https://skw.info.pmda.go.jp/notice/nt990329ol.pdf

[41] Appendix3 DCL https://skw.info.pmda.go.jp/notice/nt990329dcl.zip

[42] Appendix4 Template of SGML https://skw.info.pmda.go.jp/notice/nt990329smp.zip

[43] Appendix5 Data model (Entities and relationships) https://skw.info.pmda.go.jp/notice/nt990329dm.pdf

[44] Appendix6 List of fields of package insert information https://skw.info.pmda.go.jp/notice/nt990329lst.pdf

[45] *the Japanese Pharmacopoeia Fifteenth Edition, http://jpdb.nihs.go.jp/jp15/

[46] United States Pharmacopoeia (USP): http://www.usp.org/USPNF/faq.html#III_Q3

[47] United States Department of Agriculture's (USDA) Integrated Taxonomic Information System (ITIS) http://www.it is.usda.gov

[48] Guidance for Industry Providing Regulatory Submissions in Electronic Format — Content of Labelling http://www.fda.gov/cder/regulatory/guidance

[49] Release Notes for SPL Schema PORR_MT050020 (3.20.05) FDA SPL Schema for Implementation http://www.fda.gov/oc/datacouncil/spl.html

[50] CaCore 2.0 Technical Guide, National Cancer Institute, Center for Bioinformatics, U.S. Department of Health and Human Services ftp://ftp1.nci.nih.gov/pub/cacore/caCORE2.0_Tech_Guide.pdf

[51] A guide to RXNorm, United States National Library of Medicine, National Institute of Health http://www.nlm.nih.gov/research/umls/rxnorm_guide.pdf

[52] The Unified Code for Units of Measure http://www.fda.gov/oc/datacouncil/term.html

[53] Substance Registration System (SRS) SRS Substance Definition Manual Version 5b Final Draft.doc

[54] The U.S. Consolidated Healthcare Informatics initiative http://ushik.ahrq.gov/