

Food and Drug Administration Substance Registration System Standard Operating Procedure

Title: Substance Definition Manual¹	Version: 5c
Effective Date: June 10, 2007	Supersedes: Version 5b
Approved By: SRS Chemistry Review Committee; FDA Data Standards Council's Vocabulary Standards Working Group	

1. Purpose	1
a. Defining a Substance	1
b. Modification of Text for Computer Use	1
2. Structure Set Up and Layout	3
i. Example: ISOETHARINE	4
3. Molecules Containing Chains	4
i. Example: IOPHENDYLATE	4
4. Abbreviations	4
i. Example: ACETOPHENONE	4
ii. Example: TRIFLOCIN	5
5. Functional Groups	5
a. NITRO Groups and NITRATES	5
i. Example: NITROBENZENE	5
ii. Example: NITRATION	5
b. NITROGEN OXIDES	6
i. Example: PYRIDINE 1-OXIDE	6
c. OXIMES	6
i. Example: 2-PROPANONE, O-METHYLOXIME	6
d. AZIDES	6
i. Example: 4-AZIDOBENZAMINE	6
e. DIAZO and AZO Compounds	6
i. Example: 5-DIAZOURACIL	7
ii. Example: D&C Red No. 33	7
f. CYANIDES and ISOCYANIDES	7
i. Example: BENZOYL CYANIDE	7
ii. Example: t-BUTYLISOCYANIDE	8
g. SULFOXIDES	8
i. Example: ESOMEPRAZOLE	8
6. Mixtures and Salts	8
a. Types of Salts	9
i. Simple Salt Example: MAGNESIUM HYDROXIDE	9
ii. Quaternary Ammonium Compound Example: TETRAETHYLAMMONIUM BROMIDE	9
iii. Salt Formed by Reaction between Organic Base and an Acid Example: GLYCINE HYDROCHLORIDE	9
b. Types of Mixtures	9
i. Example of an ALL OF Mixture: AMYL NITRILE	9
7. Salts and Compounds with Ionic Bonds	10
i. Example: SODIUM ACETATE	10
ii. Example: POTASSIUM CHLORIDE	10
iii. Example: MAGNESIUM PHOSPHATE	10
8. Salts Formed by the Reaction of Ammonia with Acids	11
i. Example: AMMONIUM CHLORIDE	11
ii. Example: AMMONIUM LACTATE	11

¹ This guide was formerly titled the *Structure Drawing Guide*.

9. Salts Formed by the Reaction of Nitrogenous Organic Bases with Acids.....	11
i. Example: TRIETHYLAMINE HYDROCHLORIDE	11
ii. Example: AMINOGUANIDINE SULFATE	12
iii. Example: TETRAETHYLAMMONIUM CHLORIDE	12
10. Inner Salts.....	12
i. Example: R-5-CHLORO-2-METHYL-3-(4-SULFOBUTYL)-BENZOTHAZOLIUM INNER SALT 12	
11. Bracket Usage for Salts and Solvates	13
i. Example: 1:1 Compound to Salt Ratio	13
ii. Example: 1:2 Compound to Salt Ratio	13
iii. Example: LOSOXANTRONE HYDROCHLORIDE (DIHYDOCHLORIDE HEMIHYDRATE)	13
12. Salts Containing a Complex Substance as a Counterion	14
i. Example: FENETHAZINE TANNATE.....	14
13. Substance Of Unknown Or Variable Stoichiometry.....	14
i. Example: VIOMYCIN SULFATE.....	15
14. Partially Ionized Acids	15
i. Example: 5-SULFOSALICYLATE, MONOSODIUM SALT.....	16
ii. Example: Citrate Ion with Single Negative	16
iii. Example: MONOSODIUM CITRATE	17
iv. Example: MONOSODIUM GLUTAMATE	17
v. Example: 1H-IMIDAZOLE-4-METHANOL, DIHYDROGEN PHOSPHATE (ESTER), DISODIUM SALT	17
vi. Example: ALENDRONATE SODIUM	17
15. Complex Inorganic and Organic Compounds	18
a. Complex Inorganics.....	19
i. Example: TALC	19
b. Complex Organics.....	19
16. Botanicals	19
i. Example: FENNEL OIL.....	20
17. Foods	20
i. Example ABALONE <FOOD>.....	22
ii. Example BEEF	22
iii. Example SQUID, BOILED	23
18. Coordination compounds or complexes (Draft Proposed).....	23
i. Example: PLATINATE(2-), TETRACHLORO-, DIPOTASSIUM , (SP-4-1)-.....	24
ii. Example: CISPLATIN and TRANSPLATIN	24
iii. Example: DEXORMAPLATIN	24
iv. Example: GADOLINATE(2-), (N,N-BIS)2-(BIS(((CARBOXY-.KAPPA.O)METHYL)AMINO-.KAPPA.N)ETHYL)GLYCINATO(5-)-.KAPPA.N, KAPPA.O)-, DISODIUM	25
v. Example: IRON SUCROSE (Proposed Not Official)	25
19. Organometalic Compounds	26
i. Example: 2-.BETA.-CARBOMETHOXY-3-.BETA.-(4-TRIMETHYLSTANNYLPHENYL)TROPANE	26
a. Nitrogen, Oxygen, and Fluorine Bonded to Metallic Centers.....	26
i. Example: TRIPHENYL TIN HYDROXIDE	26
b. Other Nonmetals Bonded to Metallic Centers	26
i. Example: THIMEROSAL	26
20. Tautomerization and Resonance Form Presentation.....	27
a. Proton-Shift Tautomerization	27
b. Ketones.....	27
i. Example: ACETOPHENONE.....	27
c. Carbonyl Compounds	27
i. Example: PHENOBARBITAL	27
d. Phenols	27

i.	Example: PHENOL	28
e.	Oximes	28
i.	Example: BENZALDOXIME	28
f.	Guanidines	28
i.	Example: 1-PHENYL-3-(4-PHENYL-2-THIAZOLYL) GUANIDINE	28
ii.	Example: 1-PHENYL-3-(4-PHENYL-2-THIAZOLYL)GUANIDINE	29
iii.	Example: TRAMAZOLINE	29
g.	Tetrazoles	29
i.	Example: 5-(3-PYRIDYL)TETRAZOLE	29
h.	Pyrithione	29
i.	Thiosulfate and Thiosulfite	29
i.	Example: SODIUM THIOSULFATE	30
j.	Purines and Pyrimidines	30
i.	Example: 6-THIOPURINE	30
ii.	Example: 2-THIOURACIL	30
k.	Tetracyclines	30
i.	Example: TETRACYCLINE	31
l.	Anhydrotetracyclines	31
i.	Example: CHELOCARDIN HYDROCHLORIDE	31
m.	Rubicin Antibiotics	31
i.	Example: ANNAMYCIN	32
n.	Resonance-Stabilized Ions	32
i.	Example. PEMIROLAST POTASSIUM	32
ii.	Example: PHENOBARBITAL SODIUM	33
o.	Other Types of Tautomerization	33
p.	Phenolphthalein and Similar Compounds	33
i.	Example: PHENOLPHTHALEIN	33
ii.	Example: PHENOLPHTHALEIN, SODIUM	34
21.	Stereochemistry	34
a.	E and Z Isomerism (Proposed Not Official)	34
i.	Example: NITROFURAZONE	35
ii.	Example: RIFAPENTINE	35
iii.	Example CROTAMITON	35
b.	General Rules for Tetrahedral Stereochemistry	36
i.	Example: 4-(R-2-(ETHYLAMINO)-1-HYDROXYETHYL)BENZENE-1,2-DIOL	37
ii.	Example: PROPOXPHENE	37
iii.	Example: MEPIVACAINE	37
iv.	Example: SANTONIN	38
v.	Example: METHYLPHENIDATE	38
22.	Representation of Stereochemical Classes	38
a.	Completely Defined Enantiomer	39
i.	Example: PREDNISOLONE	39
b.	Racemic Substances (Optically Inactive) or Substances with Undefined or Unknown Stereochemistry	39
1.	A single chiral center	39
i.	Example: ISOPROTERENOL HYDROCHLORIDE	40
ii.	Example: DROFENINE	40
iii.	Example DOXPICOMINE (-)-isomer absolute configuration undefined.	40
2.	Multiple chiral centers	41
i.	Example: PENTAZOCINE HYDROCHLORIDE	41
ii.	Example: BENZQUINAMIDE	41
iii.	Example: SARPICILLIN HYDROCHLORIDE	42
iv.	Example: ENISOPROST	42
v.	Example: HEXESTROL (Meso compound)	43
c.	Chiral Substances (Optically Active) with Undetermined Configuration	44
i.	Example: LEVOPHACETOPERANE HYDROCHLORIDE	44

d. Spiro Compounds	44
i. Example: HEPTELIDIC ACID	44
e. Special Stereochemical Cases	45
i. Example: (+)-CAMPHENE	46
ii. Example: (+)-CAMPHOR	46
iii. Example: PENTAZOCINE LACTATE	46
f. Axial chirality	47
i. Example (S)-1,3-DIMETHYLALLENE	47
ii. Example ADENALLENE (+/-)	48
23. Isotopically Labeled Compounds	48
a. Position of Substitution Known	48
i. Example: IODOANTIPYRINE I	48
b. Labeled compound with the position of isotopic atom is unknown	48
i. Example: ROSE BENGAL SODIUM, I-131, Position of Substitution Not Specified	49
ii. Example: C-14 CHOLESTEROL; Position of Substitution Not Specified	49
iii. Example: DEUTERIUM OXIDE	49
24. Substituted Glycerine Compounds (Guidance needs to be developed)	49
i. Example: CALCIUM GLYCEROPHOSPHATE	49
25. Carbohydrates and Other Saccharides	50
i. Example: DEXTROSE	50
ii. Example: D-FRUCTOSE	50
iii. Example LACTOSE, ANHYDROUS	51
26. Peptides/Proteins	51
a. Description Field	52
b. Structure Field and Complete Polypeptide/Protein Substance Definitions	53
i. Example: COSYNTROPIN	53
ii. Example: BIVALIRUDIN	54
iii. Example: ENFUVIRTIDE	55
iv. Example: LUTEINIZING HORMONE	55
v. Example: INSULIN HUMAN	56
vi. Example: SOMATROPIN	57
vii. Example: EFALIZUMAB	57
viii. Example: PEGINTERFERON ALFA-2A	59
ix. Example. INFLIXIMAB	60
x. Example: TOSITUMOMAB I-131	61
c. Salts of peptides and proteins:	63
i. Example: INSULIN HUMAN ZINC	63
ii. Example: GONADORELIN ACETATE	64
27. Oligonucleotides (Proposed Not official)	65
i. Example: DNA sequence d(AGTC)	66
ii. Example: FOMIVIRSEN SODIUM	66
28. Polymers (Proposed Not Official)	67
a. Structure Drawing	67
b. Description Field	67
i. Example. Low Density POLYETHYLENE	69
ii. Example. High Density POLYETHYLENE	70
iii. Example. MICROCRYSTALLINE CELLULOSE	71
iv. Example. POWDERED CELLULOSE	72
v. Example. CELLULOSE ACETATE	74
vi. Example. CELLACEFATE	75
vii. Example. POLOXAMER 124	77
viii. Example. ACRYLONITRILE- BUTADIENE-STYRENE TERPOLYMER	78
ix. Example: PHENOL FORMALDEHYDE RESINS (Condensation polymers)	80
x. Example: NAPTHOL FORMALDEHYDE RESIN	81
xi. Example: 6,6-NYLON	83
xii. Example: HEPARIN (Not Official)	83

xiii. Example: POLACRILIN	86
29. Active Moieties	87
i. Example: ALENDRONATE SODIUM active moiety is ALENDRONIC ACID.....	88
ii. Example: CHELOCARDIN HYDROCHLORIDE active moiety is CHELOCARDIN	88

1. Purpose

This guide is used to standardize the entry of substances into the Food and Drug Administration (FDA) Substance Registration System (SRS). The primary purpose of this guide is to prevent duplicate entries of a single substance. Conventions for drawing structures and for organizing the characteristics of substances are included. The guide also provides limited aesthetic guidelines for the structures as they are intended to be shared with other databases and may be used in professional publications.

a. Defining a Substance

In the SRS, a substance is defined using two fields: Structure and Identifying Description. The preferred method to define a structure is using a two-dimensional chemical structure. Chemical structures are easily searched and stereochemistry is readily apparent to the user. Many substances cannot be described completely or at all by a single structure or sets of structures; they require other forms of representation. The Identifying Description Field is used for substances that are not easily defined using structures. Data is captured in the Identifying Description Field using a hierarchy of tags. Different classes of substances, such as polymers and peptides, have specific sets of tags used to describe substances. These tags are explained in detail under the topics in which they apply. Tags are also used to provide information that allows the classification of substances that cannot be completely described by the structure alone.²

For chemical substances, the FDA SRS differentiates substances at the molecular level. Different salts and solvates of a drug are defined as different substances. However, different tautomeric forms are considered the same substance. This guide includes conventions for standardizing the representation of tautomers in order to provide the user with a consistent view and to aid in searching. The SRS allows the registration of stereoisomers through the use of wedge bonds, the chiral flag and Identifying Description tags. Duplicate structures are permitted in the SRS only if the Identifying Description Fields differentiate the substances.

Substances can be mixtures of other substances. Mixtures that contain greater than ten chemically defined substances are considered complex substances and chemical structures are not drawn. Substances that form non-covalent interactions with other added substances are not new substances or mixtures of substances; they are defined as separate substances. Mixtures are either related compounds or compounds that are isolated together. Examples: Beta-cyclodextrin complexed with nicotine is described in the SRS system as two separate substances and is not considered a new substance (nor is it treated as a mixture). Synthetic conjugated estrogens are considered a mixture since the substances are closely related to one another.

The Description Field is used to add additional information about a substance or to describe substances for which a structure is not drawn. The Description Field is tagged with **xml** tags that are defined throughout the guide.³ The Description Field is part of a substance definition and the combination of Structure and Identifying Description must be unique for a given substance. If multiple sets of description tags are used to define a substance the order of the set of tags are alphabetical according to the primary tag of the set (i.e. <POLYMER> tags precede <STEREOCHEMISTRY> tags). The primary tag is the initial tag that defines the type of information defined USING xml tags. The following is a list of primary tags:

BOTANICAL
COMPLEX_SUBSTANCE

FOOD
NUCLEIC_ACID
POLYMER
PROTEIN
STEREOCHEMISTRY

b. Modification of Text for Computer Use

² Examples in this document include the Identifying Description field only when they contain data.

³ XML stands for Extended Markup Language and is a computer format used for internet pages.

The rules listed below are used both for Chemical Names, that are both synonyms and preferred terms and the Identifying Description Field.

- a. Convert text to uppercase.

Benzene

becomes

BENZENE

- b. Convert square brackets, [and], and braces, { and }, to parentheses, (and).

4H-[1,3]oxathio[5,4-b]pyrrole

becomes

4H-(1,3)OXATHIOLO(5,4-B)PYRROLE

- c. Spell out Greek letters. Delimit the letter's name with periods.

androstan-17-one, 3-hydroxy-, (3 α , 5 α) -

becomes

ANDROSTAN-17-ONE, 3-HYDROXY-, (3.ALPHA., 5.ALPHA.) -

- d. Convert an arrow, \rightarrow , to the two character sequence hyphen/greater than sign, $->$.

D-Glucose, O-4,6-dideoxy-4-[[(1S,4R,5S,6S)-4,5,6-trihydroxy-3-(hydroxymethyl)-2-cyclohexen-1-yl] amino]- α -D-glucopyranosyl-(1 \rightarrow 4)-O- α -D-glucopyranosyl-(1 \rightarrow 4) -

becomes

D-GLUCOSE, O-4,6-DIDEOXY-4-(((1S,4R,5S,6S)-4,5,6-TRIHIDROXY-3-(HYDROXYMETHYL)-2-CYCLOHEXEN-1-YL) AMINO) - .ALPHA. -D-GLUCOPYRANOSYL-(1 \rightarrow 4)-O- .ALPHA. -D-GLUCOPYRANOSYL-(1 \rightarrow 4) -

- e. Do not use the stereochemical descriptors d and l (for dextro and levo). Replace d with (+), l with (-), and dl with (+/-).

Camphor, d-

becomes

CAMPHOR, (+) -

- f. Represent primes using single quotes. Do not use double quotes.

' is used for a single prime.

" is used for a double prime

''' is used for a triple prime

'''' is used for a quadruple prime

Technetium-99Tc, chloro[bis[(1,2-cyclohexanedione oxime oximato-κO) (1-)] [[1,2-cyclohexanedione (oxime-κO) oximato] (2-)] methylborato (2-) -κN, κN', κN'', κN''', κN''''', κN''''''']-, (TPS-7-1-232'4'54) -

becomes

TECHNETIUM-99TC, CHLORO (BIS ((1,2-CYCLOHEXANEDIONE OXIME OXIMATO- .KAPPA.O) (1-)) ((1,2-CYCLOHEXANEDIONE (OXIME- .KAPPA.O) OXIMATO) (2-)) METHYLBORATO (2-) - .KAPPA.N, .KAPPA.N', .KAPPA.N'', .KAPPA.N''', .KAPPA.N''''', .KAPPA.N''''''')-, (TPS-7-1-232'4'54) -

- g. Enter superscripts using the term "SUP" separated from the other descriptors by a space. Place "SUP" and the superscripted information in parentheses.

1,3,5,7-Tetraazatricyclo[3.3.1.1^{3,7}]decane

becomes

1,3,5,7-TETRAAZATRICYCLO(3.3.1.1(SUP 3,7))DECANE

- h. Subscripts are entered using the term "SUB" separated from the other descriptors by a space. "SUB" and the subscripted information are placed in parentheses.

(15R)-15-methylprostaglandin E₂

becomes

(15R)-15-METHYLPROSTAGLANDIN E(SUB 2)

- i. Convert (±) to (+/-).

1,2-Benzenediol, 4-[1-hydroxy-2-[(1-methylethyl)amino]butyl]-, hydrochloride, (R*,S*)-(±)-

becomes

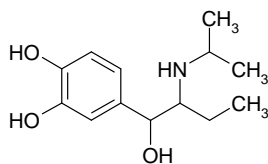
1,2-BENZENEDIOL, 4-(1-HYDROXY-2-((1-METHYLETHYL)AMINO)BUTYL)-, HYDROCHLORIDE, (R*,S*)-(+/-)-

- j. Convert superscript on radioisotopes to normal text in front of atom symbol ¹³¹I becomes 131I. This is an exception to the superscript rule above.

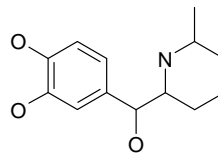
2. Structure Set Up and Layout

Structures are drawn using the landscape format. Atoms should not overlap when hydrogens are displayed on terminal carbon atoms and on heteroatoms. When using MDL software, terminal hydrogens and hydrogens on heteroatoms are not drawn but are displayed when the option is set to hetero-, terminal.⁴

- i. Example: ISOETHARINE



Correct Hydrogens Displayed

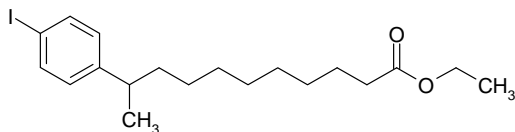


Incorrect Hydrogens Not Displayed

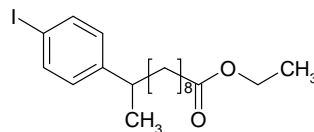
3. Molecules Containing Chains

Each atom in a substance containing a long chain is explicitly drawn. Brackets containing repeating units are not used for substances other than polymers and simple salts.⁵

- i. Example: IOPHENDYLATE



Correct



Incorrect

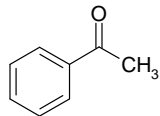
4. Abbreviations

Abbreviations for chemical groups, such as acetyl, ethyl, isopropyl, methyl, and phenyl, are not used. Explicitly draw each non-hydrogen atom of every chemical group.

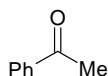
- i. Example: ACETOPHENONE

⁴ When the structure is displayed with a 1-cm bond length and a 12 pt arial font.

⁵ Polymers typically do not have a defined number of repeating units but a range of repeating units. See section 27 on Polymers.



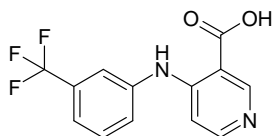
Correct



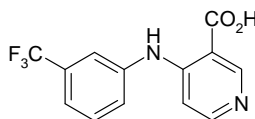
Incorrect

Condensed groups, such as CO_2H , COOH , NO_2 , SO_3H , and CF_3 , are not used. Explicitly draw each atom and bond.

ii. Example: TRIFLOXIN



Correct



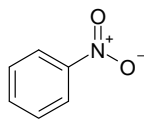
Incorrect

5. Functional Groups

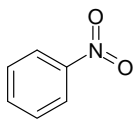
a. NITRO Groups and NITRATES

Nitro groups and nitrates are drawn with a positive charge on the nitrogen and a negative charge on one of the oxygens. They are not drawn with five bonds to the nitrogen.

i. Example: NITROBENZENE

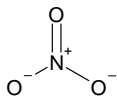


Correct

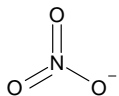


Incorrect

ii. Example: NITRATION



Correct

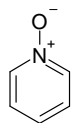


Incorrect

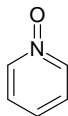
b. NITROGEN OXIDES

Nitrogen oxides are drawn with a positive charge on the nitrogen and a negative charge on the oxygen.

i. Example: PYRIDINE 1-OXIDE



Correct

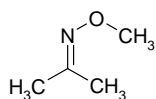


Incorrect

c. OXIMES

Oximes are drawn without a charge on any atom.

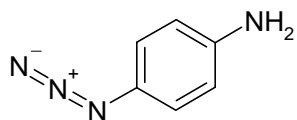
i. Example: 2-PROPANONE, O-METHYLOXIME



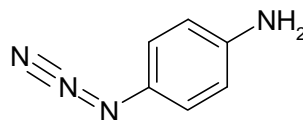
d. AZIDES

The azido group is drawn with all three nitrogen atoms in a straight line. The terminal nitrogen has a negative charge and the middle nitrogen has a positive charge.

i. Example: 4-AZIDOBENZAMINE



Correct

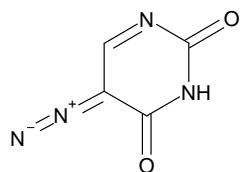


Incorrect

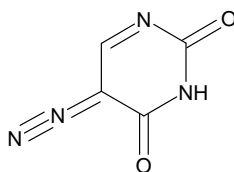
e. DIAZO and AZO Compounds

The terminal diazo group is drawn with the carbon bonded to the two nitrogen atoms in a linear configuration. The terminal nitrogen has a negative charge and the middle nitrogen has a positive charge.

i. Example: 5-DIAZOURACIL



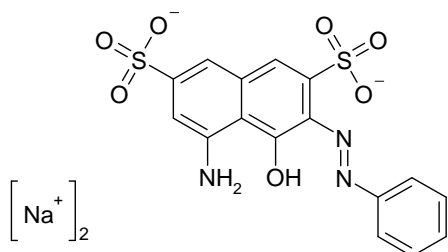
Correct



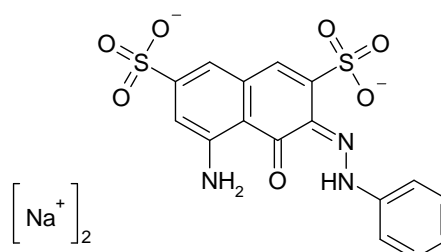
Incorrect

Diazo or azo dyes are normally thought to exist entirely in an E or trans configuration and are drawn as such unless information is provided otherwise. A Z or cis configuration usually forces the aromatic rings out of the plane and is usually a “high energy” configuration. These dyes are also drawn as diazo and not as hydrazone tautomers.

ii. Example: D&C Red No. 33



Correct

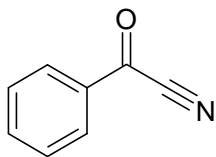


Incorrect

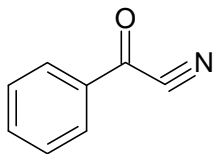
f. CYANIDES and ISOCYANIDES

Compounds with a cyano or an isocyanide group are drawn with the nitrogen and carbon atoms of the cyano group in line with the attachment atom.

i. Example: BENZOYL CYANIDE

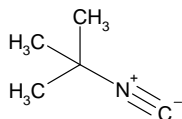


Correct

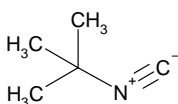


Incorrect

ii. Example: t-BUTYLISOCYANIDE



Correct

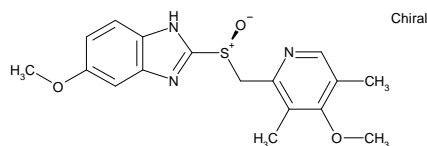


Incorrect

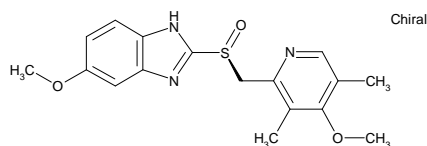
g. SULFOXIDES

The sulfoxide group is drawn in a charge-separated configuration to allow for the assignment of chirality. A single bond is drawn between the sulfur and the oxygen atoms. A negative charge is placed on the oxygen atom and a positive charge on the sulfur atom.

i. Example: ESOMEPRAZOLE



Correct



Incorrect

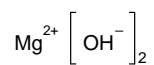
6. Mixtures and Salts

Substances that contain multiple molecules or ions are presented as either a salt or a mixture. For a salt, all the components are stored in a single record. A mixture is represented by two or more substance records that are brought together to define a new substance. The registration system has a module design for entering mixtures. When a mixture is viewed, each component is listed in a table and each substance is displayed separately.

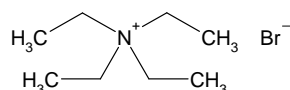
For the SRS, a simple salt is defined as an ionic compound of fixed, undefined, or variable stoichiometry. This includes simple salts, quaternary ammonium compounds, and salts formed by reaction between organic bases and acids.

a. Types of Salts

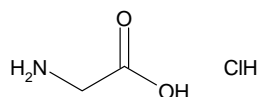
- i. Simple Salt Example: MAGNESIUM HYDROXIDE



- ii. Quaternary Ammonium Compound Example: TETRAETHYLAMMONIUM BROMIDE



- iii. Salt Formed by Reaction between Organic Base and an Acid Example: GLYCINE HYDROCHLORIDE⁶



b. Types of Mixtures

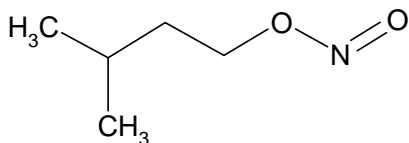
Three types of mixtures are possible: “ALL OF”, “ANY OF”, and “ONE OF”. An “ALL OF” mixture is used when two or more substances are always present. An “ANY OF” mixture is used when at least one component is present, but there is the possibility that many if not all the components are present. An ‘ONE OF’ mixture is used when only one of the components is present.

- i. Example of an ALL OF Mixture: AMYL NITRILE

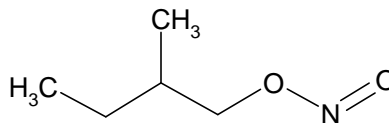
⁶ A salt formed by the reaction of an organic base with an acid is drawn as the organic base and the acid form of the counter ion, not as two ions (for more examples, see section 7 on Salts Formed by the Reaction of Nitrogenous Organic Bases with Acids).

United States Pharmacopeia (USP) defines Amyl Nitrate as a mixture of the nitrite esters of 3-methyl-1-butanol and 2-methyl-1-butanol.

Mixture Type: "ALL OF"



α -Lactose

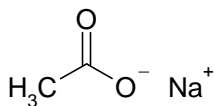


β -Lactose

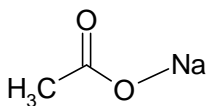
7. Salts and Compounds with Ionic Bonds

Ionic bonds are drawn by displaying a charge on the appropriate atoms.

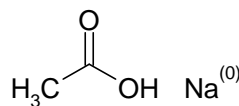
- i. Example: SODIUM ACETATE



Correct



Incorrect



Incorrect

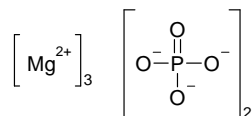
For simple salts, place the cation on the left and the anion on the right. When the stoichiometry is 1:1, simply place the components next to each other.

- ii. Example: POTASSIUM CHLORIDE



Indicate other stoichiometric ratios by placing brackets around any ion that is repeated. Place the repeat count at the bottom right hand corner of the brackets.

- iii. Example: MAGNESIUM PHOSPHATE



8. Salts Formed by the Reaction of Ammonia with Acids

A salt formed by the reaction of ammonia with an acid is represented as the ammonium ion and the conjugate base of the acid.

- i. Example: AMMONIUM CHLORIDE

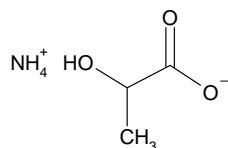


Correct

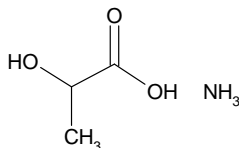


Incorrect

- ii. Example: AMMONIUM LACTATE



Correct

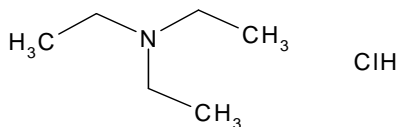


Incorrect

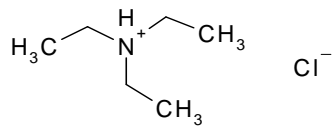
9. Salts Formed by the Reaction of Nitrogenous Organic Bases with Acids

A salt formed by the reaction of a nitrogenous organic base with an acid is represented as the free base and the acid form of the counter ion.

- i. Example: TRIETHYLAMINE HYDROCHLORIDE

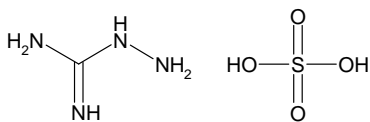


Correct

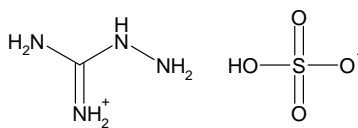


Incorrect

ii. Example: AMINOGUANIDINE SULFATE



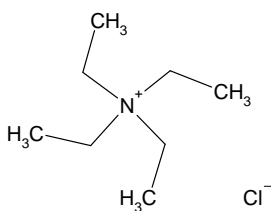
Correct



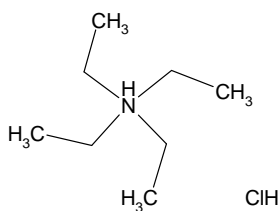
Incorrect

The nitrogen atom in a quaternary ammonium compound is drawn with a positive charge

iii. Example: TETRAETHYLAMMONIUM CHLORIDE



Correct

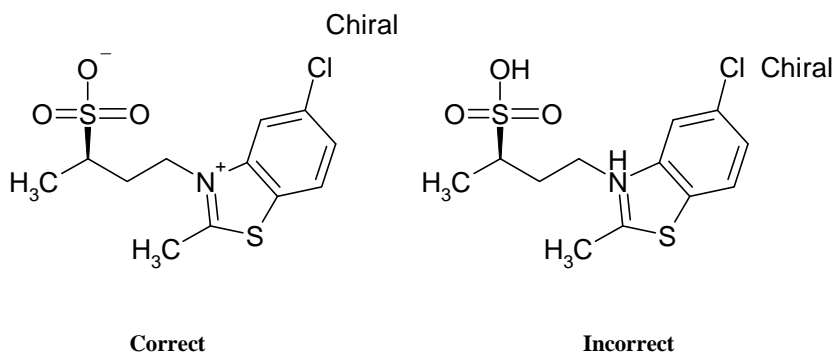


Incorrect

10. Inner Salts

Inner salts are substances in which positively and negatively charged groups are contained within the same molecule.

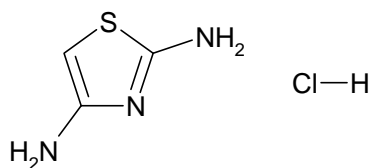
i. Example: R-5-CHLORO-2-METHYL-3-(4-SULFOBUTYL)-BENZOTHAZOLIUM INNER SALT



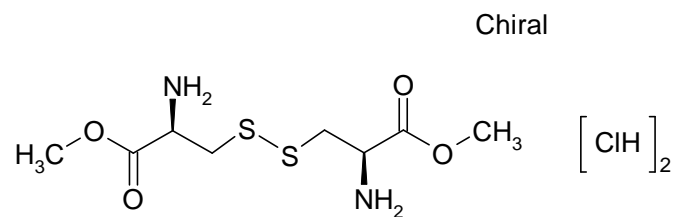
11. Bracket Usage for Salts and Solvates

Salts and solvates contain two or more molecular entities in the structure field. If the ratio of the components is 1:1, they are placed side by side. If the ratio is not 1:1, brackets are used to indicate the ratio of the components. Brackets are always labeled with whole numbers or S-text. Brackets are not labeled with fractions. For hemihydrates and similar situations use the ratio with the lowest whole numbers.

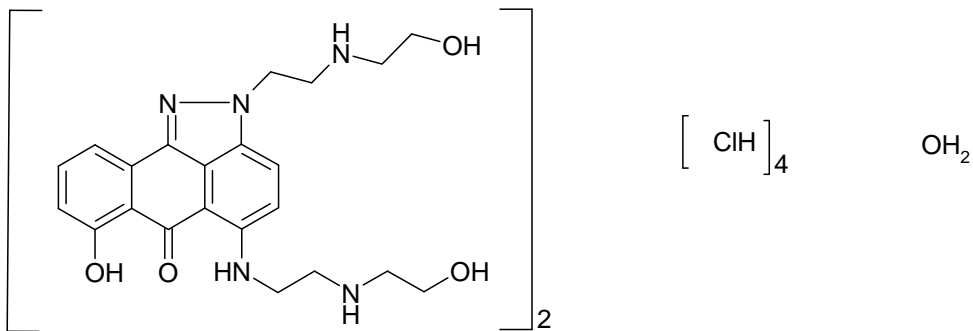
- i. Example: 1:1 Compound to Salt Ratio



- ii. Example: 1:2 Compound to Salt Ratio



- iii. Example: LOSOXANTRONE HYDROCHLORIDE (DIHYDOCHLORIDE HEMIHYDRATE)



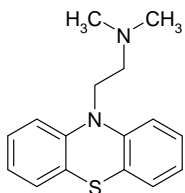
12. Salts Containing a Complex Substance as a Counterion

When a substance consists of a well-defined component and a complex counterion, the resulting salt is treated as a mixture.⁷ The mixture consists of one record for the well-defined component and one record, for the complex counter ion.

- i. Example: FENETHAZINE TANNATE

Mixture Type: "ALL OF"

Fenethazine tannate is drawn as a mixture of fenethazine and a tannic acid. This represents a mixture of a mixture (tannic acid)



FENETHAZINE

Tannic acid: See section 15.

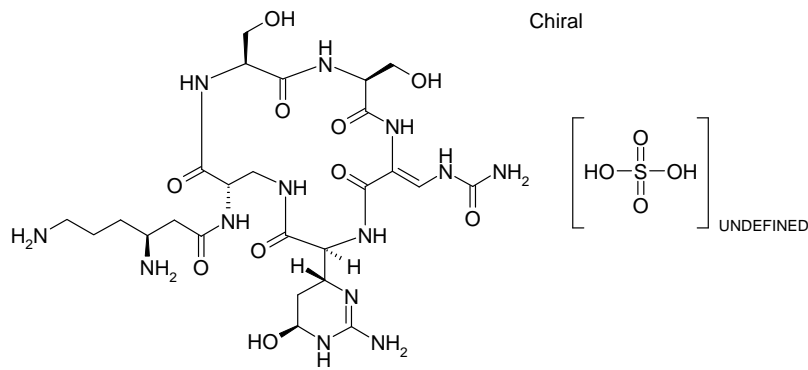
13. Substance Of Unknown Or Variable Stoichiometry

⁷ For an example of a salt with a well-defined counterion, see section on **Salts Formed by the Reaction of Nitrogenous Organic Bases with Acids**.

For substances with undefined stoichiometry, enclose the undefined component in brackets. Attach the word "UNDEFINED" to each bracket as S-text.⁸

For substances with two or more components, the main component is not placed in brackets. Only the minor component, which typically is a solvate or counter ion, is bracketed.

i. Example: VIOMYCIN SULFATE



14. Partially Ionized Acids

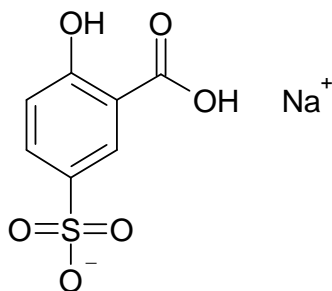
If a molecule contains multiple ionizable groups and is only partially ionized, the group that is the strongest acid is ionized first. The order of ionization for acidic functional groups is listed in the following table in order decreasing acidity:

Functional Group Acidity Table
-OSO ₃ H
-SO ₃ H
-OSO ₂ H
-SO ₂ H
-OPO ₃ H ₂
-PO ₃ H ₂
-CO ₂ H
Thiophenol
(-OPO ₃ H) ⁻
(-PO ₃ H) ⁻
Phthalimide,
CO ₃ H (peracetyl),
α-carbon-hydrogen to nitro group
-SO ₂ NH ₂ (sulfonamide)

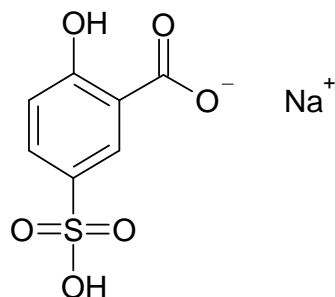
⁸ The use of "S-text", which is a text associated with the structure, is used in a limited number of situations. S-text is always attached to generic brackets.

-OBO ₂ H ₂
-BO ₂ H ₂
phenol
SH (aliphatic)
(-OBO ₂ H) ⁻
(-BO ₂ H) ⁻
Cyclopentadiene
-CONH ₂ (amide)
imidazole
-OH (aliphatic alcohol)
.alpha.carbon-hydrogen to keto group
.alpha.carbon-hydrogen to acetyl ester group
sp Carbon hydrogen
.alpha.carbon-hydrogen to sulfone group
alpha.carbon-hydrogen to sulfoxide group
-NH ₂ (amine)
benzyl hydrogen
sp ² -Carbon hydrogen
sp ³ -Carbon hydrogen

i. Example: 5-SULFOSALICYLATE, MONOSODIUM SALT



Correct



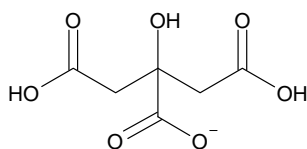
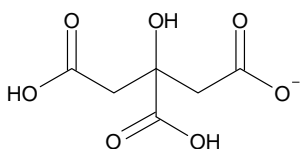
Incorrect

Salts of partially ionized acids are drawn with either all of the same functional groups completely ionized or completely unionized. The extent of ionization is sufficient to account for all the counterions (metal ions; non-hydrogen ions) present in a substance. Hydrogen ions from the extended periodic table are added to balance the charges in the substance. The only exceptions to these rules are the case where all the acidic functional groups are equivalent prior to ionization. For these substances, ionization only occurs to the extent necessary to account for all of the counterions. This exception only applies when **all** of the same functional groups are equivalent. (see **ALENDRONATE SODIUM** and **MONOSODIUM CITRATE**).

When a compound contains two or more of the same functional groups that are not equivalent, two or more structures are possible when it is partially ionized.

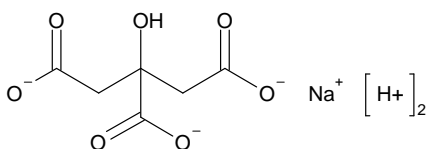
ii. Example: Citrate Ion with Single Negative

Citric acid contains three carboxylate groups. The two end groups are equivalent; the central one is not.

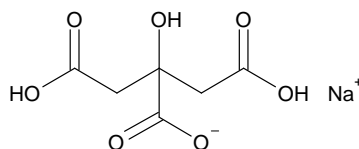


All of the carboxylate groups in citric acid would be ionized and two hydrogen ion superatoms, H⁺, from the MDL extended periodic table are added to charge balance the structure.

iii. Example: MONOSODIUM CITRATE

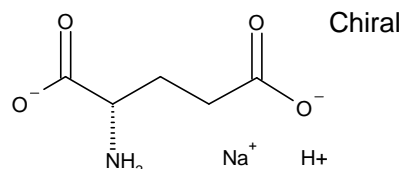


Correct

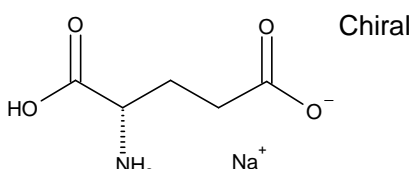


Incorrect

iv. Example: MONOSODIUM GLUTAMATE

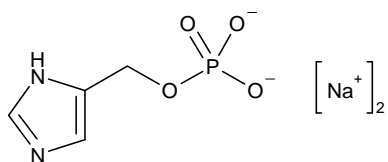


Correct

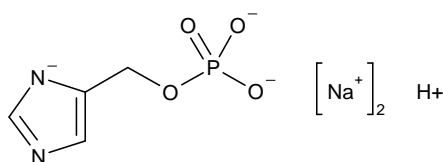


Incorrect

v. Example: 1H-IMIDAZOLE-4-METHANOL, DIHYDROGEN PHOSPHATE (ESTER), DISODIUM SALT



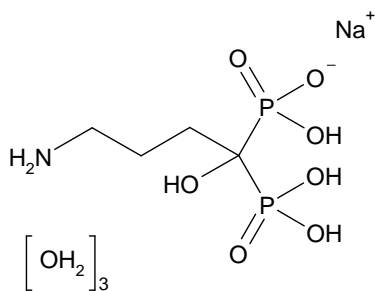
Correct



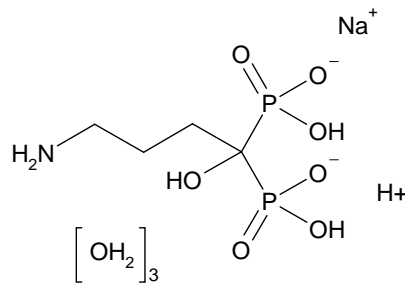
Incorrect

vi. Example: ALENDRONATE SODIUM

When acidic functional groups are equivalent before ionization, ionization only precedes to the extent necessary to account for all of the counterions. No H⁺ superatoms are added to the structure.



Correct



Incorrect

15. Complex Inorganic and Organic Compounds

Tags for Complex Substances

Unpurified or partially purified animal products, plant products that are not botanicals, substances that contain more than 10 active components or chemicals that cannot be represented by a two-dimensional structure are entered using the Identifying Description field. Complex substances can also be mixtures of several well defined active components in a complex matrix of inactive components. The following tags are used to describe complex substances.

<COMPLEX_SUBSTANCE></COMPLEX_SUBSTANCE> Parent tags indicating that the substance is complex and does not have a structure.

<SOURCE_TYPE></SOURCE_TYPE> The type of material that the substance is derived from.

Examples: PLANT, ANIMAL, MINERAL, FUNGUS, BACTERIA

<GENUS></GENUS> The genus of a given animal, fungi, or bacteria.

<SPECIES></SPECIES> The species of a given plant, animal, fungi, or bacteria.

<SUBSPECIES></SUBSPECIES> Subspecies or variant from which the species is prepared.

<PART></PART> The part of a plant or animal that is used to prepare the substance or contains the substance. The part is always singular. Examples: PANCREAS, URINE, THYROID

<AUTHOR></AUTHOR> The author of the classification of a particular species. Example: LINNAEUS. The genus, species, subspecies, and author are obtained from either the United States Pharmacopeia (USP) or the USDA's Integrated Taxonomic Information System (ITIS).

<FORMULA></FORMULA> The molecular formula of a substance. This is used for inorganic substances. Molecular formulas are listed using the Hill Order with an underscore “_” between elements. The Hill order is C followed by H, the rest of the elements are listed in alphabetical order. All letters are capitalized. A separate formula is written for each independent molecular fragment. Multiple fragments in an entity are enclosed by parenthesis followed by the number of

multiples. X following a parenthesis represents unknown or variable. Example: Quindecamine acetate which is a dihydrate would have the molecular formula

C30_H38_N4.(C2_H4_O2)2.(H2_O)2.

<PROCESS></PROCESS> Any information about process that resulted in the substance.

Examples: EXTRACT, OIL, WATER EXTRACT, ETHANOL EXTRACT, TOLUENE EXTRACT, GLYCERIN EXTRACT, METHANOL EXTRACT

<COMMENTS></COMMENTS> Information about a substance used to distinguish it from a closely related substance. Example: Details on a purification process.

a. Complex Inorganics

Many inorganic substances exist as complex crystalline or amorphous structures that cannot be easily represented using MDL software. The following is an example of how tags are used for complex inorganic substances.

i. Example: TALC

```
<COMPLEX_SUBSTANCE>
  <SOURCE_TYPE>MINERAL</SOURCE_TYPE>
  <GENUS></GENUS>
  <SPECIES></SPECIES>
  <SUBSPECIES></SUBSPECIES>
  <PART></PART>
  <AUTHOR></AUTHOR>
  <PROCESS></PROCESS>
  <FORMULA>MG_S14_O10.(OH)2.(H2_O)X</FORMULA>
  <COMMENTS></COMMENTS>
</COMPLEX_SUBSTANCE>
```

b. Complex Organics

i. Example: THYROID

```
<COMPLEX_SUBSTANCE>
  <SOURCE_TYPE>ANIMAL</SOURCE_TYPE>
  <GENUS>SUS</GENUS>
  <SPECIES>SCOFA</SPECIES>
  <SUBSPECIES></SUBSPECIES>
  <PART>THYROID</PART>
  <AUTHOR></AUTHOR>
  <PROCESS>DRIED POWDER</PROCESS>
  <FORMULA></FORMULA>
  <COMMENTS></COMMENTS>
</COMPLEX_SUBSTANCE>
```

16. Botanicals

Botanicals are represented in the SRS solely through the use of the Identifying Description Field which includes the Latin trinomial (genus, species, and variety), the author (e.g. Linnaeus, abbreviated L.), the plant part, and the process (e.g. 'extract'). If only the genus is known, the designation 'SPP' is inserted in the species tags to show general applicability to all species under that genus. If the variety, plant part, or process is unknown or not applicable, then they may be omitted. Latin binomials, trinomials, and authors are validated against accepted entries in the United States Pharmacopeia (USP) or the United States

Department of Agriculture's (USDA) Integrated Taxonomic Information System (ITIS) available at <http://www.itis.usda.gov/>. All plant parts are validated against listed terms in the Plant Part monograph in the Center of Drug Evaluation and Research (CDER) Data Standards Manual available at <http://www.fda.gov/cder/dsm/drg/drg00924.htm>. Particular extract types (e.g., 'water', 'ethanol') are not permitted for botanicals that originate from within FDA, but shall be permitted for botanicals that originate from a non-FDA source (e.g., International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), European Medicines Agency (EMA)). When the extract type must be represented, it is inserted prior to the word 'extract' (e.g., 'water extract'). Only the following extract solvents are allowed in SRS botanical nomenclature: ethanol, ether, glycerin, toluene, water, ammonia, methanol, and acetone. All plant-derived complex substances would be considered a botanical. Any characterized substances that are present, or are possibly present, in a botanical are purposely excluded from SRS botanical definition.

<BOTANICAL></BOTANICAL> Parent tags indicate the substance is a botanical.
<GENUS></GENUS> The genus of a given plant.
<SPECIES></SPECIES> The species of a given plant.
<SUBSPECIES></SUBSPECIES> Subspecies or variant from which the species is prepared.
<PART></PART> The part of a plant that is used to prepare the substance or contains the substance. The part is always singular. All plant parts are validated against listed terms in the Plant Part monograph in the CDER Data Standards Manual available at <http://www.fda.gov/cder/dsm/drg/drg00924.htm>.
<AUTHOR></AUTHOR> The author of the classification of a particular species. LINNAEUS. The genus, species, subspecies and author are derived from either the United States Pharmacopeia (USP) or the USDA's Integrated Taxonomic Information System (ITIS). The full name of the author is spelled out if available due to inconsistencies across sources.
<PROCESS></PROCESS> Any information about process that resulted in the substance. EXTRACT, OIL, WATER EXTRACT, ETHANOL EXTRACT, TOLUENE EXTRACT, GLYCERIN EXTRACT, METHANOL EXTRACT
<COMMENTS></COMMENTS> Information about a substance used to distinguish it from a closely related substance.

i. Example: FENNEL OIL

```
<BOTANICAL>
  <GENUS>FOENICULUM</GENUS>
  <SPECIES>VULGARE</SPECIES>
  <SUBSPECIES></SUBSPECIES>
  <AUTHOR>MILLER.</AUTHOR>
  <PART>FRUIT</PART>
  <PROCESS>OIL</PROCESS>
  <COMMENTS></COMMENTS>
</BOTANICAL>
```

17. Foods

Food indexing is based on the LanguaL international thesaurus for the description of foods. Information on LanguaL may be found at <http://www.languaL.org>

Authoritative sources for information on foods are:

Herbst, Sharon Tyler. FOOD LOVER'S COMPANION: COMPREHENSIVE DEFINITIONS OF OVER 6000 FOOD, DRINK, AND CULINARY TERMS .Third Edition. Hauppauge, NY: Barron's Educational Series, c.2001.

US Department of Agriculture Food Safety Inspection Service (FSIS) FACT SHEETS

<http://www.fsis.usda.gov>

Every food should be indexed with as much detail as is given in the food name. This might include extent of heat treatment, cooking method, or preservation method. Every food substance/ingredient must have a unique string of Languag codes in order to be assigned a UNII code.

The Substance Registration System indexes only food substances/ingredients, not food products (also known as "finished foods"). Examples of food substances/ingredients are "milk", "egg", "pineapple", or "cherry". Examples of food products are "eggnog" or "fruitcake". Food substances/ingredients are the building blocks for food products. Thus, "peanut", "salt", and "sugar" would be indexed in the SRS but "peanut butter" would not.

It is possible that a food may already be in the dictionary as a complex substance or botanical; an example would be some vegetable oils or some seeds or spices. For foods originating from plants it may also be worthwhile to check the description field for all substances of the same genus to determine if the food has already been entered as a complex substance. To obtain the genus of a plant, the United States Department of Agriculture's (USDA) Integrated Taxonomic Information System (ITIS), available at <http://www.itis.usda.gov/>, should be used. It can be searched by common name. Care should be taken to ensure that the complex substance or botanical refers to the same part of the plant that the food is derived. If the food is already entered as a complex substance or botanical the food tags should be added to the existing entry.

The following tags are used to define foods:

<FOOD></FOOD> First tag in description field indicates that the description contains a food.

<COMMON_FOOD_NAME></COMMON_FOOD_NAME> Indicates a common name that Languag has assigned to a particular food.

<FPT ></FPT > Indicates a Languag facet of the food name at the FPT, or product type, level. Example: SEAFOOD OR SEAFOOD PRODUCT

<FACET></FACET>Facet term used to describe a given food.

<LANGUAL_CODE></LANGUAL_CODE>Code associated with the given Languag facet term.

<FSO></FSO> Indicates a Languag facet of the food name at the FSO, or food source, level. Example: ABALONE

<FACET></FACET>

<LANGUAL_CODE></LANGUAL_CODE>

<FPP></FPP> Indicates a Languag facet of the food name at the FPP, or part of plant or animal, level.

Example: MEAT PART

<FACET></FACET>

<LANGUAL_CODE></LANGUAL_CODE>

<FPN></FPN> Indicates a Languag facet of the food name at the FPN, or extent of heat treatment, level.

<FACET></FACET>

<LANGUAL_CODE></LANGUAL_CODE>

<FKM></FKM> Indicates a Languag facet of the food name at the FKM, or cooking method, level.

<FACET></FACET>

<LANGUAL_CODE></LANGUAL_CODE>

<FTR></FTR> Indicates a Languag facet of the food name at the FTR, or treatment applied, level.

Example: IRRADIATION.

<FACET></FACET>

<LANGUAL_CODE></LANGUAL_CODE>

<FPV></FPV> Indicates a Languag facet of the food name at the FPV or preservation level. Example:

DRIED.

<FACET></FACET>

<LANGUAL_CODE></LANGUAL_CODE>

<COMMENTS></COMMENTS Used to distinguish one food ingredient from another if the Languag facets are insufficient to distinguish foods from one another or to capture a feature that is not captured using Languag.

i. Example ABALONE

```
<FOOD>
  <COMMON_FOOD_NAME>ABALONE</COMMON_FOOD_NAME>
  <FPT>
    <FACET>SEAFOOD OR SEAFOOD PRODUCT</FACET>
    <LANGUAL_CODE>A0267</LANGUAL_CODE>
  </FPT>
  <FSO>
    <FACET>ABALONE</FACET>
    <LANGUAL_CODE>B1408</LANGUAL_CODE>
  </FSO>
  <FPP>
    <FACET>MEAT PART</FACET>
    <LANGUAL_CODE>C0103</LANGUAL_CODE>
  </FPP>
  <FPN>
    <FACET></FACET>
    <LANGUAL_CODE></LANGUAL_CODE>
  </FPN>
  <FKM>
    <FACET></FACET>
    <LANGUAL_CODE></LANGUAL_CODE>
  </FKM>
  <FTR>
    <FACET></FACET>
    <LANGUAL_CODE></LANGUAL_CODE>
  </FTR>
  <FPV>
    <FACET></FACET>
    <LANGUAL_CODE></LANGUAL_CODE>
  </FPV>
  <COMMENTS></COMMENTS>
</FOOD>
```

ii. Example BEEF

```
<FOOD>
  <COMMON_FOOD_NAME>BEEF</COMMON_FOOD_NAME>
  <FPT>
    <FACET>MEAT OR MEAT PRODUCT (FROM MAMMAL)</FACET>
    <LANGUAL_CODE>A0150</LANGUAL_CODE>
  </FPT>
  <FSO>
    <FACET>CATTLE</FACET>
    <LANGUAL_CODE>B1161</LANGUAL_CODE>
  </FSO>
  <FPP>
    <FACET>SKELETAL MEAT PART</FACET>
    <LANGUAL_CODE>C0175</LANGUAL_CODE>
  </FPP>
  <FPN>
```

```

        <FACET></FACET>
        <LANGUAL_CODE></LANGUAL_CODE>
    </FPN>
    <FKM>
        <FACET></FACET>
        <LANGUAL_CODE></LANGUAL_CODE>
    </FKM>
    <FTR>
        <FACET></FACET>
        <LANGUAL_CODE></LANGUAL_CODE>
    </FTR>
    <FPV>
        <FACET></FACET>
        <LANGUAL_CODE></LANGUAL_CODE>
    </FPV>
    <COMMENTS></COMMENTS>
</FOOD>

```

iii. Example SQUID, BOILED

```

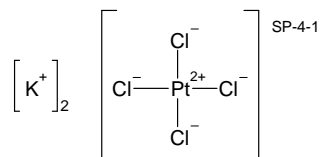
<FOOD>
    <COMMON_FOOD_NAME>SQUID, BOILED</COMMON_FOOD_NAME>
    <FPT>
        <FACET>SEAFOOD OR SEAFOOD PRODUCT</FACET>
        <LANGUAL_CODE>A0267</LANGUAL_CODE>
    </FPT>
    <FSO>
        <FACET>SQUID</FACET>
        <LANGUAL_CODE>B1205</LANGUAL_CODE>
    </FSO>
    <FPP>
        <FACET>MEAT PART</FACET>
        <LANGUAL_CODE>C0103</LANGUAL_CODE>
    </FPP>
    <FPN>
        <FACET>FULLY HEAT-TREATED</FACET>
        <LANGUAL_CODE>F0014</LANGUAL_CODE>
    </FPN>
    <FKM>
        <FACET>BOILED</FACET>
        <LANGUAL_CODE>G0014</LANGUAL_CODE>
    </FKM>
    <FTR>
        <FACET></FACET>
        <LANGUAL_CODE></LANGUAL_CODE>
    </FTR>
    <FPV>
        <FACET></FACET>
        <LANGUAL_CODE></LANGUAL_CODE>
    </FPV>
    <COMMENTS></COMMENTS>
</FOOD>

```

18. Coordination compounds or complexes (Draft Proposed)

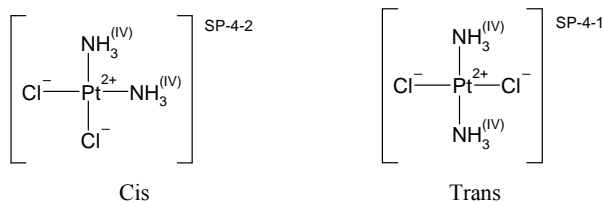
Coordination compounds are defined using the structure field and S-Text which is associated with the bracket enclosing the coordination compound. In cases where it is customary to use a coordination or dative bond, a single bond is drawn and the valence of the coordinating atom is adjusted as necessary.⁹ The coordination compound is placed in brackets and the brackets labeled using S-Text with a polyhedral descriptor. Polyhedral descriptors are described in the Chemical Abstracts Service (CAS) Index guide.

- i. Example: PLATINATE(2-), TETRACHLORO-, DIPOTASSIUM , (SP-4-1)-



The brackets and polyhedral descriptor are required to differentiate two isomers that differ only in their configuration around the central atom. The placement of ligands around the central atom is for aesthetic purposes only—MDL software does not provide a mechanism for indicating stereochemistry around the central atom.¹⁰ The polyhedral descriptor is added using S-Text and is placed next to the upper right hand corner of the brackets. The S-text field FDAREG_SGROUP is used to store the bracket information .

- ii. Example: CISPLATIN and TRANSPLATIN



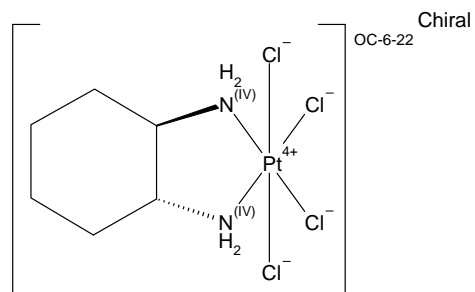
Without the brackets and coordination descriptors, the MDL software would treat these compounds as duplicates.

Wedge bonds are only used for tetrahedral stereochemistry. Tetrahedral stereochemistry on ligands is treated the same as in non-coordination compounds. The Chiral flag only applies to tetrahedral stereocenters.

- iii. Example: DEXORMAPLATIN

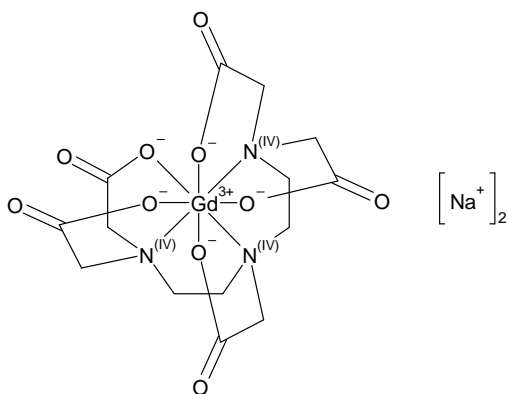
⁹ The MDL software does not include a coordination bond (dative bond).

¹⁰ MDL's software only handles stereochemistry for the following central atoms: C, N, O, S (sulfate and sulfoxide), P, Si, and pentavalent N and P (3 or 4 attachments with one double bond). Only tetrahedral stereochemistry can be represented.



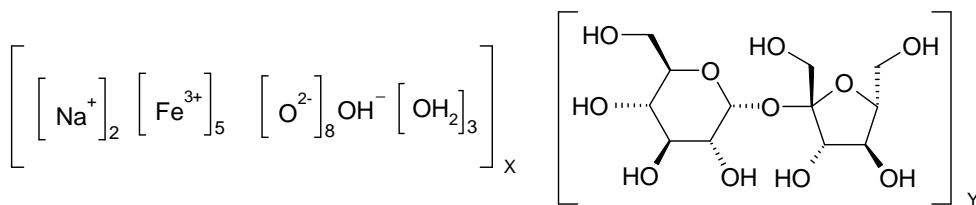
When the arrangement of atoms and/or the chirality around the central atom is not specified in the source, no polyhedral descriptor is used and no brackets are drawn around the coordination compound.

iv. Example: GADOLINATE(2-), (N,N-BIS)2-(BIS(((CARBOXY-KAPPA.O)METHYL)AMINO-KAPPA.N)ETHYL)GLYCINATO(5-)-KAPPA.N,KAPPA.O)-, DISODIUM



v. Example: IRON SUCROSE (Proposed Not Official)

Both the structure field and Complex Substance tags will be used to describe this colloidal substance.

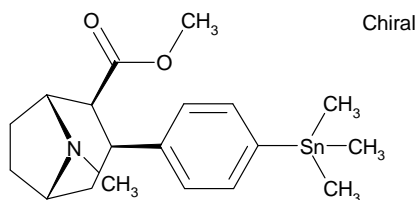


<COMPLEX_SUBSTANCE><SOURCE_TYPE>MINERAL</SOURCE_TYPE><GENUS></GENUS><SPECIES></SPECIES><SUBSPECIES></SUBSPECIES><PART></PART><AUTHOR></AUTHOR><FORMULA>(NA2_FE5_O8_(OH).(H2_O)3)X.(C12_H22_O11)Y</FORMULA><PROCESS></PROCESS><COMMENTS>POLYNUCLEAR IRON (III) HYDROXIDE/SUCROSE COMPLEX; MW: 34000 TO 60000</COMMENTS></COMPLEX_SUBSTANCE>

19. Organometallic Compounds

Carbon to metal sigma bonds are drawn as covalent bonds.

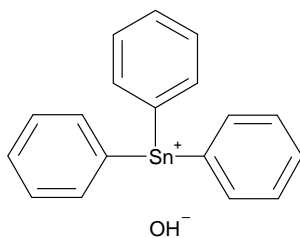
i. Example: 2-BETA.-CARBOMETHOXY-3-BETA.-(4-TRIMETHYLSTANNYLPHENYL)TROPANE



a. Nitrogen, Oxygen, and Fluorine Bonded to Metallic Centers

Nitrogen-metal, oxygen-metal, and fluorine-metal bonds are always treated as ionic bonds.

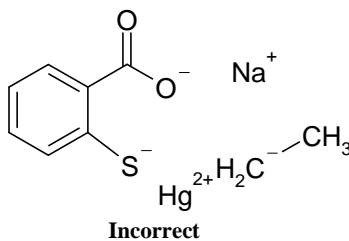
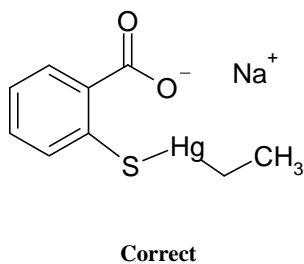
i. Example: TRIPHENYLTIN HYDROXIDE



b. Other Nonmetals Bonded to Metallic Centers

The type of bond between metal and nonmetallic elements (excluding nitrogen, oxygen, and fluorine) is determined by the elements position in the periodic table. Transition metal-nonmetal (except mercury) bonds and aluminum-nonmetal bonds are drawn as ionic bonds. The following metals are drawn with covalent bonds to nonmetals: mercury, gallium, germanium, indium, tin, antimony, thallium, lead, bismuth, and polonium.

i. Example: THIMEROSAL



20. Tautomerization and Resonance Form Presentation

In most instances, the SRS does not allow the registration of different tautomeric or resonance forms of the same compound. To maintain a more consistent set of structures, use the following conventions.¹¹

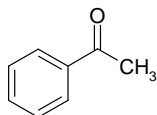
a. Proton-Shift Tautomerization

The following categories include tautomers that interconvert by proton transfer.

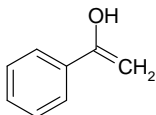
b. Ketones

The keto-form is drawn when there is potential keto-enol tautomerism.

- i. Example: ACETOPHENONE



Correct

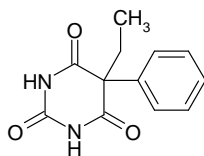


Incorrect

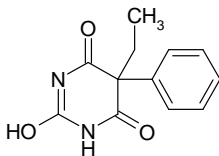
c. Carbonyl Compounds

In general, carbonyl compounds that tautomerize are drawn in the form containing the carbonyl group, not the form containing the hydroxyl group.

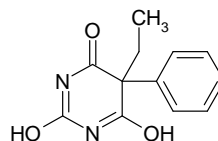
- i. Example: PHENOBARBITAL



Correct



Incorrect



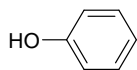
Incorrect

d. Phenols

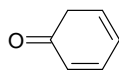
Phenols are drawn in the hydroxyl form, not the keto form.

¹¹ When the structure of a compound prevents the preferred tautomer from being formed, the compound is drawn in the alternate form.

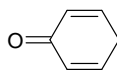
- i. Example: PHENOL



Correct



Incorrect

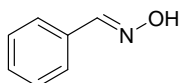


Incorrect

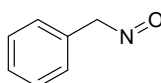
e. Oximes

Oximes are drawn in the hydroxyl form, not the nitroso form.

- i. Example: BENZALDOXIME



Correct

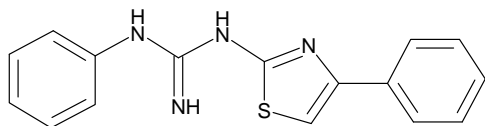


Incorrect

f. Guanidines

In acyclic guanidines, the double bond is placed between the central carbon and the least substituted nitrogen.

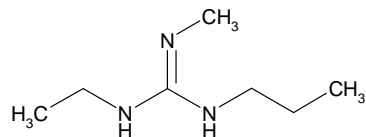
- i. Example: 1-PHENYL-3-(4-PHENYL-2-THIAZOLYL) GUANIDINE



If all nitrogens have substituents, draw the double bond to the nitrogen with the lowest priority as determined using Cahn-Ingold-Prelog sequence rules.¹²

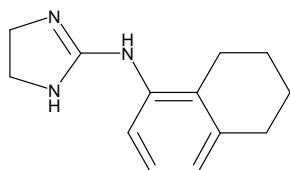
¹² Cahn-Ingold-Prelog sequence rules are outlined in the following papers: R.S. Cahn, C.K. Ingold and V. Prelog, *Angew. Chem.* 78, 413–447 (1966), *Angew. Chem. Internat. Ed. Eng.* 5, 385–415, 511 (1966); and V. Prelog and G. Helmchen, *Angew. Chem.* 94, 614–631 (1982), *Angew. Chem. Internat. Ed. Eng.* 21, 567–583 (1982). 1996, 68, 2203

ii. Example: 1-PHENYL-3-(4-PHENYL-2-THIAZOLYL)GUANIDINE



If a guanidine is incorporated into a ring, the double bond is placed inside the ring.

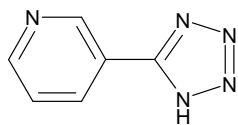
iii. Example: TRAMAZOLINE



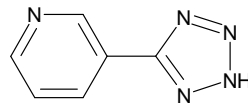
g. Tetrazoles

In tetrazoles, the hydrogen is placed at the one position on the tetrazole ring.

i. Example: 5-(3-PYRIDYL)TETRAZOLE



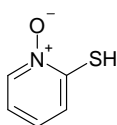
Correct



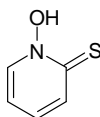
Incorrect

h. Pyrithione

Pyrithione and pyrithione salts are drawn in the thiol-nitroxide form, not the thione-hydroxylamine tautomer.



Correct

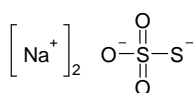


Incorrect

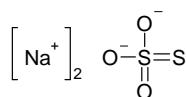
i. Thiosulfate and Thiosulfite

Thiosulfate and thiosulfite are represented with a sulfur-sulfur single bond and a hydrogen (or negative charge) on the outer sulfur.

i. Example: SODIUM THIOSULFATE



Correct

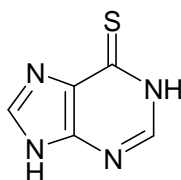


Incorrect

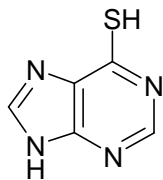
j. Purines and Pyrimidines

All purines and pyrimidines are drawn as Watson-Crick like forms with the keto (thione)-tautomer form assumed and a hydrogen atom at the one position in pyrimidines and at the nine position in purines.

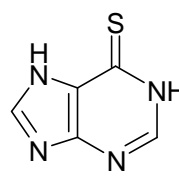
i. Example: 6-THIOPURINE



Correct

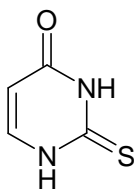


Incorrect

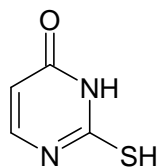


Incorrect

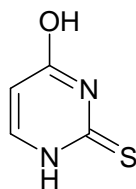
ii. Example: 2-THIOURACIL



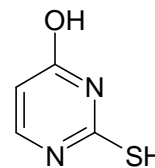
Correct



Incorrect



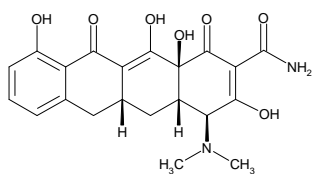
Incorrect



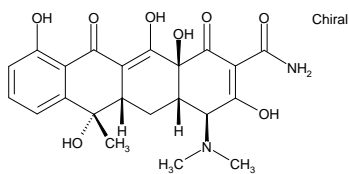
Incorrect

k. Tetracyclines

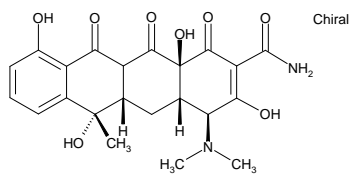
Tetracyclines can tautomerize into numerous isomers. In the SRS, tetracyclines are drawn from the following sub-structure:



i. Example: TETRACYCLINE



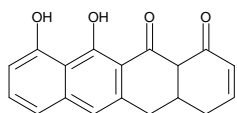
Correct



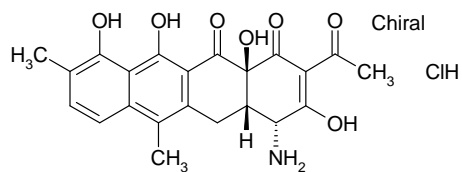
Incorrect

I. Anhydrotetracyclines

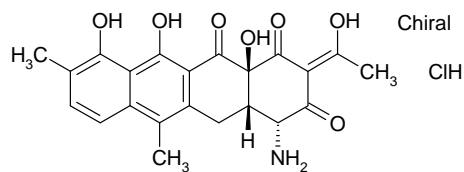
5a-6-Anhydrotetracyclines are drawn from the following sub-structure:



i. Example: CHELOCARDIN HYDROCHLORIDE



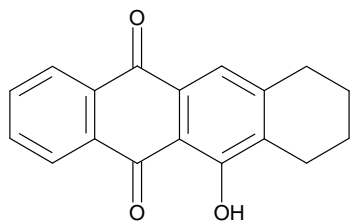
Correct



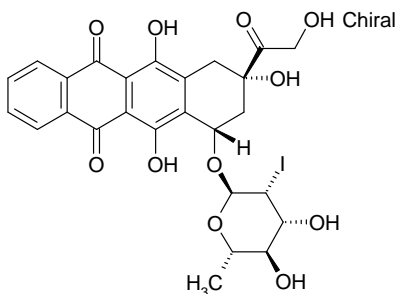
Incorrect

m. Rubicin Antibiotics

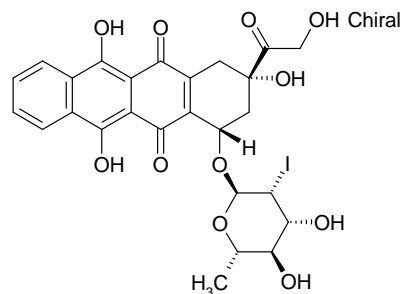
Antibiotics in the rubicin family are drawn from the following sub-structure:



i. Example: ANNAMYCIN



Correct



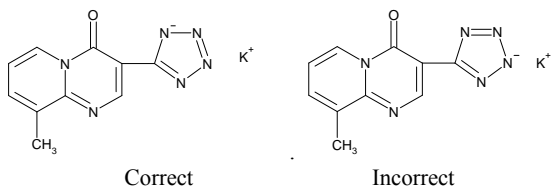
Incorrect

n. Resonance-Stabilized Ions

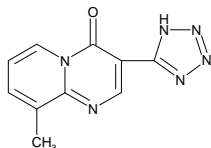
Tautomers are distinct isomers that interconvert, typically by the migration of a hydrogen atom. The same ion is produced regardless of which tautomer loses a proton. The ion produced has a resonance form that corresponds to each tautomer. Draw the resonance form that corresponds to tautomer that is drawn using the conventions listed above.

i. Example. PEMIROLAST POTASSIUM

When PEMIROLAST POTASSIUM loses a proton to form an anion, the anion has the following two resonance forms:



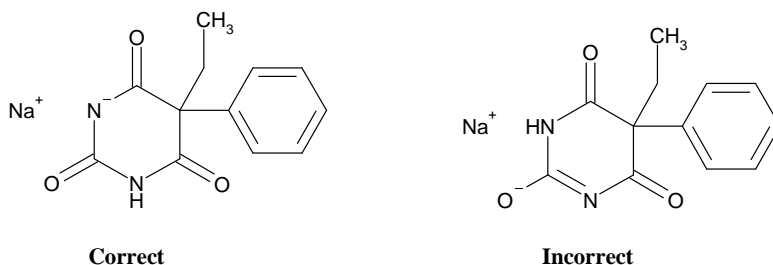
The structure on the left is the correct anion to enter as it corresponds to the following PEMIROLAST tautomer:



The negative charge in the anion and the hydrogen in the free base are both placed at the one position on the tetrazole ring.

When chirality is created in a symmetrical molecule by the loss of a proton, the stereochemistry is not described in either the structure or description field.

ii. Example: PHENOBARBITAL SODIUM



o. Other Types of Tautomerization

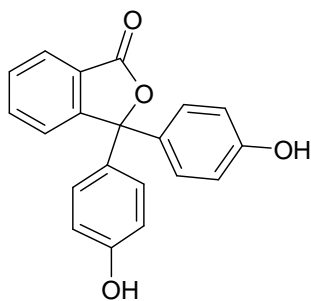
Some types of tautomerization involve more than just a proton shift. The SRS software does not recognize these tautomers as the same substance.¹³

p. Phenolphthalein and Similar Compounds

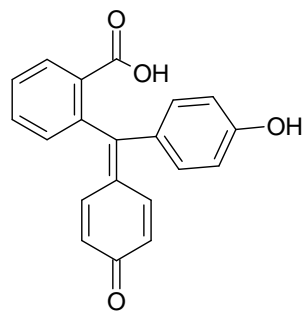
The cyclic ester is drawn for the neutral compounds. The carboxylic acid salt is drawn for the anion.

i. Example: PHENOLPHTHALEIN

¹³ The SRS registrars are responsible for ensuring that only one tautomer is entered into the system.

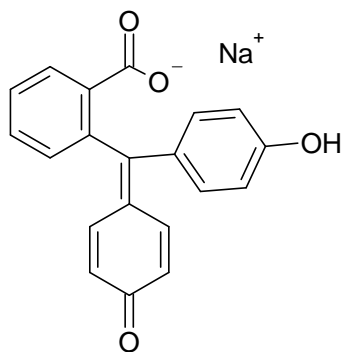


Correct

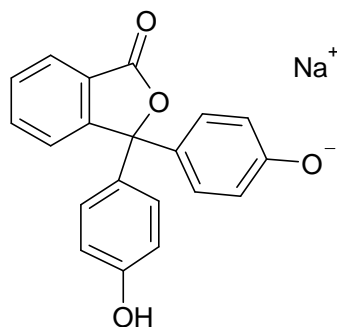


Incorrect

ii. Example: PHENOLPHTHALEIN, SODIUM



Correct



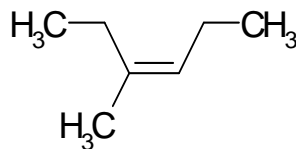
Incorrect

21. Stereochemistry

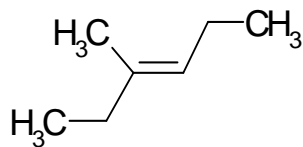
MDL software has the capability to represent stereochemistry around a double bond and around certain tetrahedral stereocenters.

a. E and Z Isomerism (Proposed Not Official)

Molecules that have a double bond are drawn with the substituents arranged according to the double bond's stereochemistry. Substances that are known to contain both E and Z isomers are handled as "ALL OF" mixtures.



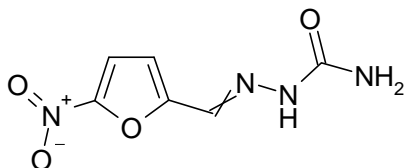
Cis (Z)



Trans (E)

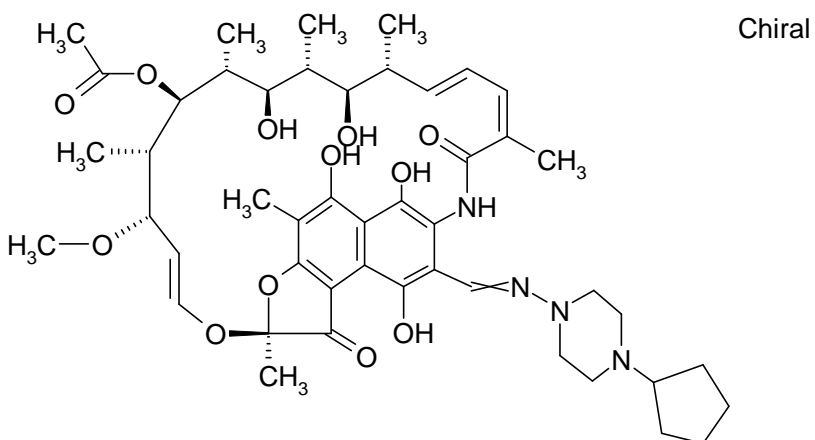
If the configuration of the double bond is unspecified or unknown in the source document, it is represented by the MDL double either bond..

i. Example: NITROFURAZONE



Identifying Description

ii. Example: RIFAPENTINE



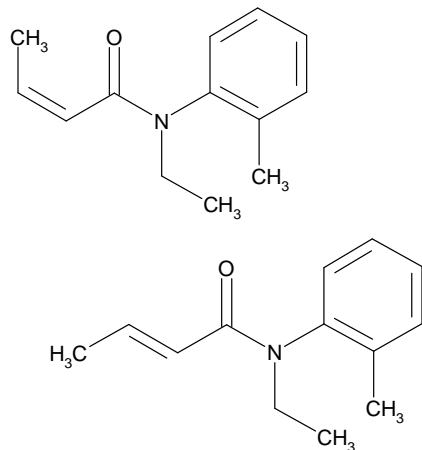
Identifying Description

If a substance is known to be a mixture of Z and E isomers both isomers will be drawn as individual substances and an "ALL OF" mixture will be created to represent the mixture.

iii. Example CROTAMITON

Crotamiton is an "ALL OF" mixture of CIS and TRANS isomers. Both isomers are drawn as separate substances.

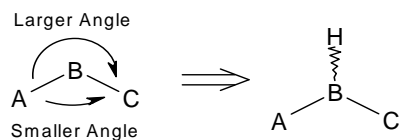
Mixture type: "All Of"



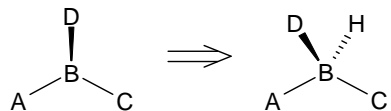
b. General Rules for Tetrahedral Stereochemistry

Chiral atoms are indicated by the presence of a solid or dashed wedge bond.¹⁴ The wider end of the bond is always drawn away from the chiral center. A solid wedge implies that the group attached to the chiral center is projected towards the viewer coming out of the plane formed by the chiral center. A dashed wedge implies the group is projected behind the plane, away from the viewer.

A hydrogen atom is usually not displayed on a chiral center.¹⁵ Their presence is inferred. The MDL software locates this "inferred" hydrogen in the larger angle formed by the central atom and the two substituents that are in the plane of the paper.¹⁶



If the substituent with the wedge bond is also in the larger angle, the hydrogen is given the opposite orientation.

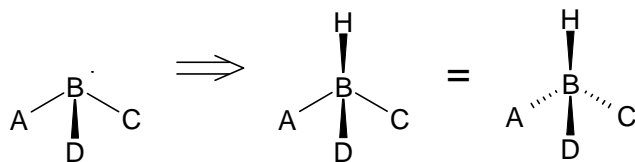


¹⁴ Racemates with only one chiral center do not contain wedge bonds at the chiral center.

¹⁵ In the examples below, hydrogen atoms are included for clarity.

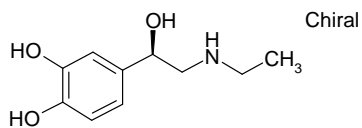
¹⁶ See Appendix A, Chemical Representation in the MDL ISIS™/Base Database Maintenance Manual.

If the substituent with the wedge bond is in the smaller angle, the hydrogen is given the same orientation.

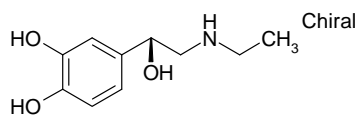


The wedge bond is placed in the larger angle whenever possible.

- i. Example: 4-(R-2-(ETHYLAMINO)-1-HYDROXYETHYL)BENZENE-1,2-DIOL



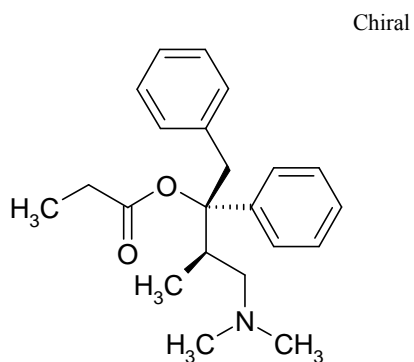
Correct



Incorrect

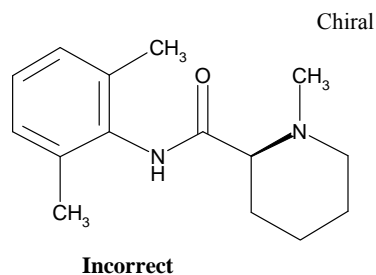
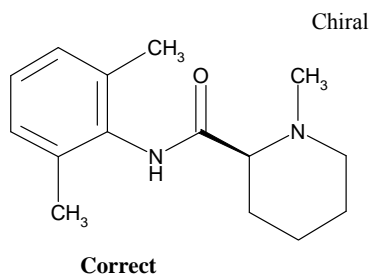
If a chiral center has nonhydrogen atoms attached, a solid wedge is drawn to one atom and dashed wedge bond to another atom; the two other atoms are attached to the center using a normal bond

- ii. Example: PROPOXPHENE



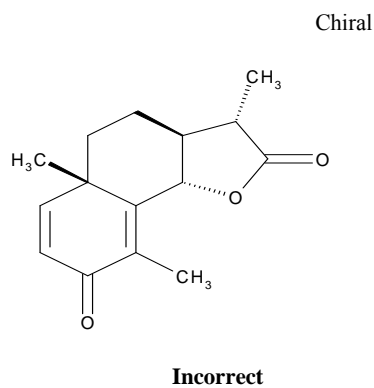
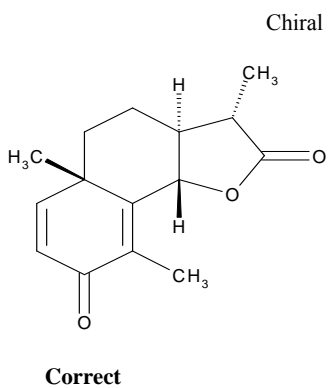
When a chiral center is part of a ring, draw the wedge bond(s) to the group(s) outside the ring.

- iii. Example: MEPIVACAINE



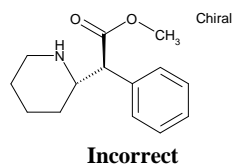
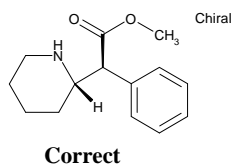
When a chiral center is located at the junction of two rings, an explicit hydrogen is added to indicate the configuration.

iv. Example: SANTONIN



A wedge bond should not be drawn to an adjacent chiral center. To prevent this from occurring, add an explicit hydrogen or reposition the wedge on a different bond.

v. Example: METHYLPHENIDATE



22. Representation of Stereochemical Classes

Substances are classified as follows: (1) enantiomers, (2) racemates, (3) substances with no or partial information about stereochemical orientation and (4) substances which are chiral but the absolute stereochemistry is not known. Substances can have single or multiple stereogenic centers. The following substances are drawn as single substances: (1) racemates, (2) substances which contain no or partial

information on stereochemical orientation and (3) chiral substances where the absolute stereochemistry is not known. Substances that are known to contain several diastereoisomers are drawn as mixtures.

The following tags are used in the Description Field to describe racemic substances and those substances that have unknown or incompletely described stereochemistry. The Substance Registration System (SRS) defines a racemic substance as any substance that contains significant amounts of both enantiomers. A significant amount is 5% of one enantiomer. Substances that contain greater than 95% of a single enantiomer are considered to be a single chiral substance. The Description Field and tags refer to all of the undefined stereochemistry present in a structure or molecular fragment, not to the individual chiral atoms or centers within the structure.

<STEREOCHEMISTRY></STEREOCHEMISTRY> Tags that surround information about stereochemistry.

<OPT_ACT></OPT_ACT>: Indicates information about the optical activity in the substance. An empty field indicates that the optical activity is unknown; (+) indicates that the substance is dextrorotary; (-) indicates that the substance is levorotary; and (+/-) indicates that the substance does not rotate polarized light (used for racemates and meso compounds).

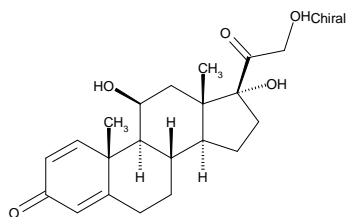
<TYPE></TYPE> Indicates the type of stereochemistry that is not completely defined by the structure. Values include UNKNOWN; RACEMIC; RELATIVE; MESO; AXIAL, UNKNOWN; AXIAL, RACEMIC; AXIAL, R; AXIAL, S. Examples below illustrate the use of each of these terms.

<COMMENTS></COMMENTS> Any comments that could help distinguish the isomers.

a. Completely Defined Enantiomer

For a chiral molecule in which the configuration of all stereocenters is known, the chiral flag is attached to the structure. The Identifying Description Field does not contain any stereochemistry tags.

i. Example: PREDNISOLONE



Identifying Description is left blank.

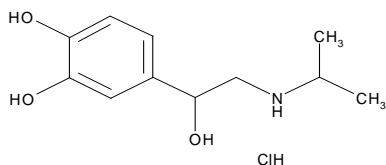
b. Racemic Substances (Optically Inactive) or Substances with Undefined or Unknown Stereochemistry

1. A single chiral center

Substances that contain a single chiral atom that is unspecified or known to be racemic are drawn using a regular (non-wedge) single bond. The Identifying Description Field distinguishes the two cases.

A single structure is drawn for a racemate with one chiral center. Wedge bonds are not used. The chiral center is linked to its neighbors by single bonds.¹⁷ The <OPT_ACT></OPT_ACT> tags contain the value “(+/-)” and the <TYPE></TYPE> tags contain the value RACEMIC.

i. Example: ISOPROTERENOL HYDROCHLORIDE

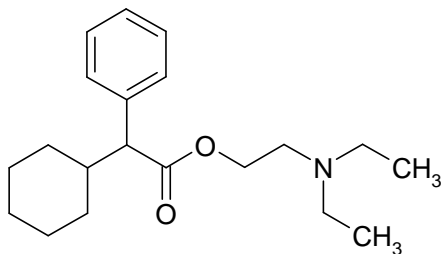


Identifying Description

<STEREOCHEMISTRY><OPT_ACT>(+/</OPT_ACT></STEREOCHEMISTRY>
<TYPE>RACEMIC</TYPE><COMMENTS></COMMENTS></STEREOCHEMISTRY>

A single structure is drawn for substances that contain a single chiral center where the stereochemistry is unknown. The <OPT_ACT></OPT_ACT> tags are left blank. The <TYPE></TYPE> tags contain the value “UNKNOWN”. Normal single bonds are drawn at the stereogenic atom.

ii. Example: DROFENINE



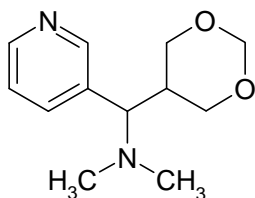
Identifying Description

<STEREOCHEMISTRY><OPT_ACT></OPT_ACT><TYPE>UNKNOWN</TYPE><COMMENTS></COMMENTS></STEREOCHEMISTRY>

A single structure is drawn containing normal single bonds for substances that are known to be chiral but where the absolute stereochemistry is not defined. The <OPT_ACT></OPT_ACT> tag contains the direction in which plane polarized light is rotated, i.e., either “(+)” or “(-)”. The <TYPE></TYPE> tag contains the value “UNKNOWN” .

iii. Example DOXPICOMINE (-)-isomer absolute configuration undefined.

¹⁷ In cases where phosphorus is the chiral atom, a double bond may also be present.



Identifying Description

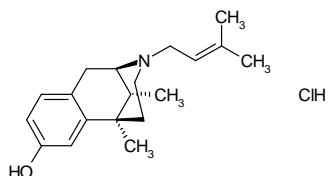
<STEREOCHEMISTRY>OPT_ACT>(-)
</OPT_ACT><TYPE>UNKNOWN<TYPE><COMMENTS></COMMENTS></STEREOCHEMISTRY>

2. Multiple chiral centers

The same tags used to describe substances with a single chiral center are used to describe substances with multiple chiral centers. The tags refer to the entire structure not individual chiral centers.

A single structure is drawn for a racemate that contains more than one chiral center and in which the relative configuration of all the centers is known. Wedge bonds are drawn at each stereocenter. A chiral flag is not used in the structure diagram. The <OPT_ACT></OPT_ACT> tags contain the value “(+/-)” and the <TYPE></TYPE> tags contain the value “RACEMIC”.¹⁸

i. Example: PENTAZOCINE HYDROCHLORIDE



Identifying Description

<STEREOCHEMISTRY><OPT_ACT>(+/-)
</OPT_ACT><TYPE>RACEMIC<TYPE><COMMENTS></COMMENTS></STEREOCHEMISTRY>

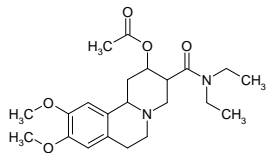
Where there is no information on relative stereochemistry (the synthesis or purification strategy used to prepare the substance is not known) a single structure is drawn that uses normal single bonds at each chiral center and tags in the Identifying Description Field indicate the lack of knowledge. The <OPT_ACT></OPT_ACT> tags do not have a value and the <TYPE></TYPE> tags contain the value “UNKNOWN”

ii. Example: BENZQUINAMIDE

2-(acetyloxy)-*N,N*-diethyl-1,3,4,6,7,11b-hexahydro-9,10-dimethoxy-2*H*-benzo[*a*]quinolizine-3-carboxamide

¹⁸ In the literature, a chemical structure drawn in this format may represent a single enantiomer. In the SRS, this structure represents a racemate or a chiral compound in which the absolute configuration has not been determined.

USAN does not indicate the configuration at any of the chiral centers.



Identifying Description

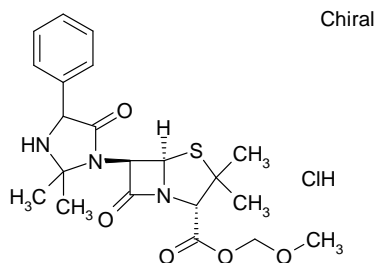
<STEREOCHEMISTRY><OPT_ACT></OPT_ACT><TYPE>UNKNOWN</TYPE><COMMENTS></COMMENTS></STEREOCHEMISTRY>

Substances in which the stereochemistry is partially defined consists of a single record in which all known stereochemistry is drawn and normal bonds are used at chiral centers in which the absolute configuration is unknown. The chiral flag is used to indicate that the stereochemistry defined in the drawing is absolute. <OPT_ACT></OPT_ACT> tags do not have a value and the <TYPE></TYPE> tags contain the value "UNKNOWN".

iii. Example: SARPICILLIN HYDROCHLORIDE

Methoxymethyl (2S,5R,6R)-6-(2,2-dimethyl-5-oxo-4-phenyl-1-imidazolidinyl)-3,3-dimethyl-7-oxo-4-thia-1-azabicyclo(3.2.0)heptane-2-carboxylate hydrochloride

The Chemical Abstract name provides absolute stereochemistry for all the chiral centers except the 4 position on the imidazole ring. The name does not indicate if this is a single enantiomer or a mixture of diastereomers. All the known chirality is drawn and the chiral flag is used. The tags refer to the chiral atoms in the substance in which the stereochemistry is not defined.



Identifying Description

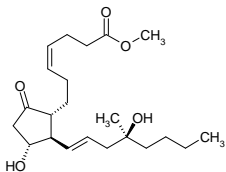
<STEREOCHEMISTRY><OPT_ACT></OPT_ACT><TYPE>UNKNOWN</TYPE><COMMENTS></COMMENTS></STEREOCHEMISTRY>

Both stereoisomers are drawn as an "ALL OF" mixture for substances known to contain two different diastereoisomers or contain racemic mixtures of two diastereoisomers.

iv. Example: ENISOPROST

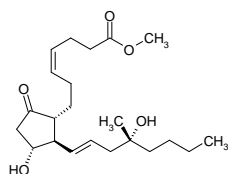
11,16-dihydroxy-16-methyl-9-oxo-, Prosta-4,13-dien-1-oic acid methyl ester, (4Z,11.alpha.,13E)-(+/-)

USAN indicates the relative stereochemistry between all the chiral centers on the cyclopentane ring. The substance contains two diastereomeric racemates and is represented by an "ALL OF" mixture.



Identifying Description

<STEREOCHEMISTRY><OPT_ACT>(+/ -)</OPT_ACT><TYPE>RACEMIC</TYPE><COMMENTS></COMMENTS></STEREOCHEMISTRY>

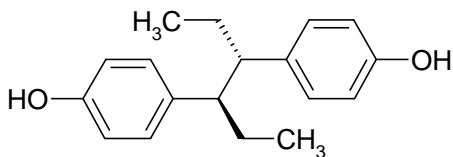


Identifying Description

<STEREOCHEMISTRY><OPT_ACT>(+/ -)</OPT_ACT><TYPE>RACEMIC</TYPE><COMMENTS></COMMENTS></STEREOCHEMISTRY>

v. Example: HEXESTROL (Meso compound)

MESO-3,4-BIS(P-HYDROXYPHENYL)-N-HEXANE



Identifying Description

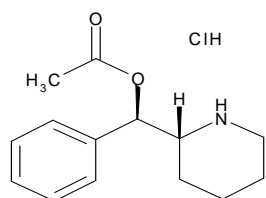
<STEREOCHEMISTRY><OPT_ACT>(+/ -)</OPT_ACT><TYPE>MESO</TYPE><COMMENTS></COMMENTS></STEREOCHEMISTRY>

c. Chiral Substances (Optically Active) with Undetermined Configuration

This section covers the following: (1) substances that are optically active and that have not been completely characterized and (2) optically active mixtures of diastereomers. These substances with undetermined composition are divided into two groups: (1) substances with more than one chiral center in which the relative configuration of all the centers is known but the absolute configuration is not and (2) substances that contain more than one chiral center in which the configuration of one or more chiral centers is unknown.

i. Example: LEVOPHACETOPERANE HYDROCHLORIDE

The Chemical Abstract name indicates that this compound is chiral and that the absolute stereochemistry is either R,R or S,S. The substance is drawn as an R,R isomer. The <TYPE></TYPE> tags indicate RELATIVE



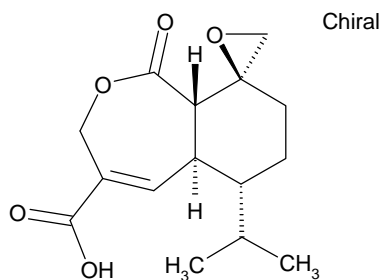
Identifying Description

<STEREOCHEMISTRY><OPT_ACT>(-)
</OPT_ACT><TYPE>RELATIVE</TYPE><COMMENTS></COMMENTS></STEREOCHEMISTRY>

d. Spiro Compounds

Molecules containing two rings joined at a single carbon are called spiro compounds. The atom at the ring junction is connected to four other atoms and may be chiral. When it is chiral, the wedge bonds are placed in a ring. The solid and dashed wedge bonds are placed in the smaller of the two rings whenever possible.

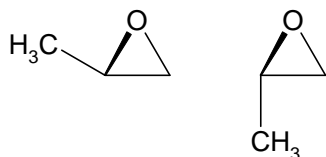
i. Example: HEPTELIDIC ACID



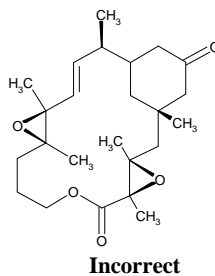
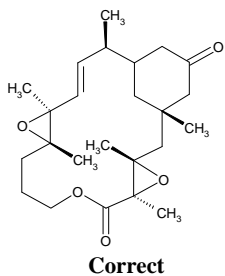
e. Special Stereochemical Cases

Due to the algorithm that MDL uses to determine the absolute configuration of a chiral center, caution must be taken when adding wedge bonds¹⁹. Two situations in which error can arise are 3-membered rings and bridgehead chiral centers.

The MDL software interprets the following epoxides as enantiomers.



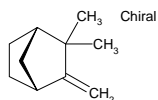
To avoid this problem, the wedge bond is drawn to a group outside the 3-membered ring.



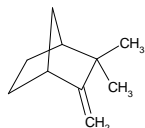
Molecules with ring systems that contain a bridge are drawn "flattened," not in a 3D perspective. (See Section 21b)

¹⁹ See section 21b on the General Rules for Tetrahedral Stereochemistry.

i. Example: (+)-CAMPHENE



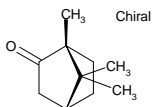
Correct



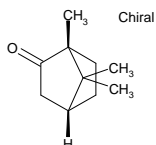
Incorrect

When there is stereochemistry at the bridgehead, the wedge bond is drawn within a ring, not to an exocyclic atom or group.

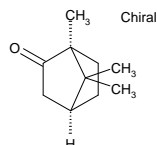
ii. Example: (+)-CAMPHOR



Correct



Incorrect

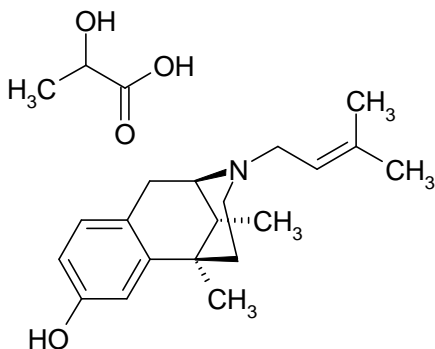


Incorrect: Opposite Enantiomer

If two molecular entities in a single structure field contain chiral centers that are not completely defined by the structure field two sets of stereochemistry tags are necessary. The first set of stereochemistry tags refer to the higher molecular weight species. The comments contain the name of the fragment.

iii. Example: PENTAZOCINE LACTATE

The structure field contains the following chemical drawing.

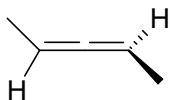


The identifying description tags contain the following tags.

```
<STEREOCHEMISTRY><OPT_ACT>(+/
-)</OPT_ACT><TYPE>RACEMIC</TYPE><COMMENTS>PENTAZOCINE</COMMENTS><
/STEREOCHEMISTRY><STEREOCHEMISTRY><OPT_ACT></OPT_ACT><TYPE>UNKN
OWN </TYPE><COMMENTS>LACTATE</COMMENTS></STEREOCHEMISTRY>
```

f. Axial chirality

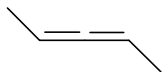
MDL database software is unable to recognize chirality resulting from restricted rotation about sp² hybridized carbons. As a result, compounds like (S)-1,3-Dimethylallene are not recognized as chiral.²⁰



Description tags are used to distinguish the stereoisomers. Consult the CAS Index guide for the methodology used to determine R or S chirality.

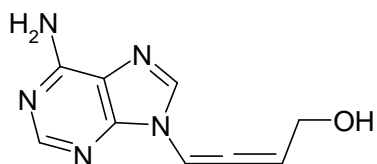
i. Example (S)-1,3-DIMETHYLALLENE

²⁰ If presented in a journal, this is a valid structure. However, the SRS software is not able to interpret the chirality.



<STEREOCHEMISTRY><OPT_ACT>(-)</OPT_ACT><TYPE>AXIAL,
S</TYPE><COMMENTS></COMMENTS></STEREOCHEMISTRY>

- ii. Example ADENALLENE (+/-)



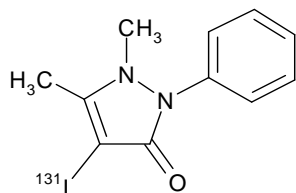
<STEREOCHEMISTRY><OPT_ACT>(+/<-)</OPT_ACT><TYPE>AXIAL,
RACEMIC</TYPE><COMMENTS></COMMENTS></STEREOCHEMISTRY>

23. Isotopically Labeled Compounds

a. Position of Substitution Known

When the position of isotopic labeling is known, the compound is drawn with the appropriate isotope at that position.

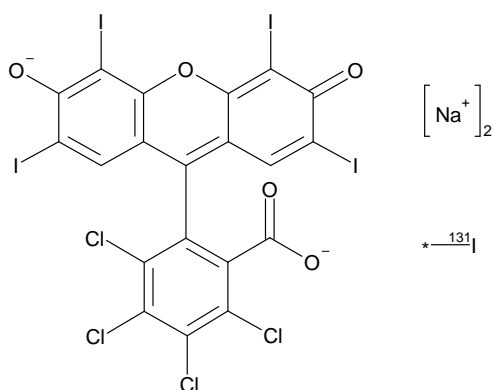
- i. Example: IODOANTIPYRINE I



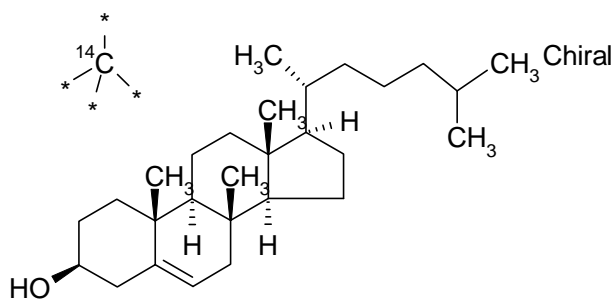
b. Labeled compound with the position of isotopic atom is unknown

When a compound is labeled, but the position of the labeled atom is unknown, the label is drawn unattached to the molecule, but linked to an asterisk(s). The substance could be a pure compound or a mixture. It might have more than one label attached.

- i. Example: ROSE BENGAL SODIUM, I-131, Position of Substitution Not Specified

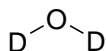


- ii. Example: C-14 CHOLESTEROL; Position of Substitution Not Specified



- iii. Example: DEUTERIUM OXIDE

Deuterium is represented by the superatom D, an isotopically labeled hydrogen is not searchable in the database.



24. Substituted Glycerine Compounds (Guidance needs to be developed)

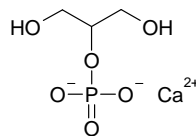
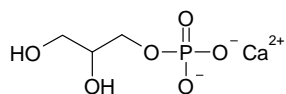
There are a number of substances that contain an ester or ether of glycerine. When no information is available on which hydroxyl group undergoes substitution, it is drawn as an “ANY OF” mixture.

- i. Example: CALCIUM GLYCEROPHOSPHATE

The commercial product is a mixture of calcium β - and (+/-)- α -glycerophosphates.

Mixture Type: "ANY OF"

Comment [s1]: This is not the best example as this should be an "ALL OF" mixture.



25. Carbohydrates and Other Saccharides

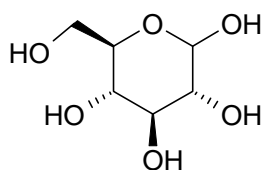
Monosaccharide and the terminal sugar in many polysaccharides are in equilibrium between a chain form and one or more cyclic isomers. These saccharides are represented as a single structure. When multiple cyclic forms are possible, the pyranose isomer is depicted. The anomeric carbon on the terminal end of a monosaccharide or polysaccharide is drawn without any indication of chirality unless the anomeric configuration is explicitly specified.

i. Example: DEXTROSE

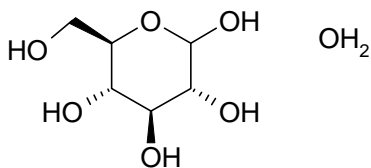
Dextrose as defined in USP can be either glucose monohydrate or glucose anhydrous, and is drawn as an "ANY OF" mixture.

Mixture Type: "ANY OF"

Chiral

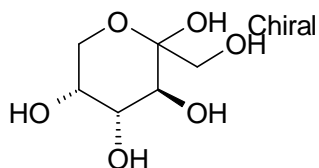


Chiral

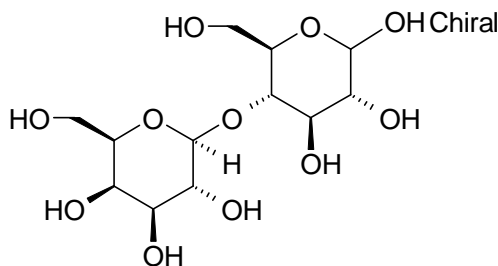


ii. Example: D-FRUCTOSE

Although fructose is a mixture of furanose and pyranose isomers, it is drawn as a single pyranose sugar.



iii. Example LACTOSE, ANHYDROUS



26. Peptides/Proteins.

Since most proteins and peptides are inherently heterogeneous, the SRS system will not attempt to capture all aspects of heterogeneity. Heterogeneity can arise from a wide variety of post-translational modifications. The SRS system will only attempt to capture end-group modification (e.g., amidation, acetylation), glycosylation in a general sense, and other modifications that are essential for activity. For example the SRS will capture γ -carboxylation that is essential for enzymatic activity in a variety of clotting factors, but it will not attempt to capture the extent of C-terminal lysine heterogeneity in antibodies. The assignment of a UNII to protein substances does not imply that two substances with the same UNII are either biosimilar or therapeutically equivalent.

Peptides containing three or less amino acid residues are drawn showing all atoms and will not require an entry into the description field. Peptides/proteins containing up to 100 residues will have an entry in the structure field and will be drawn using the standard single letter amino acid abbreviations for standard amino acids.²¹ All peptides/proteins with well defined sequences that are greater than three amino acids will be described using the description field tags. The SRS description field does not indicate if a peptide/protein was chemically synthesized, naturally derived, or produced using a recombinant expression system. For glycoproteins, the SRS does not indicate the extent of glycosylation - only that a protein is glycosylated and the type of glycosylation in a general sense. The SRS does not indicate the state of aggregation or self-aggregation due to noncovalent interactions. For peptides up to 100 amino acids in length, the structure field indicates intra- or intermolecular disulfide linkages, whereas the description field does not. For protein conjugates, the conjugate is drawn in the structure field (e.g., PEG, toxin) and

²¹ Peptides and oligonucleotides are an exception to the rule that abbreviations are not used for chemical groups.

includes the first atom of the protein to which it is linked but the protein is described in the description field.

The sequence of a protein or peptide is the most important feature in distinguishing one from another. Prior to entry of a new substance, search the description field using the protein/peptide's sequence. For monoclonal antibodies sequences outside of the complementarity determining regions (CDR) of the antibody may be common so the entire sequence should be included in the search. An effort should be made to determine information for all of the following fields, particularly when a partial sequence or no sequence is available:

<PROTEIN_TYPE>, <PROTEIN_SUBTYPE>, <SPECIES_TYPE>, <LIGAND>, <GENE_ID>.

a. Description Field.

Defined tags for peptides/proteins:

<PROTEIN></PROTEIN> Parent tags that indicate the substance is a protein or peptide.

<SEQUENCE_TYPE></SEQUENCE_TYPE> Indicates whether the sequence is "COMPLETE", "PARTIAL", OR "UNKNOWN."

<SUBUNITS></SUBUNITS> Indicates the number of subunits present in a protein. A subunit is defined as a complete linear sequence of amino acids linked through normal peptide bonds. Peptides or proteins linked through disulfide or other covalent bonds are considered separate subunits. Examples: Somatropin contains one subunit, insulin contains two subunits, and most monoclonal antibodies contain four subunits. <SUBUNIT_GROUP></SUBUNIT_GROUP> Indicates the beginning of the description for a given subunit. Subunits are listed in order of decreasing length of their amino-acid sequence. If a protein contains two identical subunits, the information is added twice to the description field.

<SUBUNIT_ID></SUBUNIT_ID> Numeric tag identifying the subunit, starting with 1.

<LENGTH></LENGTH> Second tag between <SUBUNIT_GROUP> tags indicates the length of the subunit in number of amino acid residues. This tag is used for a quality control check. When the sequence is either unknown or partial, the length tag is left blank.

<SEQUENCE></SEQUENCE> Third tag between <SUBUNIT_GROUP> tags containing the sequence using the standard twenty single-letter amino acid codes. Sequences are listed from the N-terminal residue to the C-terminal residue. X is used to indicate the presence of a non-standard amino acid or an unknown amino acid. X is identified in the <MODIFICATION> tags.

<N_TERMINAL_MOD></N_TERMINAL_MOD> Fourth tag between <SUBUNIT_GROUP> tags indicating the presence of an N-terminal modification. Use the following entries: PYROGLU, ACETYL, FORMYL, MYRISTOYL. A blank entry indicates either the absence of an N-terminal modification or that no information is available.

<C_TERMINAL_MOD></C_TERMINAL_MOD> Last tag between <SUBUNIT_GROUP> tags indicating the presence of a C-terminal modification. Use the following entries: AMIDE, ETHYL ESTER, HYDRAZIDE. A blank entry indicates either the absence of a C-terminal modification or that no information is available.

<MODIFICATION></MODIFICATION> Tags used to describe any modified amino-acids present in the sequence or any modifications of a protein that are essential for activity. Entry of modification begins with the first subunit and first residue. If the position of a modification is not specified, the modification is listed following specified modifications. Modifications are delimited by semi-colons, data elements within a modification are delimited by a comma. The extent of conjugation should be listed if known. A blank field indicates there are no modifications that are essential for activity present in a protein, or that modifications are not known to be present in a protein. Examples: 1_14=D-PHE, indicates that the fourteenth residue on the first subunit is a D-Phenylalanine residue; CONJUGATE=OZOGAMICIN, 4 CONJUGATES PER MOLECULE indicates that four ozogamicin molecules are conjugated to each protein; the location of the conjugation is not known or can vary.

<GLYCOSYLATED></GLYCOSYLATED> Indicates whether a protein is glycosylated. The field will contain the type of glycosylation. The accepted values are YEAST, YEAST (HUMANIZED), INSECT, PLANT, MAMMALIAN, MAMMALIAN (-FUCOSE) or HUMAN, when a protein is glycosylated. The type of glycosylation can be determined from the cells that the protein is produced in. Recombinant proteins produced in CHO, VERO, or MOUSE cell lines or derived from mammalian tissues will use the term MAMMALIAN. Proteins produced in human cell lines or derived from human tissues will use the

term HUMAN. If the field is blank, the protein is either not glycosylated or if it is not known whether the protein is glycosylated. The extent and sites of glycosylation will not be indicated. The SRS will also not indicate whether glycosylation is biantennary, triantennary etc., or the extent of sialylation. Proteins that contain a different sites or extent of glycosylation are not distinguished from one another and are given the same Unique Identifier (UNII).

<PEGYLATED></PEGYLATED> Indicates whether a protein is pegylated. The field contains the value "YES" if a protein is known to be pegylated. A representation of the polyethylene glycol (PEG) chain and the linking group are drawn in the structure field. A blank field indicates that the protein is not pegylated or no information is available on pegylation..

<PROTEIN_TYPE></PROTEIN_TYPE> Indicates the type of protein. Examples: MONOCLONAL ANTIBODY; ENZYME; FC FUSION PROTEIN; RECEPTOR, SOLUBLE; HORMONE.

<PROTEIN_SUBTYPE></PROTEIN_SUBTYPE>Indicates the subtype of the protein, i.e. the subtype of antibody, the name of an enzyme or ligand, or the description of heavy and light chains. Example: IGG1, KAPPA, IGG4, LAMDA, .ALPHA.-GLUCOSIDASE, TNF-RECEPTOR-FC FUSION, GROWTH FACTOR

<SPECIES_TYPE></SPECIES_TYPE> Indicates the species of origin of the protein. This is not the species of the cell-line from which it is produced. Examples: HUMAN; MURINE; CHIMERIC, HUMAN/MURINE; HUMANIZED

<LIGAND></LIGAND>The ligand or natural substrate to which the antibody or enzyme binds. For human protein targets the NCBI GENE BANK id number is used followed by the common name or title for the receptor or ligand from GENE BANK. This can be found by searching the GENE database on Entrez: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene&cmd=Retrieve&dopt=full_report&list_uids=3643 . For non-human targets a common name of the ligand can be used.

<GENE_ID></GENE_ID> Refers to the human gene that the peptide or protein was derived from. This field can be filled in for even truncated peptides or proteins. The gene ID will be obtained from NCBI genebank. The entry should consist of the NCBI genebank id number followed by the common name also from genebank

<COMMENTS></COMMENTS> Comments used to distinguish one protein from another when the tags above are not sufficient.

b. Structure Field and Complete Polypeptide/Protein Substance Definitions

In the structure field, peptides are displayed with 15 amino acids per line and are divided into groups of five.

- i. Example: COSYNTROPIN

Structure Field

```
NH2—S Y S M E   H F R W G   K P V G KChiral
      K R R P V   K V Y P—COOH
```

Description Field

```
<PROTEIN>
  <SEQUENCE_TYPE>COMPLETE</SEQUENCE_TYPE>
  <SUBUNITS>1</SUBUNITS>
  <SUBUNIT_GROUP>
    <SUBUNIT_ID>1</SUBUNIT_ID>
    <LENGTH>24</LENGTH>
    <SEQUENCE>SYSMEHFRWKGKPVGKRRRPVKVYP</SEQUENCE>
    <N_TERMINAL_MOD></N_TERMINAL_MOD>
```

```

      <C_TERMINAL_MOD></C_TERMINAL_MOD>
    </SUBUNIT_GROUP>
    <MODIFICATION></MODIFICATION>
    <GLYCOSYLATED></GLYCOSYLATED>
    <PEGYLATED></PEGYLATED>
    <PROTEIN_TYPE>HORMONE</PROTEIN_TYPE>
    <PROTEIN_SUBTYPE>CORTICOTROPIN</PROTEIN_SUBTYPE>
    <SPECIES_TYPE>HUMAN</SPECIES_TYPE>
    <LIGAND>4158, MC2R MELANOCORTIN 2 RECEPTOR (ADRENOCORTICOTROPIC
    HORMONE) </LIGAND>
    <GENE_ID>5443, POMC PROOPIOMELANOCORTIN (ADRENOCORTICOTROPIN/
    BETA-LIPOTROPIN/ ALPHA-MELANOCYTE STIMULATING HORMONE/ BETA-
    MELANOCYTE STIMULATING HORMONE/ BETA-ENDORPHIN) </GENE_ID>
    <COMMENTS></COMMENTS>
  </PROTEIN>

```

In the structure field, amino acids are assumed to be L unless otherwise specified. D-Amino acids are abbreviated using the three letter code preceded by D- (for instance, D-alanine is abbreviated D-Ala).

ii. Example: BIVALIRUDIN

Structure Field

```

  D-Phe-P R P G   G G G N G   D F E E |Chiral
      P E E Y L—COOH

```

Description Field

```

<PROTEIN>
  <SEQUENCE_TYPE>COMPLETE</SEQUENCE_TYPE>
  <SUBUNITS>1</SUBUNITS>
  <SUBUNIT_GROUP>
    <SUBUNIT_ID>1</SUBUNIT_ID>
    <LENGTH>20</LENGTH>
    <SEQUENCE>XPRPGGGGNGDFEIEPEEYL</SEQUENCE>
    <N_TERMINAL_MOD> </N_TERMINAL_MOD>
    <C_TERMINAL_MOD> </C_TERMINAL_MOD>
  </SUBUNIT_GROUP>
  <MODIFICATION>1_1=DPHE</MODIFICATION>
  <GLYCOSYLATED></GLYCOSYLATED>
  <PEGYLATED></PEGYLATED>
  <PROTEIN_TYPE></PROTEIN_TYPE>
  <PROTEIN_SUBTYPE></PROTEIN_SUBTYPE>
  <SPECIES_TYPE></SPECIES_TYPE>
  <LIGAND>2147, F2 COAGULATION FACTOR II (THROMBIN)</LIGAND>
  <GENE_ID></GENE_ID>
  <COMMENTS></COMMENTS>
</PROTEIN>

```

C-Terminal amides are abbreviated @-NH₂ where @ represents a one letter amino acid code. They are created using the ISIS Draw residue tool. The amino acid is first expanded and the residue portion eliminated. An NH₂ is then added. A new residue is then created given the abbreviation @-NH₂. N-Terminal acetyl group is handled in a like manner. The new residue is given the abbreviation Ac-@.

iii. Example: ENFUVIRTIDE

Structure Field

Ac-Y-T S L I H S L I E E S Q N QChiral
Q E K N E Q E L L E L D K W A
S L W N W F-NH₂

Description field

```
<PROTEIN>  
<SEQUENCE_TYPE>COMPLETE</SEQUENCE_TYPE>  
<SUBUNITS>1</SUBUNITS>  
<SUBUNIT_GROUP>  
  <SUBUNIT_ID>1</SUBUNIT_ID>  
  <LENGTH>36</LENGTH>  
  <SEQUENCE>YTSLIHSLIEESQNQEKNEQELLELDKWASLWNWF</SEQUENCE>  
  <N_TERMINAL_MOD>ACETYL</N_TERMINAL_MOD>  
  <C_TERMINAL_MOD>AMIDE</C_TERMINAL_MOD>  
</SUBUNIT_GROUP>  
<MODIFICATION></MODIFICATION>  
<GLYCOSYLATED></GLYCOSYLATED>  
<PEGYLATED></PEGYLATED>  
<PROTEIN_TYPE></PROTEIN_TYPE>  
<PROTEIN_SUBTYPE></PROTEIN_SUBTYPE>  
<SPECIES_TYPE></SPECIES_TYPE>  
<LIGAND>HIV-1 GP41</LIGAND>  
<GENE_ID></GENE_ID>  
<COMMENTS></COMMENTS>  
</PROTEIN>
```

Other nonstandard amino acids are drawn expanded, showing all atoms, within the peptide chain.

iv. Example: LUTEINIZING HORMONE

5- Pyroglutamate does not have a standard single letter abbreviation and is drawn expanded. The letter X is used in the <SEQUENCE></SEQUENCE> tags to indicate that pyroglutamate (PYROGLU) is not a standard amino acid.

Structure Field



Description field

```

<PROTEIN>
  <SEQUENCE_TYPE>COMPLETE</SEQUENCE_TYPE>
  <SUBUNITS>1</SUBUNITS>
  <SUBUNIT_GROUP>
    <SUBUNIT_ID>1</SUBUNIT_ID>
    <LENGTH>10</LENGTH>
    <SEQUENCE>XHWSYGLRPG</SEQUENCE>
    <N_TERMINAL_MOD>PYROGLU</N_TERMINAL_MOD>
    <C_TERMINAL_MOD>AMIDE</C_TERMINAL_MOD>
  </SUBUNIT_GROUP>
  <MODIFICATION>1_1=PYROGLU</MODIFICATION>
  <GLYCOSYLATED></GLYCOSYLATED>
  <PEGYLATED></PEGYLATED>
  <PROTEIN_TYPE>HORMONE</PROTEIN_TYPE>
  <PROTEIN_SUBTYPE></PROTEIN_SUBTYPE>
  <SPECIES_TYPE>HUMAN</SPECIES_TYPE>
  <LIGAND>3973, LHCGR LUTEINIZING HORMONE/CHORIOGONADOTROPIN
  RECEPTOR</LIGAND>
  <GENE_ID></GENE_ID>
  <COMMENTS></COMMENTS>
</PROTEIN>

```

The SRS software indicates disulfide bonds by placing x-number (e.g. x2) above the cysteine residues.

v. Example: INSULIN HUMAN

Structure Field

```

NH2—G I V E Q      Cx2 Cx3 T S I      Cx2 S L Y Q      Chiral
      L E N Y Cx1      N—COOH

NH2—F V N Q H      L Cx3 G S H      L V E A L
      Y L V Cx1 G      E R G F F      Y T P K T—COOH

```

Description field

```

<PROTEIN>
  <SEQUENCE_TYPE>COMPLETE</SEQUENCE_TYPE>
  <SUBUNITS>2</SUBUNITS>
  <SUBUNIT_GROUP>
    <SUBUNIT_ID>1</SUBUNIT_ID>
    <LENGTH>30</LENGTH>
    <SEQUENCE>FVNQHLCGSHLVEALYLVCGERGFFYTPKT</SEQUENCE>
    <N_TERMINAL_MOD></N_TERMINAL_MOD>
    <C_TERMINAL_MOD></C_TERMINAL_MOD>
  </SUBUNIT_GROUP>
  <SUBUNIT_GROUP>
    <SUBUNIT_ID>2</SUBUNIT_ID>
    <LENGTH>21</LENGTH>
    <SEQUENCE>GIVEQCCTSICSLYQLENYCN</SEQUENCE>

```

```

    <N_TERMINAL_MOD></N_TERMINAL_MOD>
    <C_TERMINAL_MOD></C_TERMINAL_MOD>
  </SUBUNIT_GROUP>
  <MODIFICATION></MODIFICATION>
  <GLYCOSYLATED></GLYCOSYLATED>
  <PEGYLATED></PEGYLATED>
  <PROTEIN_TYPE>HORMONE</PROTEIN_TYPE>
  <PROTEIN_SUBTYPE>INSULIN</PROTEIN_SUBTYPE>
  <SPECIES_TYPE>HUMAN</SPECIES_TYPE>
  <LIGAND>3643, INSULIN RECEPTOR</LIGAND>
  <GENE_ID>3630, INSULIN</GENE_ID>
  <COMMENTS></COMMENTS>
</PROTEIN>

```

Peptides with more than 100 residues are defined using only the description field. There is no entry in the structure field.

vi. Example: SOMATROPIN

Structure Field: Blank

Description field

```

<PROTEIN>
  <SEQUENCE_TYPE>COMPLETE</SEQUENCE_TYPE>
  <SUBUNITS>1</SUBUNITS>
  <SUBUNIT_GROUP>
    <SUBUNIT_ID>1</SUBUNIT_ID>
    <LENGTH>191</LENGTH>
    <SEQUENCE>FPTIPLSRLFDNAMLRAHRLHQLAFDITYQEFEEAYIPKEQKYSFLQNPQ
    TSLCFSES IPTSPNREETQQKSNLELLRISLLLIQSWLEPVQFLRSVFANSLVYGASDSN
    VYDLLKDLEEGIQTLMGRLEDGSPRTGQIFKQTYSKFDTNSHNDDALLKNYGLLYCF
    RKDMDKVETFLRIVQCRSVEGSCGF</SEQUENCE>
    <N_TERMINAL_MOD></N_TERMINAL_MOD>
    <C_TERMINAL_MOD></C_TERMINAL_MOD>
  </SUBUNIT_GROUP>
  <MODIFICATION></MODIFICATION>
  <GLYCOSYLATED></GLYCOSYLATED>
  <PEGYLATED></PEGYLATED>
  <PROTEIN_TYPE></PROTEIN_TYPE>
  <PROTEIN_SUBTYPE></PROTEIN_SUBTYPE>
  <SPECIES_TYPE></SPECIES_TYPE>
  <LIGAND> 2690, GROWTH HORMONE RECEPTOR </LIGAND>
  <GENE_ID>2688, GROWTH HORMONE 1 </GENE_ID>
  <COMMENTS></COMMENTS>
</PROTEIN>

```

vii. Example: EFALIZUMAB

Structure Field: Blank

Description field

```

<PROTEIN>
  <SEQUENCE_TYPE>COMPLETE</SEQUENCE_TYPE>

```

```

<SUBUNITS>4</SUBUNITS>
<SUBUNIT_GROUP>
  <SUBUNIT_ID>1</SUBUNIT_ID>
  <LENGTH>451</LENGTH>
  <SEQUENCE>EVQLVESGGGLVQPGGSLRLSCAASGYSFTGHWMNWVRQAPGKGLE
WVGMIHPSDSETRYNQKFKDRFTISVDKSKNTLYLQMNSLRAEDTAVYYCARGIYFY
GTTYFDYWGQGTLLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVS
WNSGALTSGVHTFPAVLQSSGLYSLSSVTVPSSSLGTQTYICNVNHKPSNTKVDKK
VEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVDVSHEDPEV
KFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKA
LPAPIEKISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNG
QPENNYKTTTPVLDSDGSFFLYSKLTVDKSRWQQGNVFCSSVMHEALHNHYTQKSL
SLSPGK</SEQUENCE>
  <N_TERMINAL_MOD></N_TERMINAL_MOD>
  <C_TERMINAL_MOD></C_TERMINAL_MOD>
</SUBUNIT_GROUP>
<SUBUNIT_GROUP>
  <SUBUNIT_ID>2</SUBUNIT_ID>
  <LENGTH>451</LENGTH>
  <SEQUENCE>EVQLVESGGGLVQPGGSLRLSCAASGYSFTGHWMNWVRQAPGKGLE
WVGMIHPSDSETRYNQKFKDRFTISVDKSKNTLYLQMNSLRAEDTAVYYCARGIYFY
GTTYFDYWGQGTLLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVS
WNSGALTSGVHTFPAVLQSSGLYSLSSVTVPSSSLGTQTYICNVNHKPSNTKVDKK
VEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVDVSHEDPEV
KFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKA
LPAPIEKISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNG
QPENNYKTTTPVLDSDGSFFLYSKLTVDKSRWQQGNVFCSSVMHEALHNHYTQKSL
SLSPGK</SEQUENCE>
  <N_TERMINAL_MOD></N_TERMINAL_MOD>
  <C_TERMINAL_MOD></C_TERMINAL_MOD>
</SUBUNIT_GROUP>
<SUBUNIT_GROUP>
  <SUBUNIT_ID>3</SUBUNIT_ID>
  <LENGTH>214</LENGTH>
  <SEQUENCE>DIQMTQSPSSLSASVGDRTITCRASKTISKYLAWYQQKPGKAPKLLI
YSGSTLQSGVPSRFRSGSGSDTDFLTISLQPEDFATYYCQQHNEYPLTFGQGTKVEIK
RTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVT
EQDSKDSYSLSTLTLSKADYEEKHKVYACEVTHQGLSSPVTKSFNRGEC</SEQUEN
CE>
  <N_TERMINAL_MOD></N_TERMINAL_MOD>
  <C_TERMINAL_MOD></C_TERMINAL_MOD>
</SUBUNIT_GROUP>
<SUBUNIT_GROUP>
  <SUBUNIT_ID>4</SUBUNIT_ID>
  <LENGTH>214</LENGTH>
  <SEQUENCE>DIQMTQSPSSLSASVGDRTITCRASKTISKYLAWYQQKPGKAPKLLI
YSGSTLQSGVPSRFRSGSGSDTDFLTISLQPEDFATYYCQQHNEYPLTFGQGTKVEIK
RTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVT
EQDSKDSYSLSTLTLSKADYEEKHKVYACEVTHQGLSSPVTKSFNRGEC</SEQUEN
CE>
  <N_TERMINAL_MOD></N_TERMINAL_MOD>
  <C_TERMINAL_MOD></C_TERMINAL_MOD>
</SUBUNIT_GROUP>
<MODIFICATION></MODIFICATION>
<GLYCOSYLATED> MAMMALIAN </GLYCOSYLATED>
<PEGYLATED></PEGYLATED>

```



```

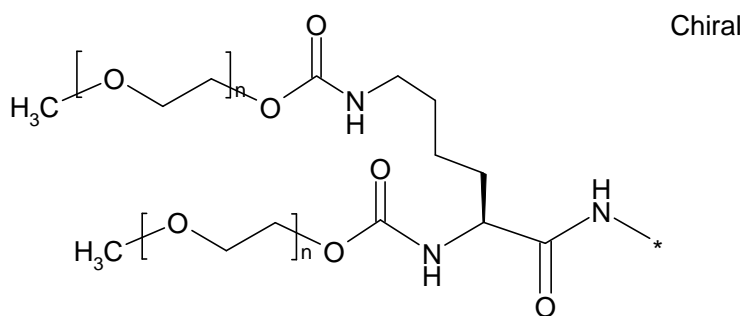
<PROTEIN_TYPE>MONOCLONAL ANTIBODY</PROTEIN_TYPE>
<PROTEIN_SUBTYPE>IGG1 KAPPA</PROTEIN_SUBTYPE>
<SPECIES_TYPE>CHIMERIC, HUMAN/MURINE</SPECIES_TYPE>
<LIGAND>3683, ITGAL INTEGRIN, ALPHA L (ANTIGEN CD11A (P180), LYMPHOCYTE
FUNCTION-ASSOCIATED ANTIGEN 1; ALPHA POLYPEPTIDE)</LIGAND>
<GENE_ID></GENE_ID>
<COMMENTS></COMMENTS>
</PROTEIN>

```

viii. Example: PEGINTERFERON ALFA-2A

This substance is described using both polymer and protein tags. When two sets of tags are present the order is alphabetical according to the parent tags. No delimiter is added between the sets of tags. The molecular weight given in the polymer tags is for the polymer portion of the molecule and not the entire molecule. For details on the polymer tags, see section 27 on Polymers.

Structure



Description field

```

<POLYMER>
<POLYMER_TYPE> HOMOPOLYMER</POLYMER_TYPE>
<NUMBER_OF_SRU>1</NUMBER_OF_SRU>
<SRU_GROUP>
  <SRU_ID>1</SRU_ID>
  <ORIENTATION_OF_POLYMERIZATION>HEAD-
  TAIL</ORIENTATION_OF_POLYMERIZATION>
  <SRU_PERCENT>
    <AVERAGE></AVERAGE>
    <LOW_LIMIT></LOW_LIMIT>
    <HIGH_LIMIT></HIGH_LIMIT>
  </SRU_PERCENT>
  <R_GROUP>
    <R_ID></R_ID>
    <LIMIT_TYPE></LIMIT_TYPE>
    <AVERAGE></AVERAGE>
    <LOW_LIMIT></LOW_LIMIT>
    <HIGH_LIMIT></HIGH_LIMIT>
  </R_GROUP>
</SRU_GROUP>

```

```

    <DEGREE_POLYMERIZATION>
      <AVERAGE></AVERAGE>
      <LOW_LIMIT></LOW_LIMIT>
      <HIGH_LIMIT></HIGH_LIMIT>
    </DEGREE_POLYMERIZATION>
    <MW_GROUP>
      <TYPE_MW> NUMBER AVERAGE </TYPE_MW>
      <MW_AVERAGE>40000</MW_AVERAGE>
      <LOW_LIMIT_MW></LOW_LIMIT_MW>
      <HIGH_LIMIT_MW></HIGH_LIMIT_MW>
    </MW_GROUP>
    <PHYSICAL_PROPERTIES_GROUP>
      <PHYSICAL_PROPERTY_TYPE></PHYSICAL_PROPERTY_TYPE>
      <AVERAGE></AVERAGE>
      <LOW_LIMIT></LOW_LIMIT>
      <HIGH_LIMIT></HIGH_LIMIT>
      <UNITS></UNITS>
    </PHYSICAL_PROPERTIES_GROUP>
    <COMMENTS></COMMENTS>
  </POLYMER>
  <PROTEIN>
    <SEQUENCE_TYPE>COMPLETE</SEQUENCE_TYPE>
    <SUBUNITS>1</SUBUNITS>
    <SUBUNIT_GROUP>
      <SUBUNIT_ID>1</SUBUNIT_ID>
      <LENGTH>165</LENGTH>
      <SEQUENCE>CDLPQTHSLGSRRTLMLLAQMRKISLFSCLKDRHDFGFPQEEFGNQFQ
      KAETIPVLHEMIQQIFNLFSTKDSSAAWDETLDDKPHYTELYQQLNDLEACVIQGVGVT
      ETPLMKEDSILAVRKYFQRITLYLKEKKYSPCAWEVVRAEIMRSFSLSTNLQESLSRSK
      E</SEQUENCE>
      <N_TERMINAL_MOD></N_TERMINAL_MOD>
      <C_TERMINAL_MOD></C_TERMINAL_MOD>
    </SUBUNIT_GROUP>
    <MODIFICATION>ONE PEG MOIETY PER MOLECULE</MODIFICATION>
    <GLYCOSYLATED></GLYCOSYLATED>
    <PEGYLATED>YES</PEGYLATED>
    <PROTEIN_TYPE>CYTOKINE</PROTEIN_TYPE>
    <PROTEIN_SUBTYPE>INTERFERON</PROTEIN_SUBTYPE>
    <SPECIES_TYPE>HUMAN</SPECIES_TYPE>
    <LIGAND>3455, IFNAR2 INTERFERON (ALPHA, BETA AND OMEGA) RECEPTOR 2
    </LIGAND>
    <GENE_ID>3440, IFNA2 INTERFERON, ALPHA 2</GENE_ID>
    <COMMENTS></COMMENTS>
  </PROTEIN>

```

ix. Example. INFLIXIMAB

Structure Field: Blank

Description field

```

  <PROTEIN>
    <SEQUENCE_TYPE>PARTIAL</SEQUENCE_TYPE>
    <SUBUNITS>4</SUBUNITS>
    <SUBUNIT_GROUP>
      <SUBUNIT_ID>1</SUBUNIT_ID>

```

```

<LENGTH></LENGTH>
<SEQUENCE>EVKLEESGGGLVQPGGSMKLSKVASGFIFSNHWMNWVRSPEKGLE
WVAEIRSKSINSATHYAESVKGRFTISRDDSKSAVYLQMTDLRTEDTGVIYCSRNY
GSTYDYWGQGTTLTVS</SEQUENCE>
<N_TERMINAL_MOD></N_TERMINAL_MOD>
<C_TERMINAL_MOD></C_TERMINAL_MOD>
</SUBUNIT_GROUP>
</SUBUNITS>
<SUBUNIT_GROUP>
  <SUBUNIT_ID>2</SUBUNIT_ID>
  <LENGTH></LENGTH>
  <SEQUENCE>EVKLEESGGGLVQPGGSMKLSKVASGFIFSNHWMNWVRSPEKGLE
WVAEIRSKSINSATHYAESVKGRFTISRDDSKSAVYLQMTDLRTEDTGVIYCSRNY
GSTYDYWGQGTTLTVS</SEQUENCE>
  <N_TERMINAL_MOD></N_TERMINAL_MOD>
  <C_TERMINAL_MOD></C_TERMINAL_MOD>
</SUBUNIT_GROUP>
<SUBUNIT_GROUP>
  <SUBUNIT_ID>3</SUBUNIT_ID>
  <SEQUENCE>DILLTQSPAILSVPGERVFSFSCRASQFVGSIIHWYQRTNGSPRLIKY
ASEMSGIPSRFSGSGGTDFTLSINTVESEDIADYICQSSHWPFTFGSGTNLEVK</S
EQUENCE>
  <N_TERMINAL_MOD></N_TERMINAL_MOD>
  <C_TERMINAL_MOD></C_TERMINAL_MOD>
</SUBUNIT_GROUP>
<SUBUNIT_GROUP>
  <SUBUNIT_ID>4</SUBUNIT_ID>
  <SEQUENCE>DILLTQSPAILSVPGERVFSFSCRASQFVGSIIHWYQRTNGSPRLIKY
ASEMSGIPSRFSGSGGTDFTLSINTVESEDIADYICQSSHWPFTFGSGTNLEVK</S
EQUENCE>
  <N_TERMINAL_MOD></N_TERMINAL_MOD>
  <C_TERMINAL_MOD></C_TERMINAL_MOD>
</SUBUNIT_GROUP>
<MODIFICATION></MODIFICATION>
<GLYCOSYLATED></GLYCOSYLATED>
<PEGYLATED></PEGYLATED>
<PROTEIN_TYPE> MONOCLONAL ANTIBODY </PROTEIN_TYPE>
<PROTEIN_SUBTYPE> IGG1, KAPPA </PROTEIN_SUBTYPE>
<SPECIES_TYPE> CHIMERIC, HUMAN/MURINE </SPECIES_TYPE>
<LIGAND> 7124, TUMOR NECROSIS FACTOR </LIGAND>
<GENE_ID></GENE_ID>
<COMMENTS></COMMENTS>
</PROTEIN>

```

Radiolabeled Proteins will be handled like other radioactive compounds. If the site of attachment is not known the radioisotope will be drawn in the structure box and linked to an asterisk.

x. Example: TOSITUMOMAB I-131



Description field

<PROTEIN>

```

<SEQUENCE_TYPE>COMPLETE</SEQUENCE_TYPE>
<SUBUNITS>4</SUBUNITS>
<SUBUNIT_GROUP>
  <SUBUNIT_ID>1</SUBUNIT_ID>
  <LENGTH>451</LENGTH>
  <SEQUENCE>QAYLQQSGAELVRPGASVKMSCKASGYTFTSYNMHWVKQTPRQ
GLEWIGAIYPNGDTSYNQKFKGKATLTVDKSSSTAYMQLSSLTSEDSAVYFCA
RVVYYNSYWFYFDVWGTGTTVTVSGPSVFPLAPSSKSTSGGTAALGCLVKDYFP
EPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVTVPSSSLGTQTYICNVNHKP
SNTKVDKKAEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCV
VVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWL
NGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLV
KGFYPSDIAVEWESNGQPENNYKTTTPVLDSGDFLYSKLTVDKSRWQQGNVF
SCSVMHEALHNHYTQKSLSLSPGK</SEQUENCE>
  <N_TERMINAL_MOD></N_TERMINAL_MOD>
  <C_TERMINAL_MOD></C_TERMINAL_MOD>
</SUBUNIT_GROUP>
<SUBUNIT_GROUP>
  <SUBUNIT_ID>2</SUBUNIT_ID>
  <LENGTH>451</LENGTH>
  <SEQUENCE>QAYLQQSGAELVRPGASVKMSCKASGYTFTSYNMHWVKQTPRQ
GLEWIGAIYPNGDTSYNQKFKGKATLTVDKSSSTAYMQLSSLTSEDSAVYFCA
RVVYYNSYWFYFDVWGTGTTVTVSGPSVFPLAPSSKSTSGGTAALGCLVKDYFP
EPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVTVPSSSLGTQTYICNVNHKP
SNTKVDKKAEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCV
VVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWL
NGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLV
KGFYPSDIAVEWESNGQPENNYKTTTPVLDSGDFLYSKLTVDKSRWQQGNVF
SCSVMHEALHNHYTQKSLSLSPGK</SEQUENCE>
  <N_TERMINAL_MOD></N_TERMINAL_MOD>
  <C_TERMINAL_MOD></C_TERMINAL_MOD>
</SUBUNIT_GROUP>
<SUBUNIT_GROUP>
  <SUBUNIT_ID>3</SUBUNIT_ID>
  <LENGTH>220</LENGTH>
  <SEQUENCE>QIVLSQSPAILSASPGEKVTMTCRASSVSYMHWYQQKPGSSPKP
WIYAPSNLASGVPARFSGSGSTSYSLTISRVEAEDAATYQCQWFSNPPTFGAG
TKLELKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQ
SGNSQESVTEQDSKSTYLSSTLTLSKADYEEKHKVYACEVTHQGLSSPVTKSFN
R</SEQUENCE>
  <N_TERMINAL_MOD></N_TERMINAL_MOD>
  <C_TERMINAL_MOD></C_TERMINAL_MOD>
</SUBUNIT_GROUP>
<SUBUNIT_GROUP>
  <SUBUNIT_ID>4</SUBUNIT_ID>
  <LENGTH>220</LENGTH>
  <SEQUENCE>QIVLSQSPAILSASPGEKVTMTCRASSVSYMHWYQQKPGSSPKP
WIYAPSNLASGVPARFSGSGSTSYSLTISRVEAEDAATYQCQWFSNPPTFGAG
TKLELKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQ
SGNSQESVTEQDSKSTYLSSTLTLSKADYEEKHKVYACEVTHQGLSSPVTKSFN
R</SEQUENCE>
  <N_TERMINAL_MOD></N_TERMINAL_MOD>
  <C_TERMINAL_MOD></C_TERMINAL_MOD>
</SUBUNIT_GROUP>
<MODIFICATION>IODINE I 131 LABELED</MODIFICATION>
<GLYCOSYLATED>MAMMALIAN</GLYCOSYLATED>

```

```

<PEGYLATED></PEGYLATED>
<PROTEIN_TYPE>MONOCLONAL ANTIBODY</PROTEIN_TYPE>
<PROTEIN_SUBTYPE>IGG2A</PROTEIN_SUBTYPE>
<SPECIES_TYPE>MOUSE</SPECIES_TYPE>
<LIGAND>931, MS4A1 MEMBRANE-SPANNING 4-DOMAINS, SUBFAMILY A,
MEMBER 1</LIGAND>
<GENE_ID></GENE_ID>
<COMMENTS></COMMENTS>
</PROTEIN>

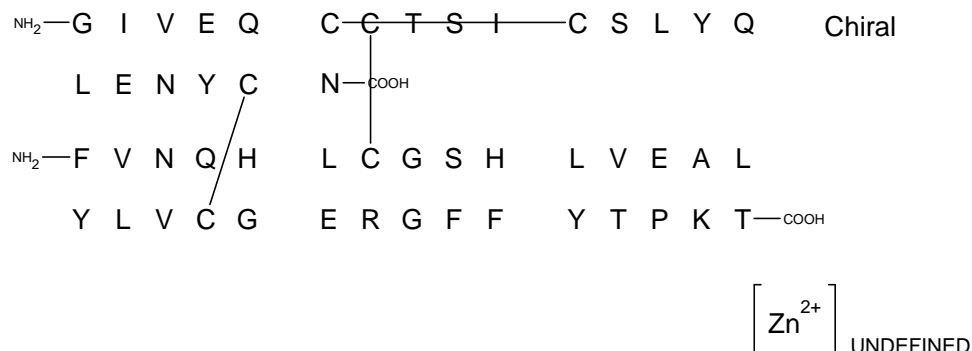
```

c. Salts of peptides and proteins:

Because of multiple sites of ionization on most peptides and proteins, peptides and proteins will not be drawn as ionized moieties. Counterions, unless the valency is well defined will be drawn within "GENERIC" brackets and UNDEFINED will be entered as stext associated with the bracket. There will only be one UNII associated with each salt of undefined or variable stoichiometry. Metal salts will be drawn as charged entities. Acid salts such as acetates, hydrochlorides and sulfates will be drawn as uncharged acids. Substances will not be distinguished by the extent of crystallinity, or release characteristics. Prompt Insulin Human Zinc and Extended Insulin Human Zinc will be given the same UNII as Insulin Human Zinc.

i. Example: INSULIN HUMAN ZINC

Structure Field



Description field

```

<PROTEIN>
<SEQUENCE_TYPE>COMPLETE</SEQUENCE_TYPE>
<SUBUNITS>2</SUBUNITS>
<SUBUNIT_GROUP>
  <SUBUNIT_ID>1</SUBUNIT_ID>
  <LENGTH>30</LENGTH>
  <SEQUENCE>FVNQHLVCGSHLVEALYLVCGERGFFYTPKT</SEQUENCE>
  <N_TERMINAL_MOD></N_TERMINAL_MOD>
  <C_TERMINAL_MOD></C_TERMINAL_MOD>
</SUBUNIT_GROUP>
<SUBUNIT_GROUP>

```

```

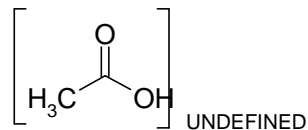
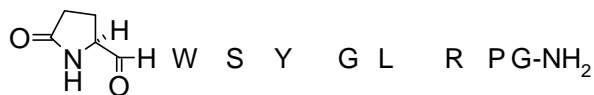
<SUBUNIT_ID>2</SUBUNIT_ID>
<LENGTH>21</LENGTH>
<SEQUENCE>GIVEQCCTSICSLYQLENYCN</SEQUENCE>
<N_TERMINAL_MOD></N_TERMINAL_MOD>
<C_TERMINAL_MOD></C_TERMINAL_MOD>
</SUBUNIT_GROUP>
<MODIFICATION></MODIFICATION>
<GLYCOSYLATED></GLYCOSYLATED>
<PEGYLATED></PEGYLATED>
<PROTEIN_TYPE>HORMONE</PROTEIN_TYPE>
<PROTEIN_SUBTYPE>INSULIN</PROTEIN_SUBTYPE>
<SPECIES_TYPE>HUMAN</SPECIES_TYPE>
<LIGAND>3643, INSULIN RECEPTOR</LIGAND>
<GENE_ID>3630, INSULIN</GENE_ID>
<COMMENTS></COMMENTS>
</PROTEIN>

```

ii. Example: GONADORELIN ACETATE

Structure Field

Chiral



Description field

```

<PROTEIN>
<SEQUENCE_TYPE>COMPLETE</SEQUENCE_TYPE>
<SUBUNITS>1</SUBUNITS>
<SUBUNIT_GROUP>
<SUBUNIT_ID>1</SUBUNIT_ID>
<LENGTH>10</LENGTH>
<SEQUENCE>XHWSYGLRPG</SEQUENCE>
<N_TERMINAL_MOD>PYROGLU</N_TERMINAL_MOD>
<C_TERMINAL_MOD>AMIDE</C_TERMINAL_MOD>
</SUBUNIT_GROUP>
<MODIFICATION>1_1=PYROGLU</MODIFICATION>
<GLYCOSYLATED></GLYCOSYLATED>
<PEGYLATED></PEGYLATED>
<PROTEIN_TYPE>HORMONE</PROTEIN_TYPE>
<PROTEIN_SUBTYPE></PROTEIN_SUBTYPE>
<SPECIES_TYPE></SPECIES_TYPE>

```

```
<LIGAND>2798, GNRHR GONADOTROPIN-RELEASING HORMONE RECEPTOR
</LIGAND>
<GENE_ID>2796, GNRH1 GONADOTROPIN-RELEASING HORMONE 1 (LUTEINIZING-
RELEASING HORMONE)</GENE_ID>
<COMMENTS></COMMENTS>
</PROTEIN>
```

27. Oligonucleotides (Proposed Not official)

Oligonucleotides with three or less nucleotides are drawn showing all atoms. Oligonucleotides up to 30 nucleotides long are drawn using a two letter code for the nucleoside and a one or two letter code for the phosphate linkage. The following abbreviations are used: dA = 2-deoxyadenosine, dC = deoxycytidine, dG = 2-deoxyguanosine, dT = thymidine, p = (3'-5')phosphoryl linkage, and ps = (3'-5')phosphorothioyl linkage.

The phosphate groups in the chain are negatively charged, so the appropriate cation is added in brackets next to the sequence.

Description Field.

Defined tags for nucleic acids.

<NUCLEIC_ACID></NUCLEIC_ACID> First tags in description field indicate that substance is a nucleic acid.

<SUBUNITS></SUBUNITS> Indicates the number of subunits or strands present in the substance.

Multiple strands can exist in a substance when two or more strands are based paired. Strands that are not based paired will be treated as separate substances.

<SUBUNIT_GROUP></SUBUNIT_GROUP> Indicates the beginning of the description for a given strand/subunit. Subunits are listed in order of decreasing length of nucleic acid sequence. If the lengths are the same then MW would be used. Each subunit is enclosed by subunit tags.

<SUBUNIT_ID></SUBUNIT_ID>Alphanumeric identifier of subunit group.

<LENGTH></LENGTH> Second tag between <SUBUNIT_GROUP> tags. Indicates the number of nucleic acid residues in a subunit. This tag is used for a quality control check.

<SEQUENCE></SEQUENCE> Third tag between <SUBUNIT_GROUP> tags. Contains the sequence abbreviated using the standard four letter DNA codes, I for inosine, and U for uracil. All non-standard bases are indicated by an X.

<5_MODIFICATION></5_MODIFICATION> Fourth tag between <SUBUNIT_GROUP> tags.

Describes any modification of the sugar at the 5' end of the strand. All strands are assumed to end in a 5' hydroxyl group. Examples: TRITYL, PHOSPHATE, and DIGOXIN.

<3_MODIFICATION></3_MODIFICATION> Fifth tag between <SUBUNIT_GROUP> tags. Describes any modification of the sugar at the 3' end of the strand. All strands are assumed to end in a 3' hydroxyl group. Examples: PHOSPHATE, DIGOXIN, and ACETATE.

<SUGAR></SUGAR> Sixth tag between <SUBUNIT_GROUP> tags. Describes modifications of the sugar throughout the strand. Sugars are assumed to be .BETA.-D-deoxyribose. Each type of modification is listed only once and the residues that have the modification are then listed starting from the 5' end; gaps are separated by commas. Modifications are listed in order of first appearance. Example: BETA.-D-RIBOSE= 1-5,8,10-16 means residues 1-5, 8 and 10-16 contain a ribose sugar.

<LINKAGE></LINKAGE> Seventh tag between <SUBUNIT_GROUP> tags. Describes modifications to the phosphate groups. Each linkage is associated with its 3' nucleoside. Each type of modified linkage is listed only once starting from the 5' end of the strand. Example: PHOSPHOTHIORATE= 1-6 means the linkages connecting residues 1-7 have a phosphothiorate linkage. For substances that contain phosphothiorate, methylphosphonate and other linkages that contain a chiral center, a single structure is drawn. Both <STEREOCHEMISTRY> and <NUCLEIC_ACID> tags are used to describe these substances. Normal bonds are used where stereochemistry is not specified and the stereochemistry <TYPE> tags contain the value "UNKNOWN".

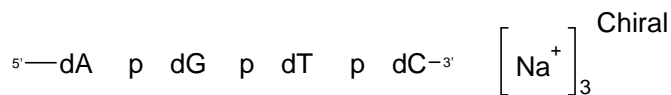
<BASE_MODIFICATIONS></BASE_MODIFICATIONS> Eighth tag between <SUBUNIT_GROUP> tags. Describes modifications to the bases—represented by X in the <SEQUENCE> tags. Each modification is listed only once. Examples: 2,6-DIAMINOPURINE= 4 this means that residue 4 contains the base 2,6-diaminopurine

<COMMENTS></COMMENTS> Comments used to distinguish one nucleic acid from another when the tags above are not sufficient.

Oligonucleotides with more than 30 residues are defined ONLY by descriptions.

i. Example: DNA sequence d(AGTC)

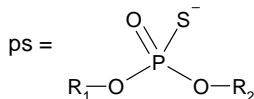
Structure Field



Description Field

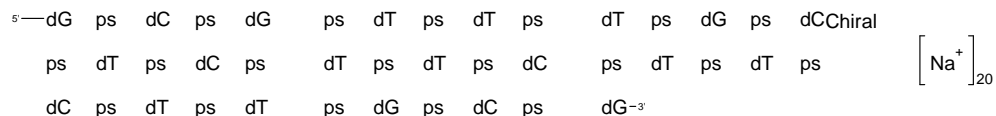
<NUCLEIC_ACID><SUBUNITS>1</SUBUNITS><SUBUNIT_GROUP><SUBUNIT_ID>1</SUBUNIT_ID><LENGTH>4</LENGTH><SEQUENCE>AGTC</SEQUENCE><5_MODIFICATION></5_MODIFICATION><3_MODIFICATION></3_MODIFICATION><SUGAR></SUGAR><LINKAGE></LINKAGE><BASE_MODIFICATIONS></BASE_MODIFICATIONS><SUBUNIT_GROUP><COMMENTS></COMMENTS></NUCLEIC_ACID>

Phosphorothioate oligonucleotides are usually a mixture of many diastereomers since the configuration of the phosphorothioate groups is not controlled during synthesis. No wedge bonds are drawn to phosphorothioate groups:



ii. Example: FOMIVIRSEN SODIUM

Structure Field



Description Field

```
<NUCLEIC_ACID><SUBUNITS>1</SUBUNITS><SUBUNIT_GROUP><SUBUNIT_ID>1</SUBUNIT_ID><LENGTH>21</LENGTH><SEQUENCE>GCGTTTGCTCTTCTTGC</SEQUENCE><5_MODIFICATION></5_MODIFICATION><3_MODIFICATION></3_MODIFICATION><SUGAR></SUGAR><LINKAGE> PHOSPHOTHIORATE=1-20</LINKAGE><BASE_MODIFICATIONS></BASE_MODIFICATIONS></SUBUNIT_GROUP><COMMENTS></COMMENTS></NUCLEIC_ACID>
```

28. Polymers (Proposed Not Official)

All polymers that contain known repeating units and/or substituents are defined using both structure and description fields. For each polymer as much detail as is available is added to the description field. Polymers that differ by molecular weight (MW) or distribution of molecular weights are considered different polymers.

a. Structure Drawing

The structure field contains the smallest repeating unit that results from the polymerization of the monomer. MDL's structure repeating unit (SRU) brackets, that are labeled with a single capital letter identifier (A, B, C, etc), are used for homopolymers and copolymers. Tags are used in the description field to indicate the orientation of polymerization for both homopolymers and copolymers.

End groups, if known, are drawn and connected to the ends of a polymer. End groups on copolymers that are not definitely connected to a given polymerized monomer are drawn and enclosed in generic brackets labeled END_GROUP using S-Text.

Cross-linking SRUs in copolymers are treated like other repeating units. When substitution on a polymer exists at multiple sites and is incomplete, the substituent is enclosed in generic brackets labeled R# using S-Text. R-groups are numbered in order of decreasing MW. The repeating unit contains an asterisk at the position of potential substitution. The substituent also contains an asterisk at the site of attachment (See examples #v, and vi).

b. Description Field

The description field contains the following tags:

<POLYMER></POLYMER> Indicates that the substance is a polymer.

<POLYMER_TYPE></POLYMER_TYPE> Indicates the type of polymer or copolymer.

Typical entries include:

HOMOPOLYMER: A polymer that has a single structural repeating unit. This term is used when tacticity does not exist.

HOMOPOLYMER UNKNOWN: A polymer that has a single structural repeating unit. This term is used when tacticity is possible, but the type of tacticity is unknown.

HOMOPOLYMER ISOTACTIC, HOMOPOLYMER SYNTACTIC, HOMOPOLYMER ATACTIC:
 A polymer that has a single structural repeating unit where tacticity is known.

COPOLYMER RANDOM: A polymer represented by multiple structural repeating units where the order of the repeating units is not controlled.

COPOLYMER BLOCK: A polymer represented by multiple structural repeating units where blocks of identical repeating units exist.

COPOLYMER GRAFT: A polymer represented by multiple structural repeating units where one polymer is attached to the main chain of another polymer.

COPOLYMER UNKNOWN: A polymer represented by multiple structural repeating units where no information about connectivity between structure repeating units is known.

<NUMBER_OF_SRU></NUMBER_OF_SRU> Indicates the number of different structural repeating units in a polymer. For a homopolymer the value is one.

<SRU_GROUP></SRU_GROUP> Tags surrounding information about a given structural repeating unit of a polymer.

<SRU_ID></SRU_ID>. The value should be identical to the bracket identifier used in the structure field. Repeating units are labeled in alphabetical order starting from the highest MW structural repeating unit. The bracket identifier is a single capital letter associated with a bracket. Attached R-Groups are not initially considered in the ordering of the subunits. If two subunits have identical MW then the MW of attached R-Groups is taken into consideration.

<ORIENTATION_OF_POLYMERIZATION></ORIENTATION_OF_POLYMERIZATION> Tags used to describe the orientation of homopolymers and block copolymers. Examples: HEAD-TAIL, HEAD-HEAD, RANDOM, UNKNOWN

<SRU_PERCENT></SRU_PERCENT> Tags that surround the percentage tags of a given repeating unit of a polymer.

<AVERAGE></AVERAGE> The average percentage of a given repeating unit that defines a polymeric substance.

<LOW_LIMIT></LOW_LIMIT> The lowest percentage of a given repeating unit that defines a polymeric substance.

<HIGH_LIMIT></HIGH_LIMIT> The highest percentage of a given repeating unit that defines a polymeric substance.

<R_GROUP></R_GROUP> Tags that surround information about substituents added to a given SRU.

<R_ID></R_ID> The value should be identical to the bracket identifier used in the structure field. R-Groups are numbered using R# (R1,R2...). R1 has the largest MW, R2 the second largest MW, and so on.

<LIMIT_TYPE></LIMIT_TYPE> Type of limit used to describe the substitution. Examples: WEIGHT PERCENT, DEGREE OF SUBSTITUTION.

<AVERAGE></AVERAGE> The average degree of substitution.

<LOW_LIMIT></LOW_LIMIT> The lower limit of substitution for the R-Group.

<HIGH_LIMIT></HIGH_LIMIT> The highest limit of substitution for the R-Group.

<DEGREE_POLYMERIZATION></DEGREE_POLYMERIZATION> Degree of polymerization is defined as the average number of monomers which make up a given polymer.

<AVERAGE></AVERAGE> The average degree of polymerization for the defined polymer.

<LOW_LIMIT></LOW_LIMIT> The lower limit for the degree of polymerization.

<HIGH_LIMIT></HIGH_LIMIT> The higher limit for the degree of polymerization.

<MW_GROUP></MW_GROUP> Tags that provide information about the molecular weight of the polymer. Polydispersity is indicated by the ratio of the weight average to number average. Both values are added if known. Tags that surround information on the polymers MW.

<TYPE_MW></TYPE_MW> The type of MW reported for the polymer. Examples: NUMBER AVERAGE, WEIGHT AVERAGE, VISCOSITY AVERAGE.

<MW_AVERAGE></MW_AVERAGE> The average MW of the polymer. Polymers are considered different if their average molecular weights differ by a significant amount.

<LOW_LIMIT_MW></LOW_LIMIT_MW > The lowest limit of the average molecular weight that describes a polymer.

<HIGH_LIMIT_MW_AVERAGE></HIGH_LIMIT_MW_AVERAGE> The highest limit of the average molecular weight that describes a polymer.

<PHYSICAL_PROPERTIES_GROUP></PHYSICAL_PROPERTIES_GROUP> Tags that provide information about the distinguishing physical properties of a polymer.

<PHYSICAL_PROPERTY_TYPE></PHYSICAL_PROPERTY_TYPE> Distinguishing physical property. Examples: DENSITY, CRYSTALLINITY, VISCOSITY, and PERMEABILITY
 [Note-Care must be taken when using physical properties to define a polymer. The addition of a new physical property does not necessarily lead to a new substance. Physical properties should only be used when they distinguish one polymer from another.]

A second physical property can be added with a second set of tags. The order should be alphabetical based on the distinguishing physical property.

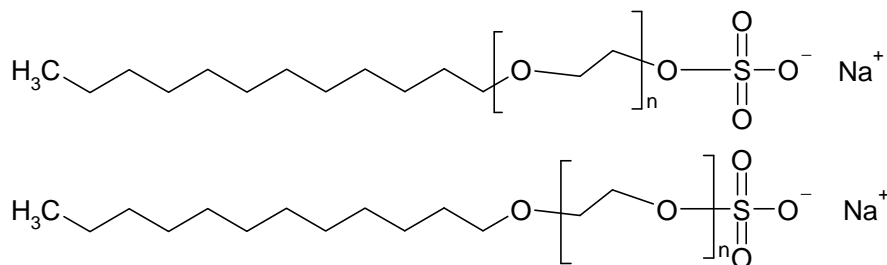
<AVERAGE></AVERAGE> The average value for a physical property.
 <LOW_LIMIT></LOW_LIMIT> The lower limit for a physical property.
 <HIGH_LIMIT></HIGH_LIMIT> The upper limit for a physical property.
 <UNIT></UNIT> Unit of measurement for a physical property. EXAMPLE: Density = G/ML

<COMMENTS></COMMENTS> General comments about a polymer that are important to the definition of a polymer. Comments should be used sparingly; only information essential to the definition of a polymer should be included.

In the description field tags, high and low limit tags either contain product specification limits or limits at the 95% confidence interval if a standard deviation is given (i.e., +/- 2 times the standard deviation).
 Example: MW of a given polymer is 5000 +/- 500. The <MW_AVERAGE> tags contain the value 5000, the <LOW_LIMIT_MW_AVERAGE> contains the value 4000, the <HIGH_LIMIT_MW_AVERAGE> contains the value 6000.

Particle size, percent water, and percent solvent are not distinguishing factors for polymers.

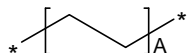
Care must be taken to ensure that two representations of the same polymer are not entered into the SRS. The SRS software allows registration of the two polymers below:



A SRS register must use substructure searches prior to entering any polymer that has two or more valid representations.

i. Example. Low Density POLYETHYLENE

Structure Field

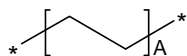


Description Field

```
<POLYMER>
  <POLYMER_TYPE>HOMOPOLYMER</POLYMER_TYPE>
  <NUMBER_OF_SRU>1</NUMBER_OF_SRU>
  <SRU_GROUP>
    <SRU_ID>A</SRU_ID>
    <ORIENTATION_OF_POLYMERIZATION>
      </ORIENTATION_OF_POLYMERIZATION>
    <SRU_PERCENT>
      <AVERAGE></AVERAGE>
      <LOW_LIMIT></LOW_LIMIT>
      <HIGH_LIMIT></HIGH_LIMIT>
    </SRU_PERCENT>
    <R_GROUP>
      <R_ID></R_ID>
      <LIMIT_TYPE></LIMIT_TYPE>
      <AVERAGE></AVERAGE>
      <LOW_LIMIT></LOW_LIMIT>
      <HIGH_LIMIT></HIGH_LIMIT>
    </R_GROUP>
    <DEGREE_POLYMERIZATION>
      <AVERAGE></AVERAGE>
      <LOW_LIMIT></LOW_LIMIT>
      <HIGH_LIMIT></HIGH_LIMIT>
    </DEGREE_POLYMERIZATION>
  </SRU_GROUP>
  <MW_GROUP>
    <TYPE_MW></TYPE_MW>
    <MW_AVERAGE></MW_AVERAGE>
    <LOW_LIMIT_MW></LOW_LIMIT_MW>
    <HIGH_LIMIT_MW></HIGH_LIMIT_MW>
  </MW_GROUP>
  <PHYSICAL_PROPERTIES_GROUP>
    <PHYSICAL_PROPERTY_TYPE>DENSITY</PHYSICAL_PROPERTY_TYPE>
    <AVERAGE></AVERAGE>
    <LOW_LIMIT>0.91</LOW_LIMIT>
    <HIGH_LIMIT>0.93</HIGH_LIMIT>
    <UNITS>G/ML</UNITS>
  </PHYSICAL_PROPERTIES_GROUP>
  <COMMENTS></COMMENTS>
</POLYMER>
```

ii. Example. High Density POLYETHYLENE

Structure Field



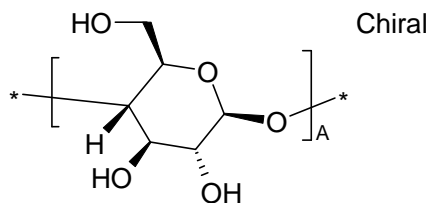
Description Field

```
<POLYMER>
  <POLYMER_TYPE>HOMOPOLYMER</POLYMER_TYPE>
  <NUMBER_OF_SRU>1</NUMBER_OF_SRU>
  <SRU_GROUP>
    <SRU_ID>A</SRU_ID>
    <ORIENTATION_OF_POLYMERIZATION>
      </ORIENTATION_OF_POLYMERIZATION>
    <SRU_PERCENT>
      <AVERAGE></AVERAGE>
      <LOW_LIMIT></LOW_LIMIT>
      <HIGH_LIMIT></HIGH_LIMIT>
    </SRU_PERCENT>
    <R_GROUP>
      <R_ID></R_ID>
      <LIMIT_TYPE></LIMIT_TYPE>
      <AVERAGE></AVERAGE>
      <LOW_LIMIT></LOW_LIMIT>
      <HIGH_LIMIT></HIGH_LIMIT>
    </R_GROUP>
    <DEGREE_POLYMERIZATION>
      <AVERAGE></AVERAGE>
      <LOW_LIMIT></LOW_LIMIT>
      <HIGH_LIMIT></HIGH_LIMIT>
    </DEGREE_POLYMERIZATION>
  </SRU_GROUP>
  <MW_GROUP>
    <TYPE_MW></TYPE_MW>
    <MW_AVERAGE></MW_AVERAGE>
    <LOW_LIMIT_MW></LOW_LIMIT_MW>
    <HIGH_LIMIT_MW></HIGH_LIMIT_MW>
  </MW_GROUP>
  <PHYSICAL_PROPERTIES_GROUP>
    <PHYSICAL_PROPERTY_TYPE>DENSITY</PHYSICAL_PROPERTY_TYPE>
    <AVERAGE></AVERAGE>
    <LOW_LIMIT>.94</LOW_LIMIT>
    <HIGH_LIMIT>.97</HIGH_LIMIT>
    <UNITS>G/CC</UNITS>
  </PHYSICAL_PROPERTIES_GROUP>
  <COMMENTS></COMMENTS>
</POLYMER>
```

iii. Example. MICROCRYSTALLINE CELLULOSE

Degree of polymerization is less than 350.

Structure Field



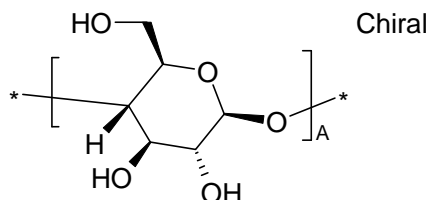
Description Field

```
<POLYMER>
  <POLYMER_TYPE>HOMOPOLYMER</POLYMER_TYPE>
  <NUMBER_OF_SRU>1</NUMBER_OF_SRU>
  <SRU_GROUP>
    <SRU_ID>A</SRU_ID>
    <ORIENTATION_OF_POLYMERIZATION>HEAD-TAIL
      </ORIENTATION_OF_POLYMERIZATION>
    <SRU_PERCENT>
      <AVERAGE></AVERAGE>
      <LOW_LIMIT></LOW_LIMIT>
      <HIGH_LIMIT></HIGH_LIMIT>
    </SRU_PERCENT>
    <R_GROUP>
      <R_ID></R_ID>
      <LIMIT_TYPE></LIMIT_TYPE>
      <AVERAGE></AVERAGE>
      <LOW_LIMIT></LOW_LIMIT>
      <HIGH_LIMIT></HIGH_LIMIT>
    </R_GROUP>
    <DEGREE_POLYMERIZATION>
      <AVERAGE></AVERAGE>
      <LOW_LIMIT></LOW_LIMIT>
      <HIGH_LIMIT>350</HIGH_LIMIT>
    </DEGREE_POLYMERIZATION>
  </SRU_GROUP>
  <MW_GROUP>
    <TYPE_MW></TYPE_MW>
    <MW_AVERAGE></MW_AVERAGE>
    <LOW_LIMIT_MW></LOW_LIMIT_MW>
    <HIGH_LIMIT_MW></HIGH_LIMIT_MW>
  </MW_GROUP>
  <PHYSICAL_PROPERTIES_GROUP>
    <PHYSICAL_PROPERTY_TYPE></PHYSICAL_PROPERTY_TYPE>
    <AVERAGE></AVERAGE>
    <LOW_LIMIT></LOW_LIMIT>
    <HIGH_LIMIT></HIGH_LIMIT>
    <UNITS></UNITS>
  </PHYSICAL_PROPERTIES_GROUP>
  <COMMENTS></COMMENTS>
</POLYMER>
```

iv. Example. POWDERED CELLULOSE

Degree of polymerization is less than 2250 and more than 440.

Structure Field



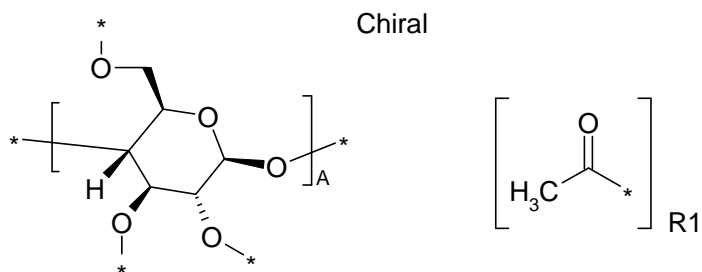
Description Field

```
<POLYMER>
  <POLYMER_TYPE>HOMOPOLYMER</POLYMER_TYPE>
  <NUMBER_OF_SRU>1</NUMBER_OF_SRU>
  <SRU_GROUP>
    <SRU_ID>A</SRU_ID>
    <ORIENTATION_OF_POLYMERIZATION>HEAD-TAIL
    </ORIENTATION_OF_POLYMERIZATION>
    <SRU_PERCENT>
      <AVERAGE></AVERAGE>
      <LOW_LIMIT></LOW_LIMIT>
      <HIGH_LIMIT></HIGH_LIMIT>
    </SRU_PERCENT>
    <R_GROUP>
      <R_ID></R_ID>
      <LIMIT_TYPE></LIMIT_TYPE>
      <AVERAGE></AVERAGE>
      <LOW_LIMIT></LOW_LIMIT>
      <HIGH_LIMIT></HIGH_LIMIT>
    </R_GROUP>
    <DEGREE_POLYMERIZATION>
      <AVERAGE></AVERAGE>
      <LOW_LIMIT>440</LOW_LIMIT>
      <HIGH_LIMIT>2250</HIGH_LIMIT>
    </DEGREE_POLYMERIZATION>
  </SRU_GROUP>
  <MW_GROUP>
    <TYPE_MW></TYPE_MW>
    <MW_AVERAGE></MW_AVERAGE>
    <LOW_LIMIT_MW></LOW_LIMIT_MW>
    <HIGH_LIMIT_MW></HIGH_LIMIT_MW>
  </MW_GROUP>
  <PHYSICAL_PROPERTIES_GROUP>
    <PHYSICAL_PROPERTY_TYPE></PHYSICAL_PROPERTY_TYPE>
    <AVERAGE></AVERAGE>
    <LOW_LIMIT></LOW_LIMIT>
    <HIGH_LIMIT></HIGH_LIMIT>
    <UNITS></UNITS>
  </PHYSICAL_PROPERTIES_GROUP>
  <COMMENTS></COMMENTS>
</POLYMER>
```

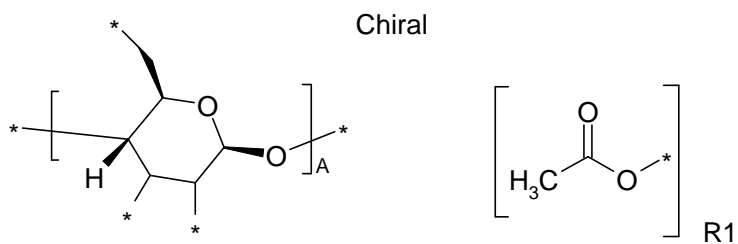
v. Example. CELLULOSE ACETATE

Cellulose acetate is defined using the USP criteria. The weight percent of acetate is not less than 29.0 percent and not greater than 44.8. Positions of potential substitution are indicated by an asterisk for both the polymer and the substituent. Hydrogen atoms are not listed but assumed to occupy positions that are not completely substituted. Polymers are drawn to retain as many of the atoms of the polymer as possible.

Structure Field



Correct



Incorrect

Description Field

```
<POLYMER>
<POLYMER_TYPE>HOMOPOLYMER</POLYMER_TYPE>
<NUMBER_OF_SRU>1</NUMBER_OF_SRU>
<SRU_GROUP>
<SRU_ID>A</SRU_ID>
```



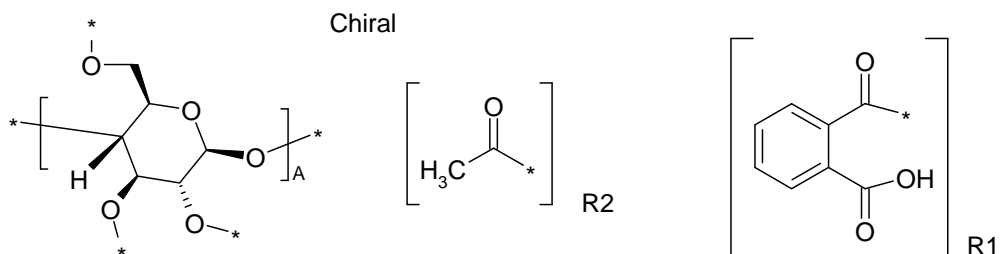
```

<ORIENTATION_OF_POLYMERIZATION>HEAD-TAIL
  </ORIENTATION_OF_POLYMERIZATION>
<SRU_PERCENT>
  <AVERAGE></AVERAGE>
  <LOW_LIMIT></LOW_LIMIT>
  <HIGH_LIMIT></HIGH_LIMIT>
</SRU_PERCENT>
<R_GROUP>
  <R_ID>R1</R_ID>
  <LIMIT_TYPE>WEIGHT PERCENT</LIMIT_TYPE>
  <AVERAGE></AVERAGE>
  <LOW_LIMIT>29.0</LOW_LIMIT>
  <HIGH_LIMIT>44.8</HIGH_LIMIT>
</R_GROUP>
<DEGREE_POLYMERIZATION>
  <AVERAGE></AVERAGE>
  <LOW_LIMIT></LOW_LIMIT>
  <HIGH_LIMIT></HIGH_LIMIT>
</DEGREE_POLYMERIZATION>
</SRU_GROUP>
<MW_GROUP>
  <TYPE_MW></TYPE_MW>
  <MW_AVERAGE></MW_AVERAGE>
  <LOW_LIMIT_MW></LOW_LIMIT_MW>
  <HIGH_LIMIT_MW></HIGH_LIMIT_MW>
</MW_GROUP>
<PHYSICAL_PROPERTIES_GROUP>
  <PHYSICAL_PROPERTY_TYPE></PHYSICAL_PROPERTY_TYPE>
  <AVERAGE></AVERAGE>
  <LOW_LIMIT></LOW_LIMIT>
  <HIGH_LIMIT></HIGH_LIMIT>
  <UNITS></UNITS>
</PHYSICAL_PROPERTIES_GROUP>
<COMMENTS></COMMENTS>
</POLYMER>

```

vi. Example. CELLACEFATE

Structure Field



Description Field

```

<POLYMER>
<POLYMER_TYPE>HOMOPOLYMER</POLYMER_TYPE>
<NUMBER_OF_SRU>1</NUMBER_OF_SRU>
<SRU_GROUP>
  <SRU_ID>A</SRU_ID>
  <ORIENTATION_OF_POLYMERIZATION>HEAD-TAIL
  </ORIENTATION_OF_POLYMERIZATION>
  <SRU_PERCENT>
    <AVERAGE></AVERAGE>
    <LOW_LIMIT></LOW_LIMIT>
    <HIGH_LIMIT></HIGH_LIMIT>
  </SRU_PERCENT>
  <R_GROUP>
    <R_ID>R1</R_ID>
    <LIMIT_TYPE>WEIGHT PERCENT</LIMIT_TYPE>
    <AVERAGE></AVERAGE>
    <LOW_LIMIT>30.0</LOW_LIMIT>
    <HIGH_LIMIT>36.0</HIGH_LIMIT>
  </R_GROUP>
  <R_GROUP>
    <R_ID>R2</R_ID>
    <LIMIT_TYPE>WEIGHT PERCENT</LIMIT_TYPE>
    <AVERAGE></AVERAGE>
    <LOW_LIMIT>21.5</LOW_LIMIT>
    <HIGH_LIMIT>26.0</HIGH_LIMIT>
  </R_GROUP>
  <DEGREE_POLYMERIZATION>
    <AVERAGE></AVERAGE>
    <LOW_LIMIT></LOW_LIMIT>
    <HIGH_LIMIT></HIGH_LIMIT>
  </DEGREE_POLYMERIZATION>
</SRU_GROUP>
<MW_GROUP>
  <TYPE_MW></TYPE_MW>
  <MW_AVERAGE></MW_AVERAGE>
  <LOW_LIMIT_MW></LOW_LIMIT_MW>
  <HIGH_LIMIT_MW></HIGH_LIMIT_MW>
</MW_GROUP>
<PHYSICAL_PROPERTIES_GROUP>

```

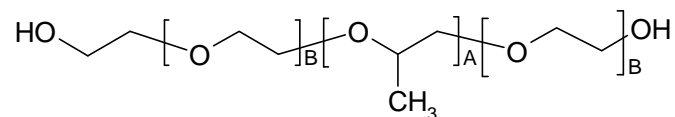
```

    <PHYSICAL_PROPERTY_TYPE></PHYSICAL_PROPERTY_TYPE>
    <AVERAGE></AVERAGE>
    <LOW_LIMIT></LOW_LIMIT>
    <HIGH_LIMIT></HIGH_LIMIT>
    <UNITS></UNITS>
  </PHYSICAL_PROPERTIES_GROUP>
  <COMMENTS></COMMENTS>
</POLYMER>

```

vii. Example. POLOXAMER 124

Structure Field



Description Field

```

<POLYMER>
  <POLYMER_TYPE>COPOLYMER, BLOCK</POLYMER_TYPE>
  <NUMBER_OF_SRU>2</NUMBER_OF_SRU>
  <SRU_GROUP>
    <SRU_ID>A</SRU_ID>
    <ORIENTATION_OF_POLYMERIZATION>HEAD-TAIL
      </ORIENTATION_OF_POLYMERIZATION>
    <SRU_PERCENT>
      <AVERAGE></AVERAGE>
      <LOW_LIMIT>51.4</LOW_LIMIT>
      <HIGH_LIMIT>55.2</HIGH_LIMIT>
    </SRU_PERCENT>
    <R_GROUP>
      <R_ID></R_ID>
      <LIMIT_TYPE></LIMIT_TYPE>
      <AVERAGE></AVERAGE>
      <LOW_LIMIT></LOW_LIMIT>
      <HIGH_LIMIT></HIGH_LIMIT>
    </R_GROUP>
    <DEGREE_POLYMERIZATION>
      <AVERAGE>20</AVERAGE>
      <LOW_LIMIT></LOW_LIMIT>
      <HIGH_LIMIT></HIGH_LIMIT>
    </DEGREE_POLYMERIZATION>
  </SRU_GROUP>
  <SRU_GROUP>
    <SRU_ID>B</SRU_ID>
    <ORIENTATION_OF_POLYMERIZATION>HEAD-TAIL
      </ORIENTATION_OF_POLYMERIZATION>
    <SRU_PERCENT>
      <AVERAGE></AVERAGE>
      <LOW_LIMIT>44.8</LOW_LIMIT>

```

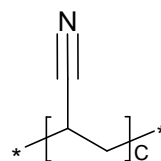
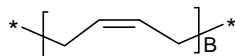
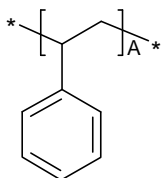
```

      <HIGH_LIMIT>48.6</HIGH_LIMIT>
    </SRU_PERCENT>
  <R_GROUP>
    <R_ID></R_ID>
    <LIMIT_TYPE></LIMIT_TYPE>
    <AVERAGE></AVERAGE>
    <LOW_LIMIT></LOW_LIMIT>
    <HIGH_LIMIT></HIGH_LIMIT>
  </R_GROUP>
<DEGREE_POLYMERIZATION>
  <AVERAGE>12</AVERAGE>
  <LOW_LIMIT></LOW_LIMIT>
  <HIGH_LIMIT></HIGH_LIMIT>
</DEGREE_POLYMERIZATION>
</SRU_GROUP>
<MW_GROUP>
  <TYPE_MW>NUMBER AVERAGE</TYPE_MW>
  <MW_AVERAGE></MW_AVERAGE>
  <LOW_LIMIT_MW>2090</LOW_LIMIT_MW>
  <HIGH_LIMIT_MW>2360</HIGH_LIMIT_MW>
</MW_GROUP>
<PHYSICAL_PROPERTIES_GROUP>
  <PHYSICAL_PROPERTY_TYPE></PHYSICAL_PROPERTY_TYPE>
  <AVERAGE></AVERAGE>
  <LOW_LIMIT></LOW_LIMIT>
  <HIGH_LIMIT></HIGH_LIMIT>
  <UNITS></UNITS>
</PHYSICAL_PROPERTIES_GROUP>
<COMMENTS></COMMENTS>
</POLYMER>

```

viii. Example. ACRYLONITRILE- BUTADIENE-STYRENE TERPOLYMER

Structure Field



Description Field

```

<POLYMER>
  <POLYMER_TYPE>COPOLYMER, GRAFT</POLYMER_TYPE>
  <NUMBER_OF_SRU>3</NUMBER_OF_SRU>
  <SRU_GROUP>
    <SRU_ID>A</SRU_ID>
    <ORIENTATION_OF_POLYMERIZATION>

```

```

    </ORIENTATION_OF_POLYMERIZATION>
  <SRU_PERCENT>
    <AVERAGE></AVERAGE>
    <LOW_LIMIT>40</LOW_LIMIT>
    <HIGH_LIMIT>60</HIGH_LIMIT>
  </SRU_PERCENT>
  <R_GROUP>
    <R_ID></R_ID>
    <LIMIT_TYPE></LIMIT_TYPE>
    <AVERAGE></AVERAGE>
    <LOW_LIMIT></LOW_LIMIT>
    <HIGH_LIMIT></HIGH_LIMIT>
  </R_GROUP>
  <DEGREE_POLYMERIZATION>
    <AVERAGE></AVERAGE>
    <LOW_LIMIT></LOW_LIMIT>
    <HIGH_LIMIT></HIGH_LIMIT>
  </DEGREE_POLYMERIZATION>
</SRU_GROUP>
<SRU_GROUP>
  <SRU_ID>B</SRU_ID>
  <ORIENTATION_OF_POLYMERIZATION>
    </ORIENTATION_OF_POLYMERIZATION>
  <SRU_PERCENT>
    <AVERAGE></AVERAGE>
    <LOW_LIMIT>5</LOW_LIMIT>
    <HIGH_LIMIT>30</HIGH_LIMIT>
  </SRU_PERCENT>
  <R_GROUP>
    <R_ID></R_ID>
    <LIMIT_TYPE></LIMIT_TYPE>
    <AVERAGE></AVERAGE>
    <LOW_LIMIT></LOW_LIMIT>
    <HIGH_LIMIT></HIGH_LIMIT>
  </R_GROUP>
  <DEGREE_POLYMERIZATION>
    <AVERAGE></AVERAGE>
    <LOW_LIMIT></LOW_LIMIT>
    <HIGH_LIMIT></HIGH_LIMIT>
  </DEGREE_POLYMERIZATION>
</SRU_GROUP>
<SRU_GROUP>
  <SRU_ID>C</SRU_ID>
  <ORIENTATION_OF_POLYMERIZATION>
    </ORIENTATION_OF_POLYMERIZATION>
  <SRU_PERCENT>
    <AVERAGE></AVERAGE>
    <LOW_LIMIT>15</LOW_LIMIT>
    <HIGH_LIMIT>30</HIGH_LIMIT>
  </SRU_PERCENT>
  <R_GROUP>
    <R_ID></R_ID>
    <LIMIT_TYPE></LIMIT_TYPE>
    <AVERAGE></AVERAGE>
    <LOW_LIMIT></LOW_LIMIT>
    <HIGH_LIMIT></HIGH_LIMIT>

```

```

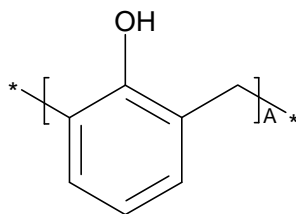
    </R_GROUP>
    <DEGREE_POLYMERIZATION>
      <AVERAGE></AVERAGE>
      <LOW_LIMIT></LOW_LIMIT>
      <HIGH_LIMIT></HIGH_LIMIT>
    </DEGREE_POLYMERIZATION>
  </SRU_GROUP>
  <MW_GROUP>
    <TYPE_MW></TYPE_MW>
    <MW_AVERAGE></MW_AVERAGE>
    <LOW_LIMIT_MW></LOW_LIMIT_MW>
    <HIGH_LIMIT_MW></HIGH_LIMIT_MW>
  </MW_GROUP>
  <PHYSICAL_PROPERTIES_GROUP>
    <PHYSICAL_PROPERTY_TYPE></PHYSICAL_PROPERTY_TYPE>
    <AVERAGE></AVERAGE>
    <LOW_LIMIT></LOW_LIMIT>
    <HIGH_LIMIT></HIGH_LIMIT>
    <UNITS></UNITS>
  </PHYSICAL_PROPERTIES_GROUP>
  <COMMENTS>STYRENE AND ACRYLONITRILE ARE GRAFTED ONTO A BUTADIENE
  BACKBONE FREE STYRENE-ACRYLONITRILE COPOLYMER CHAINS ALSO
  PRESENT</COMMENTS>
</POLYMER>

```

ix. Example: PHENOL FORMALDEHYDE RESINS (Condensation polymers)

Phenol formaldehyde resins are usually three-dimensional network polymers. Only one structure is drawn for each condensation polymer. The polymer structure is simplified with only two positions of substitution indicated. Substitution of the methylene group will either be ortho to the primary functional group or at the position nearest to the primary functional group.

Structure Field



Description Field

```

<POLYMER>
  <POLYMER_TYPE>CONDENSATION, NETWORK</POLYMER_TYPE>

```

```

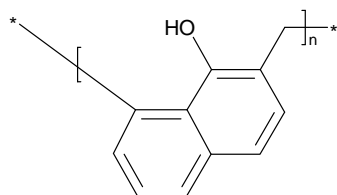
<NUMBER_OF_SRU>1</NUMBER_OF_SRU>
<SRU_GROUP>
  <SRU_ID>A</SRU_ID>
  <ORIENTATION_OF_POLYMERIZATION>
  </ORIENTATION_OF_POLYMERIZATION>
  <SRU_PERCENT>
    <AVERAGE></AVERAGE>
    <LOW_LIMIT></LOW_LIMIT>
    <HIGH_LIMIT></HIGH_LIMIT>
  </SRU_PERCENT>
  <R_GROUP>
    <R_ID></R_ID>
    <LIMIT_TYPE></LIMIT_TYPE>
    <AVERAGE></AVERAGE>
    <LOW_LIMIT></LOW_LIMIT>
    <HIGH_LIMIT></HIGH_LIMIT>
  </R_GROUP>
  <DEGREE_POLYMERIZATION>
    <AVERAGE></AVERAGE>
    <LOW_LIMIT></LOW_LIMIT>
    <HIGH_LIMIT></HIGH_LIMIT>
  </DEGREE_POLYMERIZATION>
</SRU_GROUP>
<MW_GROUP>
  <TYPE_MW></TYPE_MW>
  <MW_AVERAGE></MW_AVERAGE>
  <LOW_LIMIT_MW></LOW_LIMIT_MW>
  <HIGH_LIMIT_MW></HIGH_LIMIT_MW>
</MW_GROUP>
<PHYSICAL_PROPERTIES_GROUP>
  <PHYSICAL_PROPERTY_TYPE></PHYSICAL_PROPERTY_TYPE>
  <AVERAGE></AVERAGE>
  <LOW_LIMIT></LOW_LIMIT>
  <HIGH_LIMIT></HIGH_LIMIT>
  <UNITS></UNITS>
</PHYSICAL_PROPERTIES_GROUP>
  <COMMENTS>FORMS A THREE DIMENSIONAL NETWORK, EXTENT OF NETWORK
  DEPENDS ON RATIO OF FORMALDEHYDE TO PHENOL; PHENOL PARA, META-
  POSITIONS CAN ALSO UNDERGO SUBSTITUTION </COMMENTS>
</POLYMER>

```

In cases where symmetrical substitution is not possible the methylene group is attached to the ortho position or to the nearest position of substitution. The polymer attachment site would be at the nearest site on the opposite side of the functional group.

x. Example: NAPTHOL FORMALDEHYDE RESIN

Structure Field



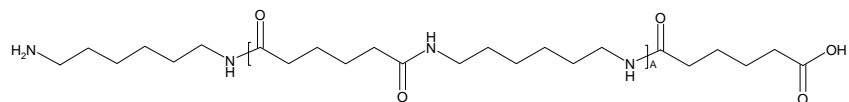
Description Field

```
<POLYMER>
  <POLYMER_TYPE>CONDENSATION, NETWORK</POLYMER_TYPE>
  <NUMBER_OF_SRU>1</NUMBER_OF_SRU>
  <SRU_GROUP>
    <SRU_ID>A</SRU_ID>
    <ORIENTATION_OF_POLYMERIZATION>
    </ORIENTATION_OF_POLYMERIZATION>
    <SRU_PERCENT>
      <AVERAGE></AVERAGE>
      <LOW_LIMIT></LOW_LIMIT>
      <HIGH_LIMIT></HIGH_LIMIT>
    </SRU_PERCENT>
    <R_GROUP>
      <R_ID></R_ID>
      <LIMIT_TYPE></LIMIT_TYPE>
      <AVERAGE></AVERAGE>
      <LOW_LIMIT></LOW_LIMIT>
      <HIGH_LIMIT></HIGH_LIMIT>
    </R_GROUP>
    <DEGREE_POLYMERIZATION>
      <AVERAGE></AVERAGE>
      <LOW_LIMIT></LOW_LIMIT>
      <HIGH_LIMIT></HIGH_LIMIT>
    </DEGREE_POLYMERIZATION>
  </SRU_GROUP>
  <MW_GROUP>
    <TYPE_MW></TYPE_MW>
    <MW_AVERAGE></MW_AVERAGE>
    <LOW_LIMIT_MW></LOW_LIMIT_MW>
    <HIGH_LIMIT_MW></HIGH_LIMIT_MW>
  </MW_GROUP>
  <PHYSICAL_PROPERTIES_GROUP>
    <PHYSICAL_PROPERTY_TYPE></PHYSICAL_PROPERTY_TYPE>
    <AVERAGE></AVERAGE>
    <LOW_LIMIT></LOW_LIMIT>
    <HIGH_LIMIT></HIGH_LIMIT>
    <UNITS></UNITS>
  </PHYSICAL_PROPERTIES_GROUP>
  <COMMENTS>FORMS A THREE DIMENSIONAL NETWORK, EXTENT OF NETWORK
  DEPENDS ON RATIO OF FORMALDEHYDE TO NAPTHOL; OTHER POSITIONS ON RINGS
  CAN ALSO UNDERGO SUBSTITUTION </COMMENTS>
</POLYMER>
```


xi. Example: 6,6-NYLON.

Each of the polymerized monomers is fully drawn out and linked together .

Structure Field



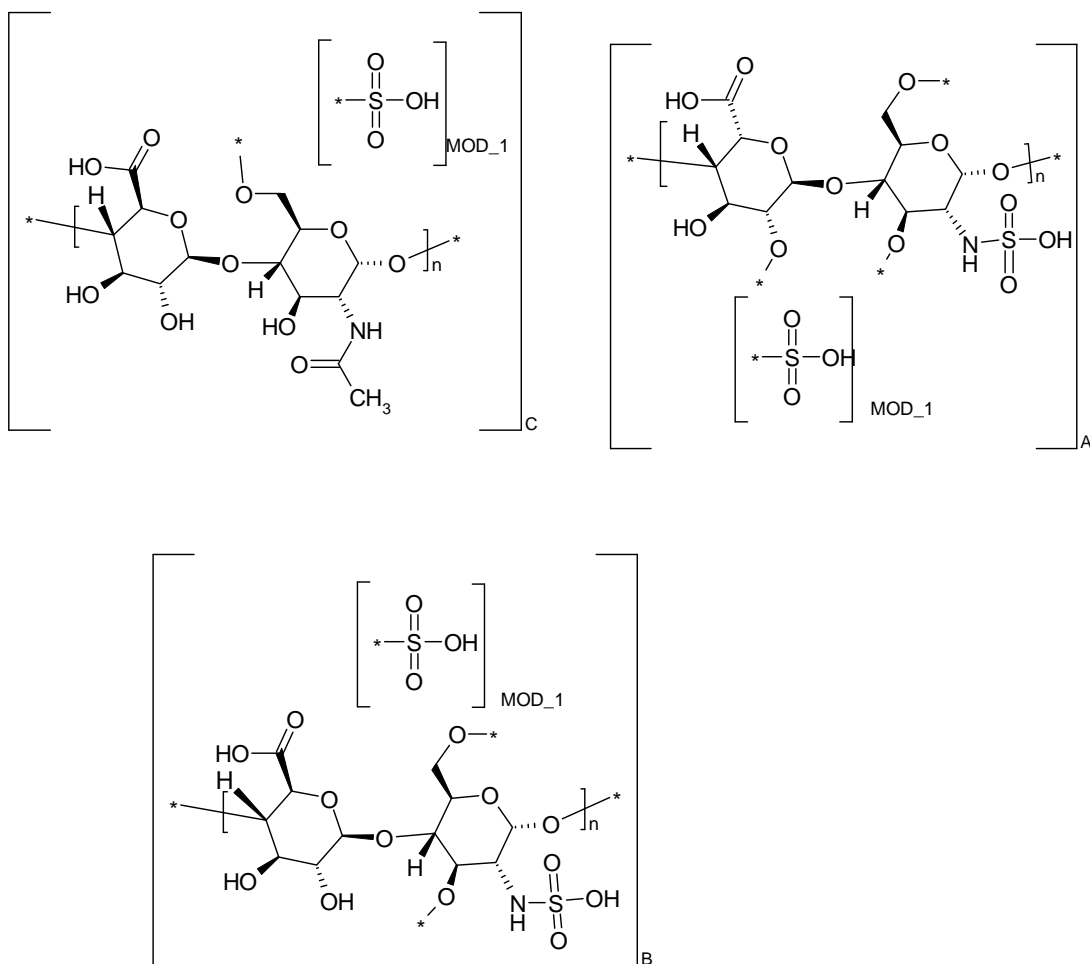
Description Field

```
<POLYMER>
  <POLYMER_TYPE>CONDENSATION</POLYMER_TYPE>
  <NUMBER_OF_SRU>1</NUMBER_OF_SRU>
  <SRU_GROUP>
    <SRU_ID>A</SRU_ID>
    <ORIENTATION_OF_POLYMERIZATION>
      </ORIENTATION_OF_POLYMERIZATION>
    <SRU_PERCENT>
      <AVERAGE></AVERAGE>
      <LOW_LIMIT></LOW_LIMIT>
      <HIGH_LIMIT></HIGH_LIMIT>
    </SRU_PERCENT>
    <R_GROUP>
      <R_ID></R_ID>
      <LIMIT_TYPE></LIMIT_TYPE>
      <AVERAGE></AVERAGE>
      <LOW_LIMIT></LOW_LIMIT>
      <HIGH_LIMIT></HIGH_LIMIT>
    </R_GROUP>
    <DEGREE_POLYMERIZATION>
      <AVERAGE></AVERAGE>
      <LOW_LIMIT></LOW_LIMIT>
      <HIGH_LIMIT></HIGH_LIMIT>
    </DEGREE_POLYMERIZATION>
  </SRU_GROUP>
  <MW_GROUP>
    <TYPE_MW>NUMBER AVERAGE</TYPE_MW>
    <MW_AVERAGE>10,000</MW_AVERAGE>
    <LOW_LIMIT_MW></LOW_LIMIT_MW>
    <HIGH_LIMIT_MW></HIGH_LIMIT_MW>
  </MW_GROUP>
  <PHYSICAL_PROPERTIES_GROUP>
    <PHYSICAL_PROPERTY_TYPE></PHYSICAL_PROPERTY_TYPE>
    <AVERAGE></AVERAGE>
    <LOW_LIMIT></LOW_LIMIT>
    <HIGH_LIMIT></HIGH_LIMIT>
    <UNITS></UNITS>
  </PHYSICAL_PROPERTIES_GROUP>
  <COMMENTS></COMMENTS>
</POLYMER>
```

xii. Example:.HEPARIN (Not Official).

When two moieties can substitute on a given atom such as acetyl and sulfate groups on the amino-group of the sugar, both moieties will be drawn as defined repeating units (see B and C below). When the polymeric

moiety and substituents within a bracket have the same overall molecular weight, the repeating unit that has the most positions of potential substitution will be given priority (see A, B below).



```

<POLYMER>
  <POLYMER_TYPE>COPOLYMER, NATURAL PRODUCT, RANDOM</POLYMER_TYPE>
  <NUMBER_OF_SRU>3</NUMBER_OF_SRU>
  <SRU_GROUP>
    <SRU_ID>A</SRU_ID>
    <ORIENTATION_OF_POLYMERIZATION>
      </ORIENTATION_OF_POLYMERIZATION>
    <SRU_PERCENT>
      <AVERAGE></AVERAGE>
      <LOW_LIMIT></LOW_LIMIT>
      <HIGH_LIMIT></HIGH_LIMIT>
    </SRU_PERCENT>
  </SRU_GROUP>
  <R_GROUP>
    <R_ID></R_ID>
  </R_GROUP>

```

```

        <LIMIT_TYPE></LIMIT_TYPE>
        <AVERAGE></AVERAGE>
        <LOW_LIMIT></LOW_LIMIT>
        <HIGH_LIMIT></HIGH_LIMIT>
    </R_GROUP>
    <DEGREE_POLYMERIZATION>
        <AVERAGE></AVERAGE>
        <LOW_LIMIT></LOW_LIMIT>
        <HIGH_LIMIT></HIGH_LIMIT>
    </DEGREE_POLYMERIZATION>
</SRU_GROUP>
<SRU_GROUP>
    <SRU_ID>B</SRU_ID>
    <ORIENTATION_OF_POLYMERIZATION>
        </ORIENTATION_OF_POLYMERIZATION>
    <SRU_PERCENT>
        <AVERAGE></AVERAGE>
        <LOW_LIMIT></LOW_LIMIT>
        <HIGH_LIMIT></HIGH_LIMIT>
    </SRU_PERCENT>
    <R_GROUP>
        <R_ID></R_ID>
        <LIMIT_TYPE></LIMIT_TYPE>
        <AVERAGE></AVERAGE>
        <LOW_LIMIT></LOW_LIMIT>
        <HIGH_LIMIT></HIGH_LIMIT>
    </R_GROUP>
    <DEGREE_POLYMERIZATION>
        <AVERAGE></AVERAGE>
        <LOW_LIMIT></LOW_LIMIT>
        <HIGH_LIMIT></HIGH_LIMIT>
    </DEGREE_POLYMERIZATION>
</SRU_GROUP>
<SRU_GROUP>
    <SRU_ID>C</SRU_ID>
    <ORIENTATION_OF_POLYMERIZATION>
        </ORIENTATION_OF_POLYMERIZATION>
    <SRU_PERCENT>
        <AVERAGE></AVERAGE>
        <LOW_LIMIT></LOW_LIMIT>
        <HIGH_LIMIT></HIGH_LIMIT>
    </SRU_PERCENT>
    <R_GROUP>
        <R_ID></R_ID>
        <LIMIT_TYPE></LIMIT_TYPE>
        <AVERAGE></AVERAGE>
        <LOW_LIMIT></LOW_LIMIT>
        <HIGH_LIMIT></HIGH_LIMIT>
    </R_GROUP>
    <DEGREE_POLYMERIZATION>
        <AVERAGE></AVERAGE>
        <LOW_LIMIT></LOW_LIMIT>
        <HIGH_LIMIT></HIGH_LIMIT>
    </DEGREE_POLYMERIZATION>
</SRU_GROUP>
<MW_GROUP>
    <TYPE_MW>NUMBER AVERAGE</TYPE_MW>

```

```

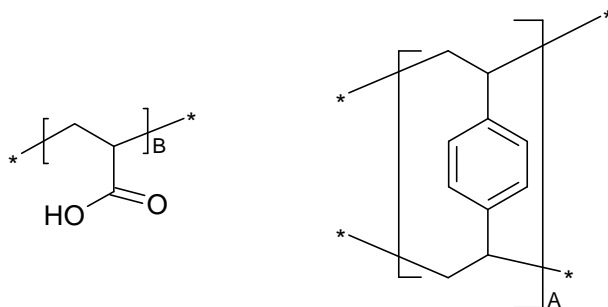
    <MW_AVERAGE></MW_AVERAGE>
    <LOW_LIMIT_MW>12,000</LOW_LIMIT_MW>
    <HIGH_LIMIT_MW>15,000</HIGH_LIMIT_MW>
  </MW_GROUP>
  <PHYSICAL_PROPERTIES_GROUP>
    <PHYSICAL_PROPERTY_TYPE></PHYSICAL_PROPERTY_TYPE>
    <AVERAGE></AVERAGE>
    <LOW_LIMIT></LOW_LIMIT>
    <HIGH_LIMIT></HIGH_LIMIT>
    <UNITS></UNITS>
  </PHYSICAL_PROPERTIES_GROUP>
  <COMMENTS>AVERAGE NEGATIVE CHARGE PER RESIDUE 2.3</COMMENTS>
</POLYMER>

```

xiii. Example: POLACRILIN

Polacrillin is a synthetic polymer consisting of acrylate residues crosslinked with divinylbenzene. The amount of divinylbenzene present is between 2 and 10%. In the structure field, divinylbenzene is represented by the para-moiety even though it is often a mixture of ortho, meta, and para isomers. The following statement "THE CROSSLINKING MOIETY DIVINYLBENZENE IS A MIXTURE OF THE META, ORTHO AND PARA ISOMERS" is placed in the comments. A similar statement is used if the crosslinking agent is a single isomer.

Structure Field



Description Field

```

<POLYMER>
  <POLYMER_TYPE>COPOLYMER</POLYMER_TYPE>
  <NUMBER_OF_SRU>2</NUMBER_OF_SRU>
  <SRU_GROUP>
    <SRU_ID>A</SRU_ID>
    <ORIENTATION_OF_POLYMERIZATION>
      </ORIENTATION_OF_POLYMERIZATION>
    <SRU_PERCENT>
      <AVERAGE></AVERAGE>
      <LOW_LIMIT>2</LOW_LIMIT>
      <HIGH_LIMIT>10</HIGH_LIMIT>
    </SRU_PERCENT>
  </SRU_GROUP>
  <R_GROUP>
    <R_ID></R_ID>
    <LIMIT_TYPE></LIMIT_TYPE>
  </R_GROUP>
</POLYMER>

```

```

        <AVERAGE></AVERAGE>
        <LOW_LIMIT></LOW_LIMIT>
        <HIGH_LIMIT></HIGH_LIMIT>
    </R_GROUP>
    <DEGREE_POLYMERIZATION>
        <AVERAGE></AVERAGE>
        <LOW_LIMIT></LOW_LIMIT>
        <HIGH_LIMIT></HIGH_LIMIT>
    </DEGREE_POLYMERIZATION>
</SRU_GROUP>
<SRU_GROUP>
    <SRU_ID>B</SRU_ID>
    <ORIENTATION_OF_POLYMERIZATION>
        </ORIENTATION_OF_POLYMERIZATION>
    <SRU_PERCENT>
        <AVERAGE></AVERAGE>
        <LOW_LIMIT>90</LOW_LIMIT>
        <HIGH_LIMIT>98</HIGH_LIMIT>
    </SRU_PERCENT>
    <R_GROUP>
        <R_ID></R_ID>
        <LIMIT_TYPE></LIMIT_TYPE>
        <AVERAGE></AVERAGE>
        <LOW_LIMIT></LOW_LIMIT>
        <HIGH_LIMIT></HIGH_LIMIT>
    </R_GROUP>
    <DEGREE_POLYMERIZATION>
        <AVERAGE></AVERAGE>
        <LOW_LIMIT></LOW_LIMIT>
        <HIGH_LIMIT></HIGH_LIMIT>
    </DEGREE_POLYMERIZATION>
</SRU_GROUP>
<MW_GROUP>
    <TYPE_MW></TYPE_MW>
    <MW_AVERAGE></MW_AVERAGE>
    <LOW_LIMIT_MW></LOW_LIMIT_MW>
    <HIGH_LIMIT_MW></HIGH_LIMIT_MW>
</MW_GROUP>
<PHYSICAL_PROPERTIES_GROUP>
    <PHYSICAL_PROPERTY_TYPE></PHYSICAL_PROPERTY_TYPE>
    <AVERAGE></AVERAGE>
    <LOW_LIMIT></LOW_LIMIT>
    <HIGH_LIMIT></HIGH_LIMIT>
    <UNITS></UNITS>
</PHYSICAL_PROPERTIES_GROUP>
<COMMENTS>THE CROSSLINKING MOIETY DIVINYLBENZENE IS A MIXTURE OF
THE META, ORTHO AND PARA ISOMERS</COMMENTS>
</POLYMER>

```

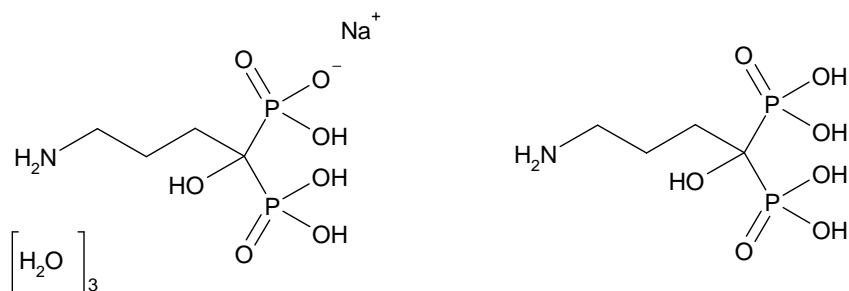
29. Active Moieties

Active moieties are included in the FDA/USP Substance Registration System (SRS). An active moiety is defined in 21CFR316.3 as “the molecule or ion, excluding those appended portions of the molecule that cause the drug to be an ester, salt (including a salt with hydrogen or coordination bonds), or other noncovalent derivative (such as a complex, chelate, or clathrate) of the molecule, responsible for the physiological or pharmacological action of the drug substance.”

Active moieties will be added to the SRS without regard to the actual charged state of the molecule in-vivo. By convention in the FDA/USP SRS, the active moiety of a hydrochloride salt of a base will be the free base and not the protonated form of the base. Likewise, the active moiety of a metal acid salt will be the free acid. This definition applies when a molecule can be neutralized by addition or loss of a proton. For quaternary ammonium cations that contain pharmacological activity, the active moiety is the charged cation. This ensures there will be only one active moiety for each single drug substance.

Care should be taken to ensure that an ester is not essential for the pharmacological activity of a substance. An active moiety can be contained within either the alcohol or acidic portion of the molecule. The reviewing division within FDA should provide individual guidance for each ester to determine the active moiety or whether an active moiety exists.

- i. Example: ALENDRONATE SODIUM active moiety is ALENDRONIC ACID



- ii. Example: CHELOCARDIN HYDROCHLORIDE active moiety is CHELOCARDIN

