Guidance for the graphic representation of pharmaceutical substances in the publications of International Nonproprietary Names and The International Pharmacopoeia

Please submit your comments through the online platform, PleaseReview™ (https://who.pleasereview.net/Main/Default.aspx?action=loaddocument&reviewid=270) If not registered or included in our mailing list, kindly submit your request with your full name, email address and organization/affiliation to nsp@who.int.

For any technical questions, you may contact **Dr Herbert Schmidt**, Technical Officer, Norms and Standards for Pharmaceuticals, Technical Standards and Specifications (schmidth@who.int), with a copy to Ms Bezawit Kibret (kibretb@who.int)

Comments should be submitted through the online platform by **24 September 2024**. Please note that only comments received by this deadline will be considered for the preparation of this document.

Our working documents are sent out electronically and uploaded into PleaseReview™. The working documents are also placed on the WHO Medicines website (https://www.who.int/teams/health-product-and-policy-standards/standards-and-specifications/pharmaceuticals/working-documents-public-consultation) under “Working documents in public consultation”.

If you wish to receive all our draft guidelines during the course of the year, please send your full name, organization/affiliation and email address to jonessi@who.int, nsp@who.intand your name will be added to our electronic mailing list and review platform.

Please send any request for permission to: Ms Bezawit Kibret, Norms and Standards for Pharmaceuticals, Technical Standards and Specifications, Department of Health Products Policy and Standards, World Health Organization, CH-1211 Geneva 27, Switzerland, email: kibretb@who.int.

Introduction and background

During the Fiftieth meeting of the World Health Organization (WHO) Expert Committee on Specifications for Pharmaceutical Preparations (ECSPP), a possible revision of the general guidance on how to represent graphic formulae was discussed. The revision of guidance on representation of graphic formulae on how to represent graphic formulae in *The International Pharmacopoeia* (Ph.Int.) and within the International Nonproprietary Names (INN) list was developed and adopted by the Expert Committee at its Thirty-fourth meeting (Technical Report Series (TRS) 863, Annex 1, 1996). A discussion took place on whether an update of this guidance would be useful to bring it into line with current practices. Such updated guidance could promote convergence in this area. The Expert Committee supported the proposal and recommended that work should start promptly to update WHO guidance on representation of graphic formulae.

During the Fifty-first ECSPP meeting, an update was given on the revision of this general guidance. Different approaches for graphical representation used by different pharmacopoeias may result in a risk of confusion and mistakes. There is a need to identify a harmonized approach that can be recommended for use by all organizations. The Expert Committee expressed its support for ongoing work in this regard which will enable an update of WHO guidance on representation of graphic formulae.

The work on the general guidance was resumed in March 2024. The resulting working document is presented for feedback.

This guide is intended to describe rules for drawing structures and to supersede the previous version, entitled *Guidelines for the graphic representation of chemical formulae*, published in the 34th report of the ECSPP, TRS 863, Annex 1, 1996, concerning conventions for chemical structure representation of pharmaceutical substances.

SCHEDULE FOR DRAFT WORKING DOCUMENT QAS/17.726:

Guidance for the graphic representation of pharmaceutical substances in the publications of International Nonproprietary Names and The International Pharmacopoeia

Description of Activity	**Date**
Presentation for discussion on possible revision of existing document on *Guidelines for the graphic representation of chemical formulae* during the 50th meeting of the WHO Expert Committee on Specifications for Pharmaceutical Preparations (ECSPP).	12–16 Oct 2015
Preparatory work and discussions.	15 Oct 2016
Presentation of the update and provision of practical examples to the 51st ECSPP.	10–16 Oct 2016
Preparation of new draft working document. .	Oct 2016–May 2017
Circulation for comments to INN experts, Expert Advisory Panel, world pharmacopoeias, WHO Collaborating Centres, involved in the establishment of monographs for the Ph.Int.	Jul–Sept 2017
Collation of feedback.	Oct 2017
Presentation of update to Fifty-second meeting of the ECSPP.	16–20 Oct 2017
Preparation of a revised version.	Oct–Dec 2017
Circulation for public consultation.	Jan–Mar 2018
Discussion of feedback in Working Group.	Apr–May 2018
Revision of working document.	May–Jun 2018
Circulation of working document for public consultation.	Jul–Aug 2018
Collation of feedback.	Sept 2018
Presentation to Fifty-third meeting of the ECSPP.	Oct 2018
Work on the guidance documents resumed.	March 2024
Circulation of working document for public consultation	July–August 2024

Presentation to the Fifty-eight meeting of the ECSPP.	Oct 2024
Any follow-up action, as needed.	

Contents

Preface

The programme on International Nonproprietary Names[1] (INN), The International Pharmacopoeia and other pharmacopoeias all over the world, use both chemical structures and chemical names with the aim of defining molecular entities as precisely and as unambiguously as possible.

As far as pharmaceutical substances are concerned, the molecular structure is a key factor in realizing a particular pharmacological function and, therefore, its pictorial representation, provided by sequence alignments or 2-D graphical plots, should not only give relevant information but also be relatively easy to understand.

The last publication[2] of the World Health Organization (WHO) treating structure representation know-how entitled “*The guidelines for the graphic representation of chemical formulae*” appeared in 1995. Only from 2006 on, official recommendations were published by the International Union of Pure and Applied Chemistry (IUPAC) concerning the graphical representation of chemicals[3,4] as well as their preferred naming[5]. The IUPAC conventions for drawing techniques cover a broad range of chemicals with an exhaustive list of rules and specific styles which are not all applicable and sometimes in contradiction with the way pharmaceutical substances are currently depicted.

Whereas several decades ago substances for publication were hand drawn, nowadays they are exclusively processed by computer-aided techniques. As such, and with time, the practice of drawing brought to light certain depiction styles that became more accepted than others, mostly because of convenience for interpretation, whether or not assisted with computer-aided analysis. The 1995 specifications were thus reconsidered in view of some of the published IUPAC rules applied to pharmaceutical substances and the progress made since. Furthermore, several new issues were addressed in order to permit ongoing harmonization of drawing practices.

Specifications described in this guideline are guided by three basic principles. First, a chemical structural diagram should reflect its associated chemical name. Second, rules should be generally applicable as much as possible and not be confined to a particular class of structures. Third, compounds belonging to the same chemical class should be drawn in a similar way to make clear that they are structurally analogous.

Throughout the document, diagrams are shown with the intention to be representative of the topic being discussed. Although depiction styles are described, no recommendations are made of several minor qualitative (font, colour …) and quantitative parameters (bond lengths, line thickness …). Only bond angles are sometimes described with numerical precision but they need not be strictly interpreted. A precise description is used only to achieve a consistent style and uniform format and may be interpreted in practice as being approximate.

INN stems, if ever of any relevance, have been provided for substances cited as examples in order to facilitate searches on the WHO Mednet[6] platform for compounds belonging to the same chemical category within a pharmacological related group of substances.

Furthermore, hyperlinks have been introduced into the electronic version of this document for the ease of reading and switching between different sections, background literature, references, websites, etc.

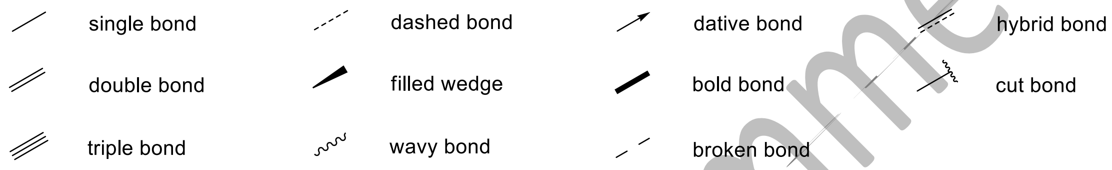
I. General issues

1.1. Bond types

Links between atoms are represented by dashes: a single bond is shown as a single dash, a double bond as a double dash, and a triple bond as a triple dash.

Fig. 1.1-1

Bond types



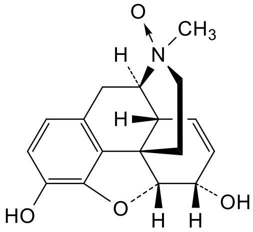
To specify the stereochemistry of a centre of asymmetry, a single bond can be replaced by a dashed bond (bond projecting behind the plane of depiction) or a filled wedge (bond projecting in front of that plane) whereas unknown stereochemistry can be indicated with a wavy bond. The rules for stereochemical representation are described in more detail in *section 1.5.* of this guide.

In some rare cases concerning mostly perspective drawings of bridged structures, bold bonds, as well as broken lines, may be used to avoid crossovers being misinterpreted as tetravalent connecting atoms.

In the case of coordination complexes *(section 1.4.3.)* and *N*-oxides (*section 1.3.2*), a dative bond may be used which is illustrated by an arrow pointing from the atom donating the lone pair to the atom accepting it.

Fig. 1.1-2

Structural diagram of *(17S)-morphine N-oxide* illustrated with single, double and dashed bonds, filled wedges and a broken, a bold and a dative bond



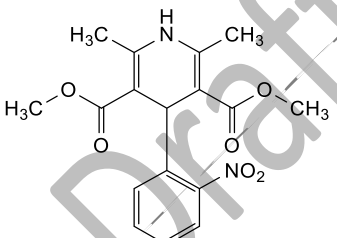
In order to suggest, to a certain degree, configurational information for any atom in a structure in general, at least one, but mostly two, bonds of normal thickness whether single, multiple, dative or broken, are usually supposed to be in the plane of the depiction surface of that atom.

Hybrid bonds may be used to display delocalization of electrons (*section 1.4.2.1.*) A cut bond may be used to indicate free valences of radicals (*section 1.6.1.*).

Bonds are drawn using a single consistent length but for clarity reasons some sterically congested structures might require singular bond lengths.

Fig. 1.1-3

Non-standard bond length as illustrated with *nifedipine*



1.2. Atomic symbols and structural abbreviations

Elements are indicated by their approved atomic symbols except for carbon and hydrogen atoms which are, most of the time, intentionally omitted to simplify the appearance of the main structure, except for terminal carbon atoms with the corresponding hydrogens. Atom labels may be modified by adding either charges (ionic structures) or mass numbers (isotopically modified compounds). These cases are discussed separately in *section 1.4.2.* and *section 1.4.4.* respectively of this guide.

Although characteristic groups should be shown in full, certain common groups of atoms are shown in a condensed form.

Fig. 1.2-1

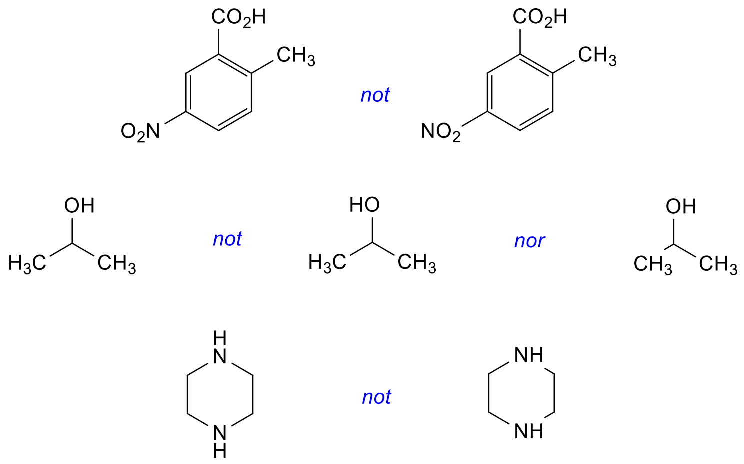
Groups shown in condensed form

$-CH_3$	methyl	$-CHO$	formyl
$-CN$	cyano	$-NC$	isocyano
$-CO_2H$	carboxy	$-CO_2^-$	carboxylate
$-OH$	hydroxy	$-OCH_3$	methoxy
$-NH_2$	amino	$-NO_2$	nitro
$-SO_3H$	sulfo	$-SO_3^-$	sulfonate
$-N_3$	azido	$-PO_3H_2$	phosphono

A polyatomic group is set out such that the atom directly attached to the rest of the structure is shown closest to the connecting dash. With a group oriented towards the right or vertically, all other atoms should preferably be on the right of the connecting atom and, if not, the atom label should be inversed. The atom label should always be oriented to minimize its overlap with any part of structure.

Fig. 1.2-2

Examples of different ways of representing polyatomic labels



Single- and *three-letter* amino acid abbreviations and nonstandard symbols for groups used in (poly)peptides are discussed in *section 2.1*.of this guide. Information on abbreviations characteristic for nucleotides can be found in references addressed in *section 2.3.*

1.3. Chemical groups

1.3.1. Expanded form

Except for some functional groups which are preferably depicted in condensed form (*section 1.2*), characteristic groups should be shown in full. The expanded form should, to some extent, suggest the corresponding configuration of that group.

Chemical groups containing an element with three explicit adjacent atoms should ideally be drawn with three equal angles (120 °) separating the bonds to suggest either trigonal planar or, in the case of an implicit hydrogen, tetrahedral configuration.

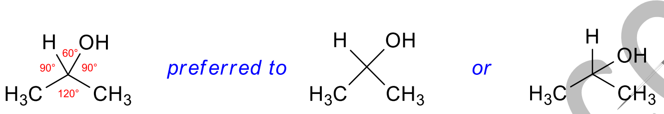
Fig. 1.3.1-1

Chemical groups drawn with three bonds

Chemical groups containing an element with four explicit adjacent atoms in tetrahedral configuration are best drawn with two adjacent bonds separated by a 60 ° angle, an opposing 120 ° angle and, in between, two 90 ° angles. A depiction with four equal angles is not recommended to avoid confusion with square planar configuration, neither graphics with two angles of 120 °, which display geometrically less accurate information.

Fig. 1.3.1-2

Chemical groups drawn with four bonds



Some chemical groups need special attention when depicted.

1.3.2. Representation of *N*-oxides

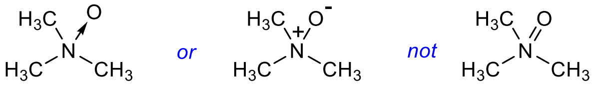
Although an *N*-oxide may be depicted with a single N-O bond, this representation complicates drawing as separate charges have to be displayed. Using a N=O double bond is clearly inappropriate from a chemical point of view as nitrogen is not an element with variable valences.[1] Using a dative bond (arrow pointing from the N to the O) not only clearly indicates the amine donating two electrons to an oxygen atom but implicitly suggests separated charges without the need of drawing them and, as such, this may seem the most obvious way to visualize *N*-oxides.

---

[1] Official IUPAC naming might however refer to hyper valence in the case when the oxide is on a nitrogen atom of a substituent of the parent structure; when, on the other hand, the oxide is located on a nitrogen atom of the parent structure, amine oxides are named by functional class nomenclature using the class name 'oxide'.

Fig. 1.3.2-1

Representation of *N*-oxides

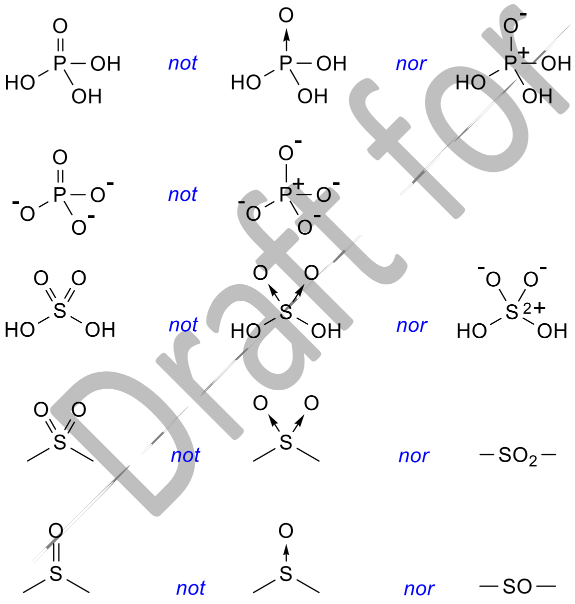


1.3.3. Representation of functional groups containing elements with variable valences

Whereas second-row atoms such as nitrogen and oxygen are supposed to be limited by the octet rule, third-row atoms (and beyond), such as phosphorus and sulfur, can present nonstandard bonding patterns and may form more than four covalent bonds. Characteristic groups containing these heteroatoms are fully depicted using single and double bonds without the addition of charges.

Fig. 1.3.3-1

Representation of groups with nonstandard bonding patterns



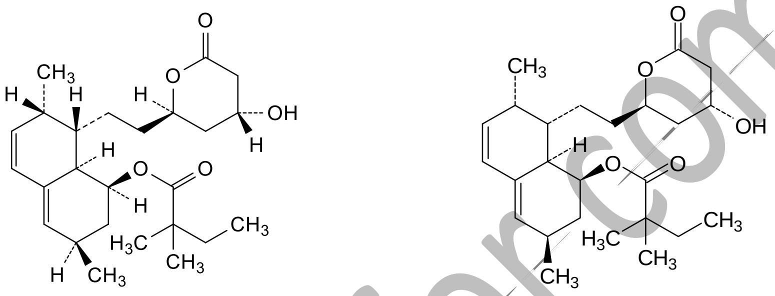
1.4. Structures

1.4.1. General aspects

The main skeleton should, by and large, be shown in full. Skeletal bonds are to be represented in the form of lines at an angle to one another where each corner represents a carbon atom unless explicitly labelled as a heteroatom. Where, for reasons of clarity, the skeletal carbon and hydrogen atoms need to be drawn explicitly - such as in terminal groups, collinear double bonds, or to indicate stereochemistry - they are usually omitted to simplify the appearance of the main structure.

Fig. 1.4.1-1

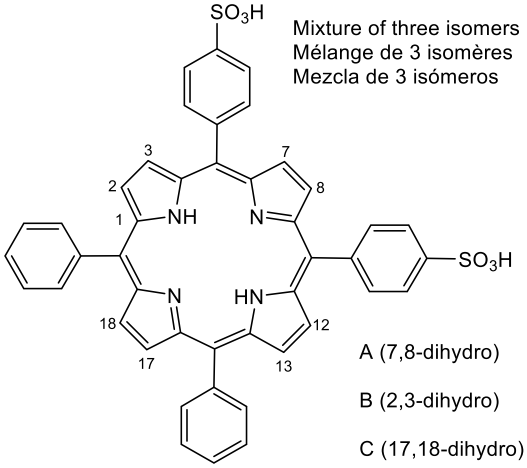
Representation of *simvastatin* with and without explicit hydrogen atom labels



The numerical indexing of skeletal atoms concerning either linear or cyclic structures is consistent with the established chemical nomenclature and is usually not shown, except for some very rare cases where it is important to emphasize certain structural features.

Fig. 1.4.1-2

Representation illustrating numerical indexing of skeletal atoms; the example shown is *fimaporfin*, a mixture of three isomers



Mixture of three isomers
Mélange de 3 isomères
Mezcla de 3 isómeros

A (7,8-dihydro)

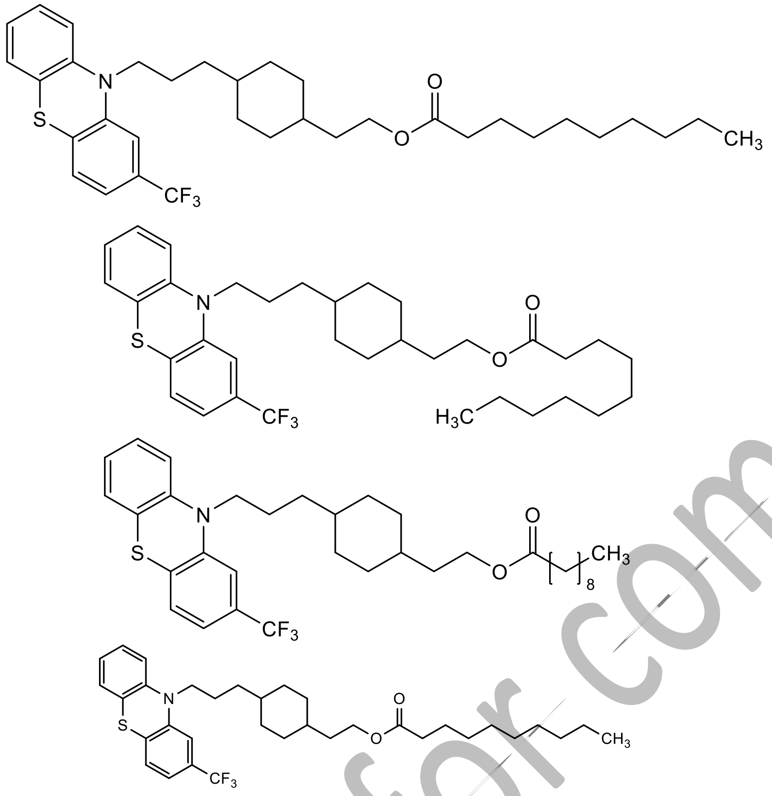
B (2,3-dihydro)

C (17,18-dihydro)

Large molecules may require size reduction to fit within the available space but, when possible, it is preferable to rotate a portion of the structure around a single bond if doing so will allow the diagram to fit without resizing. For structures with many identical groups, it may be useful to use square brackets rather than folding back the structure in the plane. The repeating units, either linear or branched, are to be placed in square brackets, their number being indicated by a subscript on the right. If one of these groups terminates the chain or bears a heteroatom, it is shown outside the square brackets.

Fig. 1.4.1-3

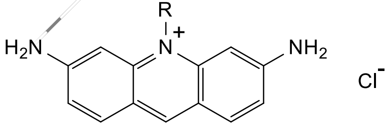
Different representations of *fluphenazine decanoate* illustrating various ways of handling structure sizes



In some rare cases, a pharmaceutical substance may be defined as a mixture of different but relatively similar structures that are mostly depicted through a generic formula (*section 1.7.*) shown along with information on the various compounds making up the mixture.

Fig. 1.4.1-4

Representation of a generic structure and accompanying information for *acriflavinium chloride*

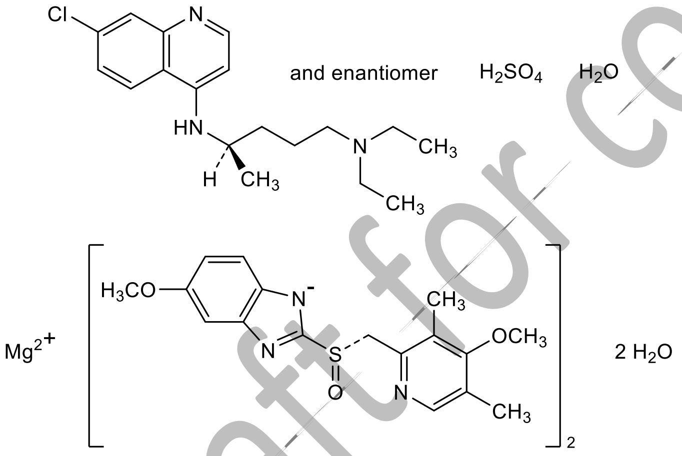


R	Mol. Formula	M_r
H	$C_{13}H_{12}ClN_3$	245,7
CH_3	$C_{14}H_{14}ClN_3$	259,7

Different components are sometimes associated with a parent structure to represent a substance. Those fragments mainly concern water of crystallization, substances of solvation, counter-ions, acids and bases. Such fragments are normally separated from each other and the principal structure by a distance approximately equivalent to the structure's standard bond length. Numerals used to indicate multiple occurrences of components are always preceded and followed by a space. Bracketed fragments are subscripted on the right-hand side bracket to indicate multiple occurrences of that fragment. The parent structure is usually followed in a specific order with the fragment representing an acid (or base) then water. If any textual information concerning stereochemistry is to be added, it is inserted straight after the parent structure and before any other component. Positioning of ionic species will be discussed in *section 1.4.2.* of this guide.

Fig. 1.4.1-5

Examples of "multicomponent structures" as illustrated with *chloroquine sulfate hydrate* (top) and *esomeprazole magnesium dihydrate* (bottom)

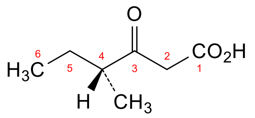


1.4.1.1. Distinctive aspects of acyclic structures

Wherever possible, structures are set out horizontally rather than vertically and are orientated so that the highest numbered atom in an acyclic structure is on the left with the numbering of atoms decreasing from left to right.

Fig. 1.4.1.1-1

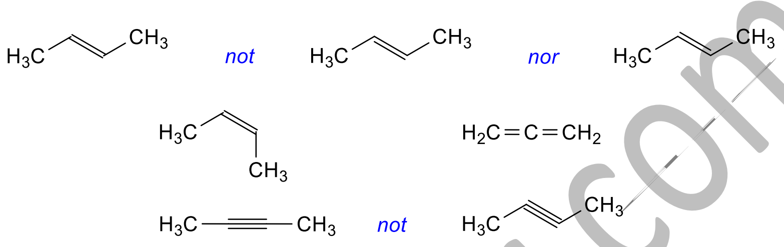
Orientation of linear structures



Simple bonds are to be represented in the form of lines at a regular angle of 120 ° to one another, whereas double bonds are shown as centered between the two atoms with varying angles depending on the geometry and hybridization of the concerned atoms. Triple bonds are to be drawn in a linear fashion rather than in a zigzag format because triple bonds involve sp-hybridized atoms and have linear geometry.

Fig. 1.4.1.1-2

Representation of different linear chain types

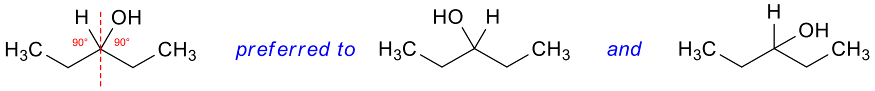


When only one substituent attached to an atom (carbon or heteroatom) occupying a position in the chain is depicted in order to represent either a trigonal planar or tetrahedral (with one implicit hydrogen atom) configuration, it should ideally be drawn with three equal angles (120 °) separating the bonds as in *Figure 1.3.1-1*.

Where two substituents are attached to the same chain atom, they should generally both be at the same angle to the bisector of the bonds formed by the "connecting" and adjacent chain atoms. Substituents should preferably be at a right angle to the chain (substituents separated by a 60 ° angle with an opposing 120 ° angle of the chain). If both substituents are not identical, the larger atom label of the two should preferably be on the right.

Fig. 1.4.1.1-3

Representation of substituted chains



Adjustments of chain angles may sometimes be necessary in congested structures to prevent overlap, or simply for aesthetic reasons, depending on the overall appearance of the drawing. In those cases, it is often sufficient to modify one particular angle of that structure by orienting a bond either horizontally or vertically or sometimes somewhere in between, mostly nearby a (multiple of a) 30 ° angle.

Fig. 1.4.1.1-4

Orientation of a linear fragment in congested structures as illustrated with *losartan potassium*

1.4.1.2. Distinctive aspects of cyclic structures

Rings are shown in full as polygons. The symbol of the carbon atoms that form the ring are not shown but are represented by the vertices of the rings. The hydrogen atoms attached to them are not represented unless they are needed to show stereochemistry. The symbols of atoms other than carbon are shown with all the hydrogen atoms attached to them but without linking dashes. Cyclic double bonds are shown with a dash inside the ring rather than being centred between atoms and triple bond have a dash outside and inside the ring and are centred between two atoms.

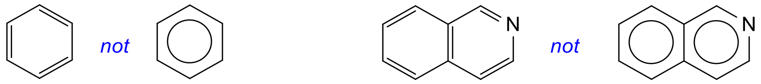
Fig. 1.4.1.2-1

Representation of ring skeletons

In aromatic systems, a circle should not be used to depict delocalized electrons; instead, alternating single and double bonds are shown (Kekulé representation).

Fig. 1.4.1.2-2

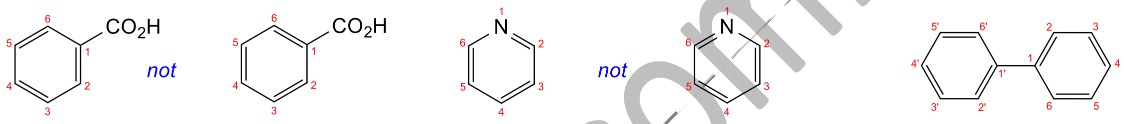
Representation of delocalized electrons in ring systems



In monocyclic unsaturated compounds, double bonds should be arranged to have the lowest possible numbering, where practicable, and rings should be numbered in a clockwise direction.

Fig. 1.4.1.2-3

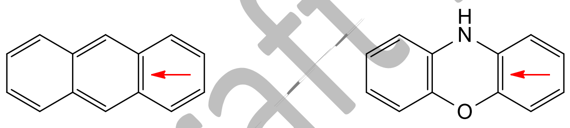
Lowest numbering for ring double bonds



In fused polycyclic systems, a double bond should form the fusion bond nearest to the right-hand side.

Fig. 1.4.1.2-4

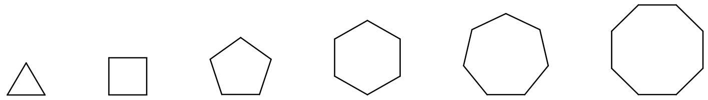
Positioning of double bonds in fused polycyclic systems



Rings are shown as regular polygons when they consist of up to 8 carbon atoms.

Fig. 1.4.1.2-5

Regular polygons of up to eight atoms



Whereas isolated rings of 3,4,5,7 and 8 atoms are preferably drawn with a horizontal bond at the bottom of the ring, six-membered rings should preferably be represented with a vertex at the base.

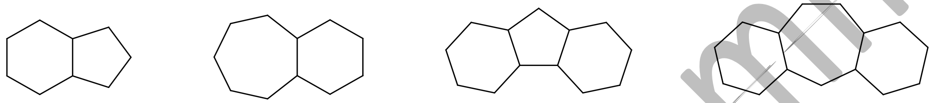
Fig. 1.4.1.2-6

Orientation of six-membered rings

Wherever possible, the regularity of the polygons is maintained in the drawing of fused cyclic compounds.

Fig. 1.4.1.2-7

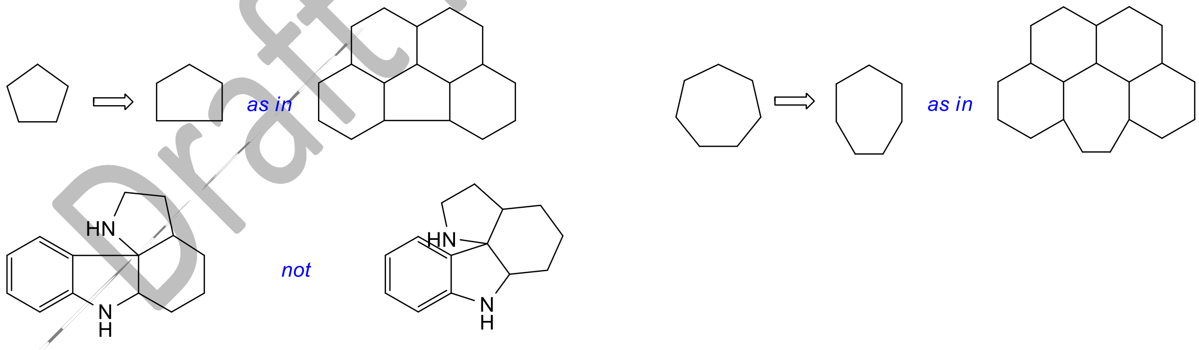
Fused ring systems with regular polygons



However, in fused polycyclic systems the polygons may often be distorted to maintain the symmetry of the structure or to prevent an overlap of rings.

Fig. 1.4.1.2-8

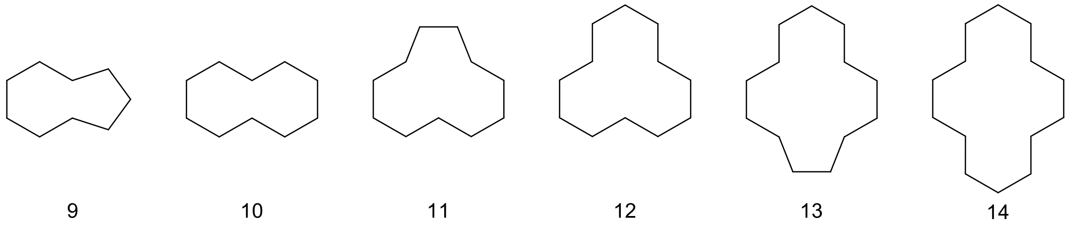
Distortion of rings



Rings with more than 8 vertices are often shown with re-entrant angles. IUPAC recommends that they should be drawn like amalgamated rings with 5, 6 or 7 vertices, encircling a maximum of virtual hexagons.

Fig. 1.4.1.2-9

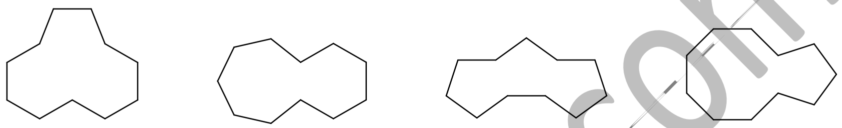
Rings with more than eight vertices



As overlap of groups is to be avoided, and double bond configuration needs to be truly depicted, other conformations may be considered depending on the constraints imposed by the presence of substituents and double bonds.

Fig. 1.4.1.2-10

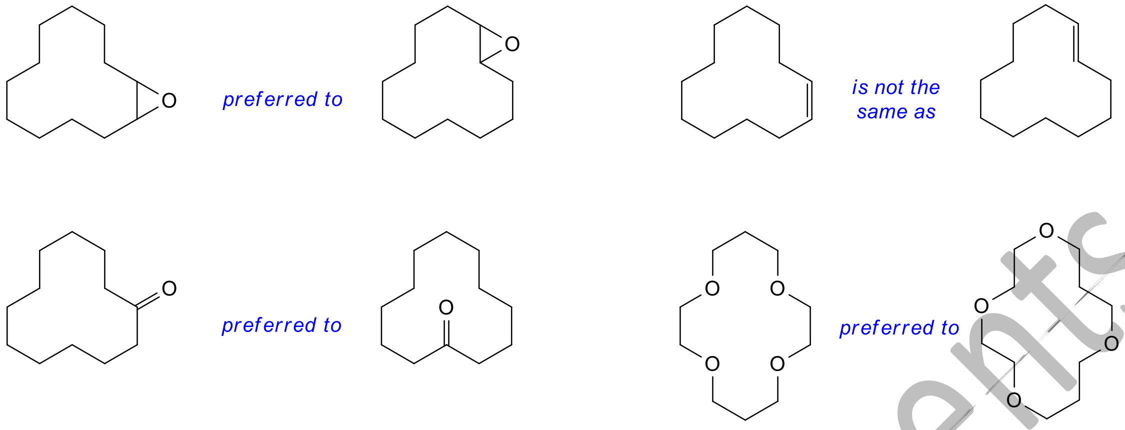
Alternative conformations for 11-membered rings



Care should be given as to avoid depictions with substitution at re-entrant positions. Substituents and fused rings should be outside the ring as far as possible. Furthermore, double bonds in any large rings should always be depicted in the correct stereochemical configuration. Finally, large heterocycles, in the absence of other constraints, are to be depicted if possible, with heteroatoms at the re-entrant positions.

Fig. 1.4.1.2-11

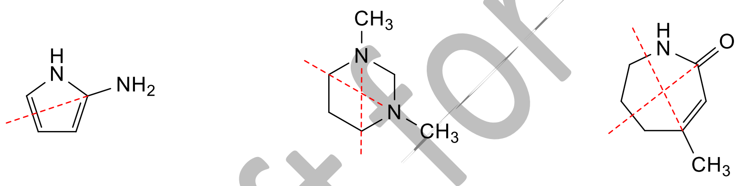
Depiction of different types of large rings



When a substituent is attached to an atom occupying a position in a ring (carbon or heteroatom), the direction to be taken by the dash, linking it to that atom, can be found by extending the line bisecting the cycle.

Fig. 1.4.1.2-12

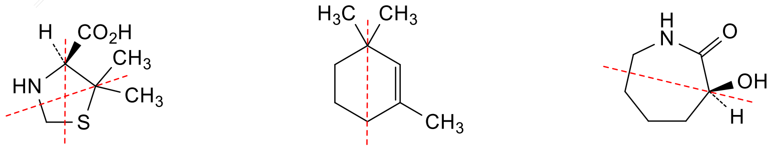
Orientation of ring substituents



Where two substituents are attached to the same ring atom, they should generally both be at the same angle to the bisector and preferably at a right angle to the adjacent side.

Fig. 1.4.1.2-13

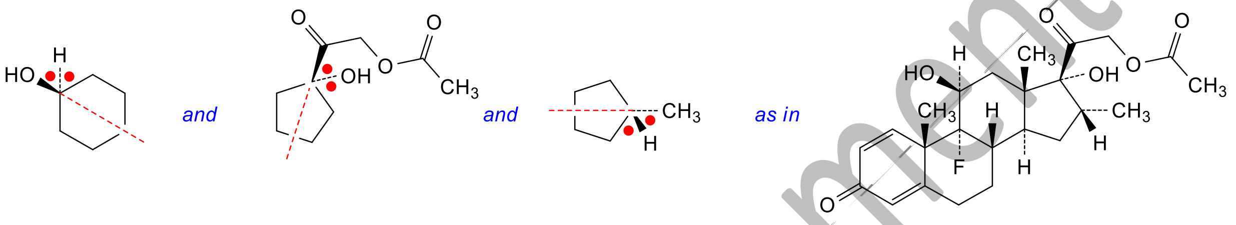
Orientation of multiple ring substituents



The positioning of substituents in the case of certain congested structures, like steroids, may however require adjustments, in comparison to the recommendations mentioned above, to avoid an overlap of structure fragments or simply for aesthetic reasons. In these cases, one (preferably the larger) substituent may be oriented by extending the line bisecting the cycle to which it is attached, and the second (smaller) one, to divide in equal parts the angle in between the first substituent and the ring.

Fig. 1.4.1.2-14

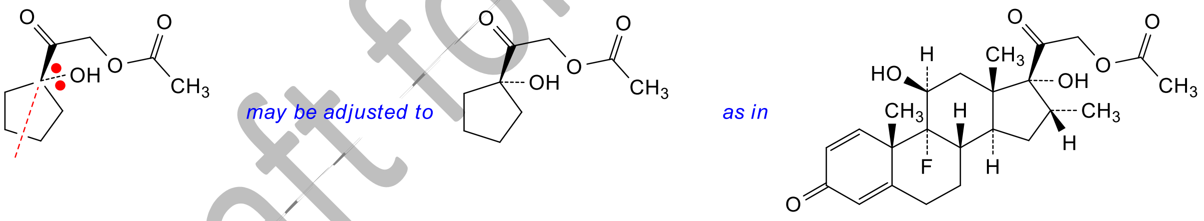
Orientation of multiple ring substituents in congested structures



The orientation of substituents that are nearly horizontal, vertical or nearby a (multiple of a) 30 ° angle when bisecting the adjacent bonds may be adjusted to such an angle for aesthetic reasons.

Fig. 1.4.1.2-15

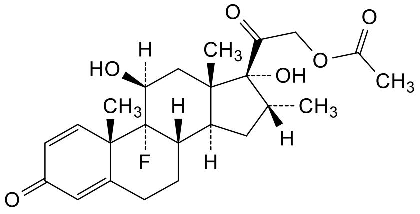
Adjustment of orientation of ring substituents in congested structures



Substituents and explicit hydrogens are normally placed outside cyclic systems, except for crowded structures such as steroids and some alkaloids (see *sections 3.5.2.* and *3.2.*) where substituents and explicit hydrogens attached at bridge heads can, if not enough space, be displayed inside the rings of polycyclic structures.

Fig. 1.4.1.2-16

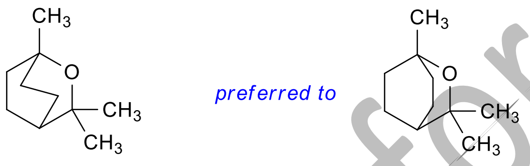
Positioning of substituents and explicit hydrogens on crowded ring systems

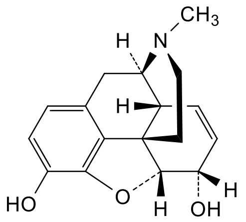


In bridged structures, a non-atomic bridge (direct bond) is represented by a straight line and an atomic bridge by lines at a clear angle to one another. The symbols for carbon atoms are not shown, however, if the bridge contains one or more heteroatoms, the atomic symbols for those atoms are shown. To give some perspective to the figure or to represent stereochemical features, wedges, thickened lines or broken lines can be used.

Fig. 1.4.1.2-17

Representation of bridged structures

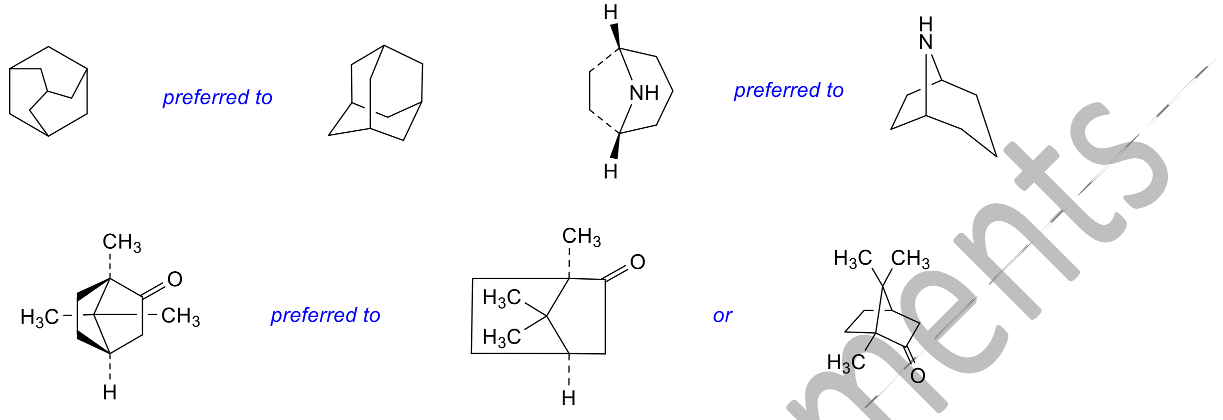




Planar representations, if considered clear enough, are in general preferred to perspective drawings.

Fig. 1.4.1.2-18

Planar representations as compared to perspective drawings with *adamantane* (top left), *tropane* (top right) and *D-camphor* (bottom)



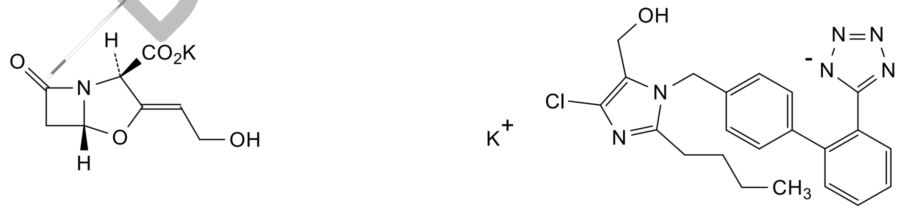
1.4.2. Ionic structures

1.4.2.1. General aspects

Ionic species are drawn either as neutral components or as structures with explicit positive and negative charges. In the latter case, the cationic part is placed on the left and the anionic fragment on the right. Cations and anions are separated by space rather than a comma or a dot with a separation distance approximately equivalent to the structure's standard bond length.

Fig. 1.4.2.1-1

Representation of the ionic species, *potassium clavulanate* (left) and *losartan potassium* (right) respectively without and with explicit charge symbols



Ionic charges are represented by either a *plus* or a *minus* sign for respectively positive and negative charges which are drawn as such without being encircled. Multiple charges are indicated by using a numerical multiplier *n* as in *n*+ or *n*- and not by writing the + or - symbol *n* times.

Fig. 1.4.2.1-2

Numerical multiplier for the representation of multiple charges of *calcium*.

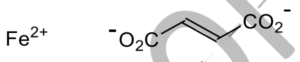
Ca^{2+} *not* Ca^{++}

Charge symbols associated with specific elements are usually shown on the top right-handed side of the (poly)atomic label, unless the label is reversed, in which case the charge is shown as a superscript on the left.

Fig. 1.4.2.1-3

Positioning of charge symbols on atom labels; the example shown is *ferrous fumarate*

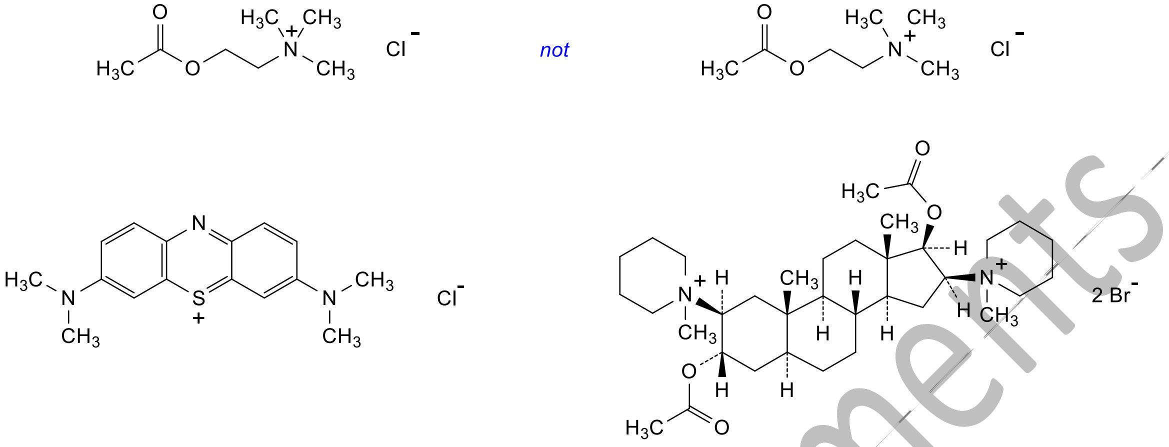
Fe^{2+} *or* Fe^{2+}



In some cases, if there is little or no space for a superscripted charge on the right-hand side of the atom, the charge can be shown immediately above the atom concerned. When a ring is involved, the charge is usually placed outside the ring unless it is impossible to place the charge without ambiguity, in which case it may then be shown inside the ring.

Fig. 1.4.2.1-4

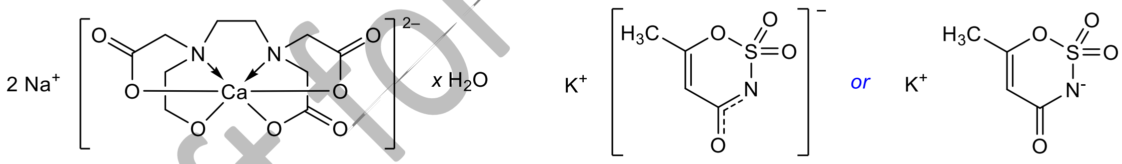
Positioning of charge symbols in congested structures, *acetyl choline chloride* (top), *methylthioninium chloride* (bottom left) and *pancuronium bromide* (bottom right)



In structures where the positions of the charges cannot be indicated precisely, the structure is put in square brackets, with the overall charge placed outside them as a superscript on the right.

Fig. 1.4.2.1-5

Representation of *sodium calcium edetate* (left) and *acesulfame potassium* (right)



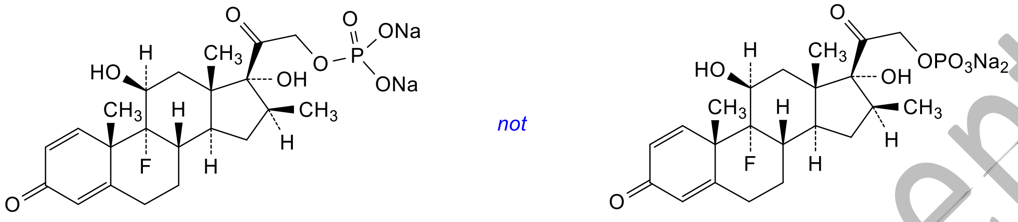
1.4.2.2. Distinctive aspects of inorganic metal salts

Metal salts (for example, $KMnO_4$) of inorganic acids are represented by their molecular formula without net charges. Structural formulas are normally not given for these substances except in the case of coordination compounds for which the stereochemistry of the complexes should be specified (*section 1.4.3.*). If salts include several metals, the symbols for the metals are shown in alphabetical order (for example, $AlK(SO_4)_2$). In acid salts, the metal precedes the hydrogen as in $NaHCO_3$. Molecules of water of crystallization or of substances of solvation follow the formula of the salt, from which they are separated by a middle dot, without a preceding or following space (for example, $AlK(SO_4)_2{\cdot}12H_2O$ and $NaH_2PO_4{\cdot}2H_2O$).

Metal salts of phosphoric and sulfuric esters are preferably shown in expanded form with the metal replacing the acid hydrogen.

Fig. 1.4.2.2-1

Representation of the metal salt of *betamethasone sodium phosphate*

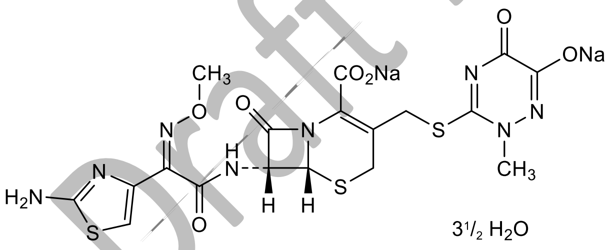


1.4.2.3. Distinctive aspects of organic metal salts

In the metal salts of organic (carboxylic) acids and the metal compounds of alcohols, phenols (and their sulfur, selenium and tellurium analogues), the metal symbol usually replaces the acid hydrogen but neither charges nor bonds are shown.

Fig. 1.4.2.3-1

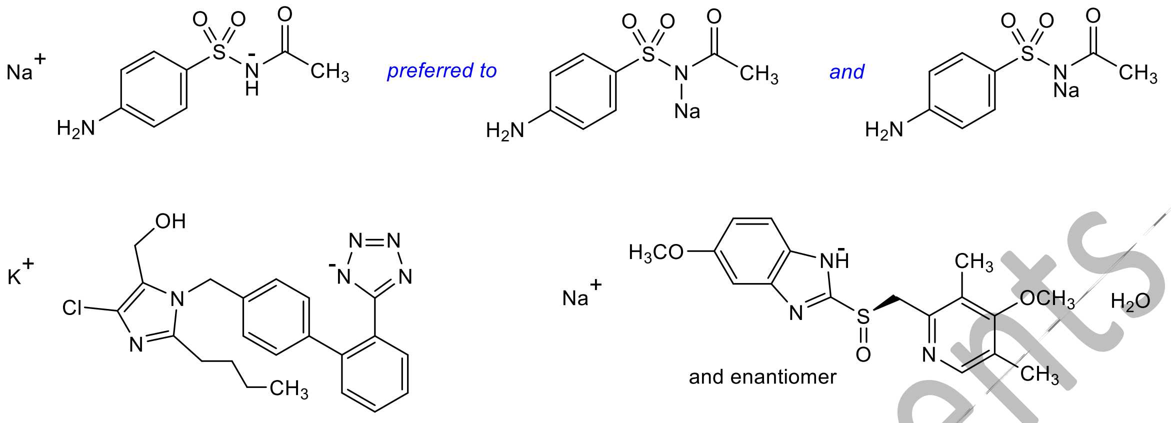
Representation of a metal salt of an organic acid and alcohol, *ceftriaxone sodium*



Metal salts of nitrogen-containing acidic groups like amides are preferably shown with explicit charges and are usually depicted as their preferred tautomeric neutral counterpart in which the mobile hydrogen is replaced with a negative charge or the metal atom (*section 1.4.5*).

Fig. 1.4.2.3-2

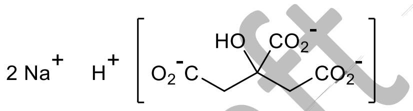
Representation of nitrogen-containing metal salts, *sulfacetamide sodium* (top), *losartan potassium* (bottom left) and *omeprazole sodium* (bottom right)



Ionic forms may be used in case of mixed and acidic salts when substances contain several anionic groups to which various cations cannot easily be attributed. The larger charged group should be enclosed in square brackets to emphasize that there is no precise structural association between the ionic fragments. Associated counter ions are separated from each other by a space. In structural formula, numerals used to indicate multiple occurrences of counter ions (but also water of crystallization, substances of solvation and acids) are likewise preceded and followed by a space.

Fig. 1.4.2.3-3

Representation of the acidic salt *disodium hydrogen citrate*

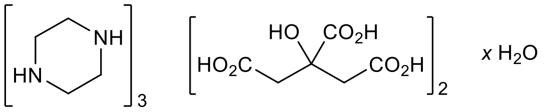


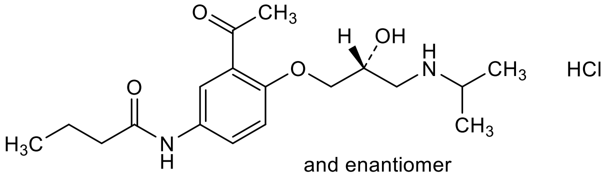
1.4.2.4. Distinctive aspects of amine salts

Amine salts are shown with the structure of the neutral amine on the left (as if it were in the ammonium form) and, after a space, the formula of the organic or inorganic acid on the right.

Fig. 1.4.2.4-1

Representation of the amine salts *piperazine citrate* (top) and *acebutolol hydrochloride* (bottom)

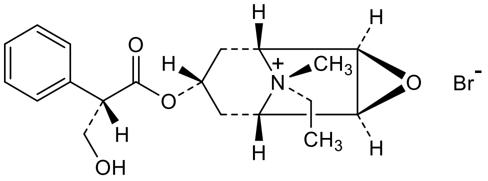




Quaternary ammonium salts and other compounds with a positive charge on a heteroatom (P, As, Sb, O, S, Se and Te) are shown in ionic form with explicit charges, the two ions being separated by space.

Fig. 1.4.2.4-2

Representation of the quaternary amine salt *oxitropium bromide*

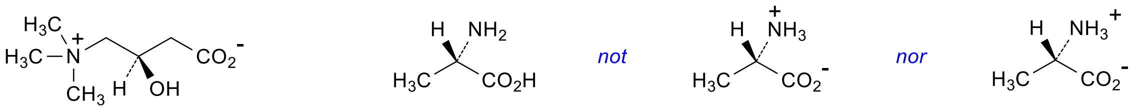


1.4.2.5. Distinctive aspects of zwitterions

For zwitterions, the *plus* and *minus* charges are drawn and placed according to the rules described above, provided that these charges are indispensable. If a choice is possible, a neutral form is preferred.

Fig. 1.4.2.5-1

Representation of zwitterions *levocarnitine* (left) and *alanine* (right)



1.4.3. Coordination compounds[7]

Coordination complexes, other than metal salts discussed above (*section 1.4.2*), are either depicted with condensed linear formulas or drawn with expanded diagrams, depending on the complexity of the structure.

1.4.3.1. Linear formulas of non-cyclic coordination compounds

In the linear formula of a coordination complex, the symbol of the central atom is placed on the left and is followed by the ionic ligands and then the neutral ligands.

Polyatomic ligands are placed in parentheses, with the atom linked to the central atom on the left. If several identical ligands are attached to the central atom, their number is indicated as a subscript to the right. In each class of ligands, the symbols of the linking atoms, and then of any other atoms, are shown in alphabetical order. The complete formula of the coordination entity (neutral group or complex ion) is placed in square brackets.

If the entire structure consists of ions, the positive ions are placed on the left and the negative ions on the right. No spaces should be left between representations of ionic species within the formula of a coordination compound. If the charge of the coordination entity needs to be specified, it is placed outside the square bracket as a right superscript. The individual charges usually carried by the central atom and the ligands are normally not shown. If a compound contains water, it is cited last and separated from the rest of the formula by a middle dot without a preceding and following space.

Fig. 1.4.3.1-1

Representation of *diamminesilver(1+) fluoride* (left), *sodium nitroprusside dihydrate* (middle) and the ion *ferrocyanide* (right)

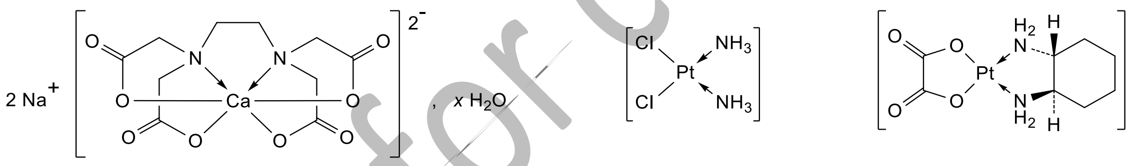
$[Ag(NH_3)_2]F$ $Na_2[Fe(CN)_5(NO)]\cdot 2H_2O$ $[Fe(CN)_6]^{4-}$

1.4.3.2. Structural formulas of coordination complexes

Some coordination complexes, whether cyclic or not, are best represented with a full depiction. Where possible, the metal atom is placed in the centre of the complex. Square brackets are placed around every coordination entity, even if the charge is zero. Coordination bonds are depicted with plain bonds unless filled wedges and dashed bonds are used to indicate stereo chemical configurations. Dative bonds may be used to avoid depicting atoms with non-standard valences. Individual charges carried by the central atom and the ligands are normally not indicated in structural formulas unless it is difficult to show all the coordination links.

Fig. 1.4.3.2-1

Representation of *sodium calcium edetate* (left), *cisplatin* (middle) and *oxaliplatin* (right)



“Sandwich” structures are shown with the rings connected to the central atom by a solid line starting from inside the cycle and passing through one side. Benzene rings and condensed benzene systems in “sandwich” compounds are drawn with alternating single and double bonds. Pentagonal and heptagonal rings are shown with a circle inside.

Fig. 1.4.3.2-2

Representation of “sandwich” structures *bis(benzene)chromium* (left) and *ferrocene* (right)

1.4.3.3. Stereochemistry of coordination complexes

The stereochemistry of mononuclear complexes is expressed by means of special descriptors formed from an abbreviation (system indicator) for the central atom geometry and the coordination number.

Configurations of polyhedral complexes are depicted by using different bond styles in perspective drawings. Both normal and dative bonds that are linked to the central atom (symbolized by “M” in the following figures) are usually supposed to be in the plane of depiction. Perspective drawings of chiral structures should preferably be drawn with stereobonds to orient coordination entities above and behind the depiction plane unless configuration is unknown, in which case only plain bonds are used. Perspective drawings of achiral structures should preferably be drawn with stereobonds to emphasize three-dimensional configuration.

A clear depiction of chelating groups in perspective drawings sometimes necessitates structures to be rotated and/or angles to be slightly distorted as compared to the following theoretical examples of polyhedral complexes.

T-4: tetrahedral complexes

These are described by the chirality symbols (*R*) and (*S*); they are shown in the same way as stereogenic carbon atoms, a dashed line denoting a bond projecting behind the plane of the depiction and a filled wedge one projecting in front of that plane.

Fig. 1.4.3.3-1

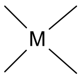
Representation of tetrahedral complexes

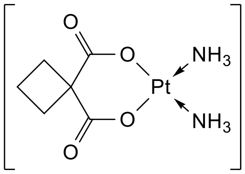
SP-4: square planar complexes

The four coordination links are shown in the plane of the depiction.

Fig. 1.4.3.3-2

Representation of square planar complexes; the example shown below is *carboplatin*



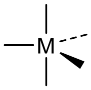


TBPY-5: trigonal bipyramidal complexes

The reference axis is shown in the plane of the depiction; of the three other equatorial ligands, one is assumed to be also in the plane of the depiction, one in front of it and the other behind it.

Fig. 1.4.3.3-3

Representation of trigonal bipyramidal complexes

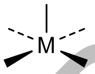


SPY-5: square pyramidal complexes

The reference axis with its lone coordinating atom is shown in the plane of the depiction and four coordination links are assumed to be in a plane perpendicular to the reference axis, two in front and two behind the plane of the depiction.

Fig. 1.4.3.3-4

Representation of square pyramidal complexes

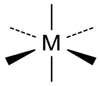


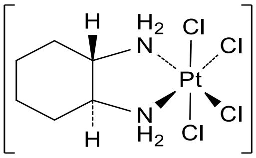
OC-6: octahedral complexes

Two coordination links are shown as the axis in the plane of the depiction and four are assumed to be in a plane perpendicular to the reference axis: two in front of and two behind the plane of the depiction.

Fig. 1.4.3.3-5

Representation of octahedral complexes; the example shown is *ormaplatin*



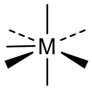


PBPY-7: pentagonal bipyramidal complexes

Two coordination links are shown as the axis in the plane of the depiction and the five other coordination links are shown as their projection onto the plane perpendicular to this axis: one in the plane of the depiction, two in front of it and two behind it.

Fig. 1.4.3.3-6

Representation of pentagonal bipyramidal complexes

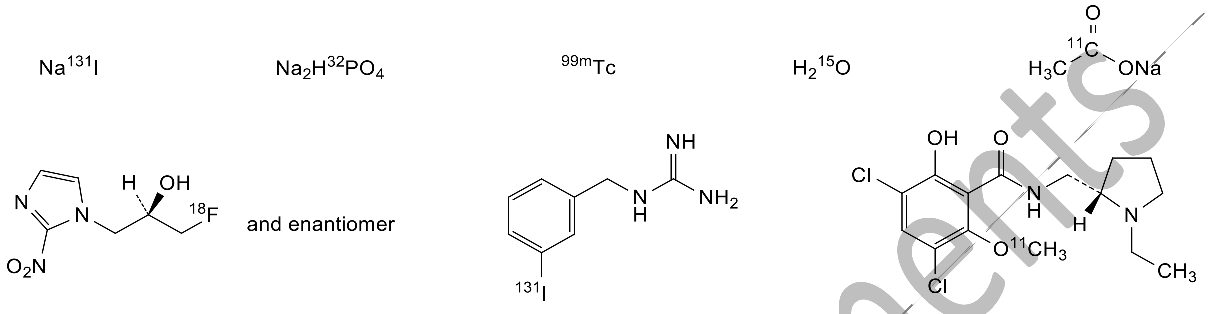


1.4.4. Isotopically modified compounds[8]

In an isotopically modified compound, the isotope used is indicated by its mass number placed as a superscript on the left of the symbol of the element concerned. Deuterium and tritium are respectively written ^{2}H (*not* D) and ^{3}H (*not* T). The carbon atom in a ring or a simplified angular-chain representation is explicitly designated when its mass number is shown.

Fig. 1.4.4-1

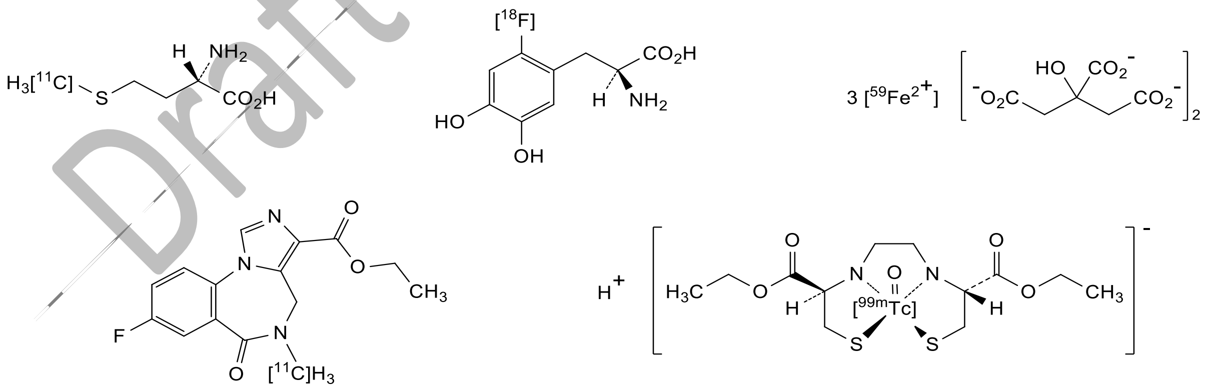
Examples of isotope-substituted compounds with (top, from left to right) ^{131}I-substituted *sodium iodide*, ^{32}P-substituted *sodium phosphate*, ^{99m}*technetium*, ^{15}O-substituted *water*, ^{11}C-substituted *sodium acetate* and (bottom, left to right) ^{18}F-substituted *analoque of misonidazole*, ^{131}I-substituted *iobenguane* and ^{11}C-substituted *raclopride*



When atomic symbols and formulas are drawn without square brackets, the compounds are assumed to be isotopically substituted (i.e. the atom concerned is completely replaced by the nuclide shown). To indicate isotopic labelling (partial replacement of the atom by the nuclide shown), atomic symbols in formulas should be enclosed in square brackets.

Fig. 1.4.4-2

Examples of isotope-labelled compounds with (top, from left to right) L-[^{11}C]*methionine*, [^{18}F]*fluorodopa*, [^{59}Fe]*ferrous citrate* and [^{11}C]*flumazenil* (bottom left) and [^{99m}Tc]*technetium bicisate* (bottom right)



1.4.5. Tautomeric compounds

Graphical depictions of compounds that may be subject to tautomeric rearrangements represent the predominant form as based on available structural information concerning the position of mobile hydrogens and double bonds.

Fig. 1.4.5-1

The preferred depiction (left) of *famotidine* corresponds to the predominant tautomer

Some exceptions may exist though, taking into consideration that compounds belonging to the same chemical class should be drawn in a similar way to make clear that they are structurally analogous.

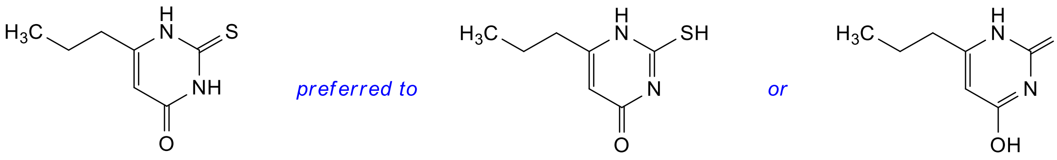
Fig. 1.4.5-2

The preferred depiction (left) of *sulfathiazole* shows analogy with the chemical class of sulfonamides but does not correspond to the predominant imine tautomer which is shown on the right

Compounds that can undergo tautomeric rearrangements are, in the absence of precise structural information, represented with a single structure which is usually selected among the tautomeric forms based on the seniority of both principal groups and substituents. Such compounds, mostly oxygen-, sulfur- and nitrogen-containing, are depicted with double bonds to oxygen and sulfur atoms rather than to nitrogen atoms.

Fig. 1.4.5-3

The preferred depiction (left) of *propylthiouracil* shows both oxo-and thioxo-groups which are respectively senior to hydroxy- and sulfanyl-groups



If criteria of seniority do not permit to select a tautomeric form, then locants may be considered, the lower being preferred.

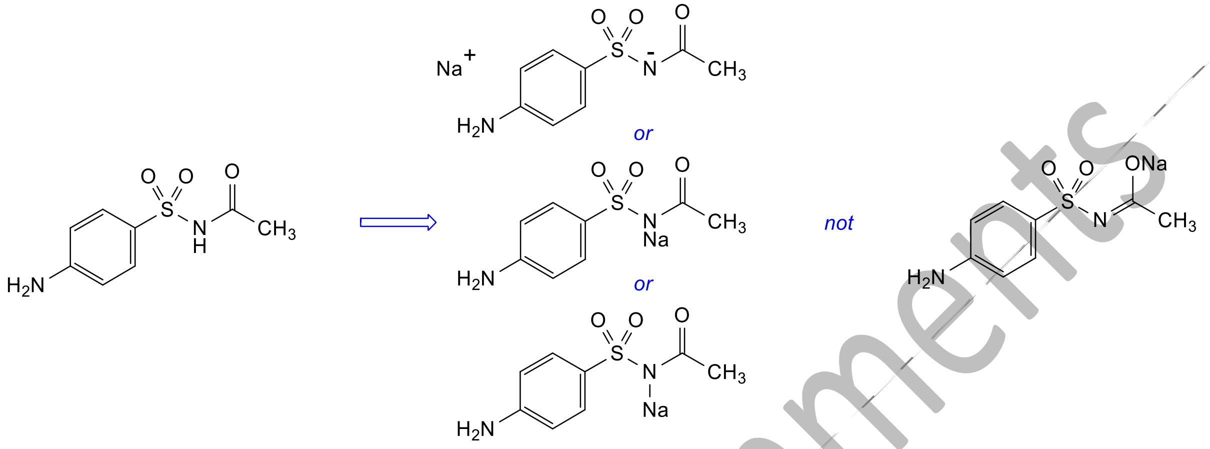
Fig. 1.4.5-4

The preferred depiction (left) of *omeprazole* has the lower locant for the indicated hydrogen of the benzimidazole moiety (1*H* instead of 3*H*)

Metal salts of nitrogen-containing acidic groups like amides are usually depicted as their preferred tautomeric neutral counterpart, in which case the mobile hydrogen is replaced with a negative charge or the metal atom.

Fig. 1.4.5-5

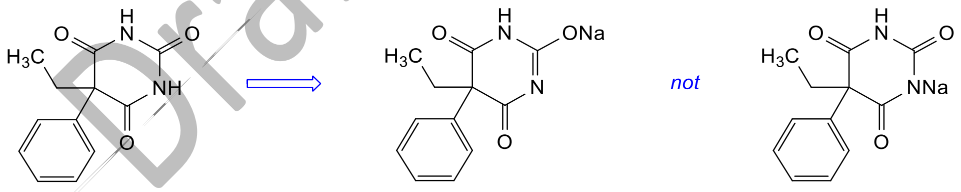
The preferred depiction (middle) of *sulfacetamide sodium* has the mobile hydrogen of the tautomeric preferred form of its neutral counterpart (left) replaced with a negative charge or metal atom



Sometimes structure characteristics may lead to a different representation. The keto-enol tautomerism in barbituric acid, for example, occurs preferably at the C-2 carbonyl due to the presence of two electronegative amido nitrogens flanking this group. As the enol (lactim) form is favoured in alkaline solution, a resulting salt is thus depicted with the metal being associated with the C-2 oxygen and not one of the adjacent nitrogens.

Fig. 1.4.5-6

The preferred depiction (middle) of *phenobarbital sodium* has the metal associated with the oxygen of the favoured enol form



1.5. Stereochemistry[4]

1.5.1. General aspects

In the absence of stereobonds, single plain bonds (—) surrounding either stereogenic or non-stereogenic atoms can represent bonds orientated in any direction, but double bonds(═), triple bonds(≡) and, with some rare exceptions, dative bonds(→) are usually considered to be in the plane of depiction of the atom concerned.

Stereobonds are mostly used to indicate configuration of stereogenic atoms in flattened drawings and, more rarely, in the case of perspective drawings to emphasize three-dimensional configuration of a molecule as a whole. If any stereobonds are attached to a stereogenic centre of tetrahedral configuration, then two of the additional normal bonds attached to this centre are supposed to be (approximatively) in the plane of depiction.

Hashed wedged bonds (┉) are used to represent a bond projecting behind the plane of depiction, while a filled wedge (◀) designates that the atom at the wide end of the wedge is above the atom at the narrow end. In both cases the narrow end of the wedge points at the stereocenter. Unknown configuration can be explicitly indicated with an asterisk (*) or, in the case of some natural products (*sections* 2.2. and 3.5.2.), by using a wavy bond (〰).

*[**Note from the Secretariat**. In particular, comments are sought on the proposal to use hashed wedges to denote a bond that projects behind the plane of representation. In structures shown in this guideline, dashed lines (-----) or hollow wedges (◁) are also used for this purpose. If comments are received supporting the proposal to use only hashed wedges to indicate a bond projecting behind the plane of representation, the chemical structures in this guideline that do not currently follow this proposal will be redrawn.]*

If configuration of ring atoms is to be indicated, then stereobonds should preferably be restricted to the acyclic substituents and not the ring skeleton. Furthermore, stereobonds should not be used between two stereogenic centres unless unavoidable.

When needed to define the stereochemical configuration, a hydrogen connected to a stereogenic atom is explicitly represented by its symbol "H".

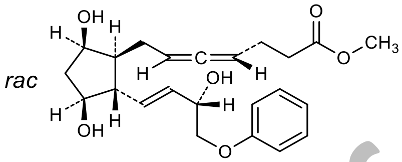
1.5.2. Cis- and trans-isomerism

1.5.2.1. (*EZ*) isomerism: carbon-carbon double bonds and imines

Unless constraints exist, such as in cyclic or congested structures, compounds containing double bonds with two sp^2-hybridized atoms are mostly depicted by lines at an angle of 120 ° to one another and hydrogens, if any, are omitted.

Fig. 1.5.2.1-1

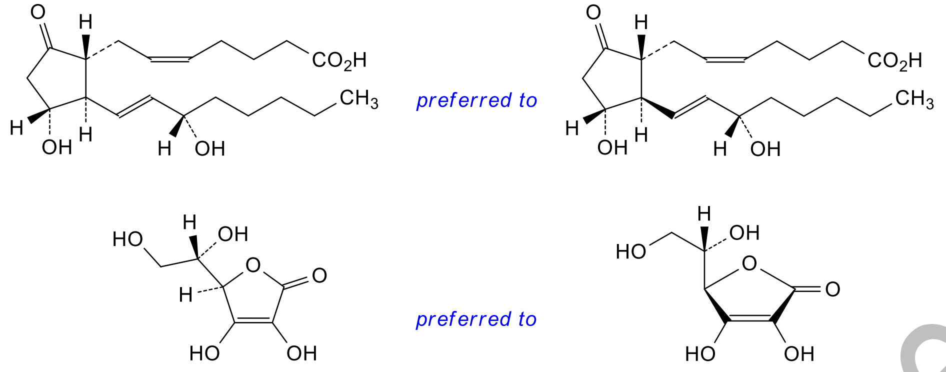
Structure of *fenprostalene* containing one isolated bond of known configuration and two consecutive double bonds with unknown axial chirality



For compounds containing double bonds, it is customary to draw the formula so that the double bond and atoms attached to it are lying in the plane of depiction with bonds designated with lines of normal thickness. Although stereobonds are sometimes used in the case of perspective drawings, they should usually be avoided whenever a double bond is linked to a stereogenic centre.

Fig. 1.5.2.1-2

Structure of *dinoprostone*(top) and *ascorbic acid* (bottom) illustrating customary depiction of double bonds

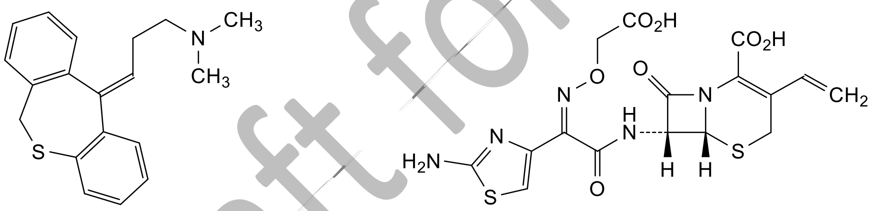


Specified configuration: (*E*)-isomers, (*Z*)-isomers and mixtures of (*E*)- and (*Z*)-isomers

Isomers are shown with the two sequence-rule-preferred atoms or groups (each attached to one atom of the double bond) placed on the same side (*cis*) of the double bond for the (*Z*)-isomer and on the opposite side (*trans*) of this plane for the (*E*)-isomer.

Fig. 1.5.2.1.1-1

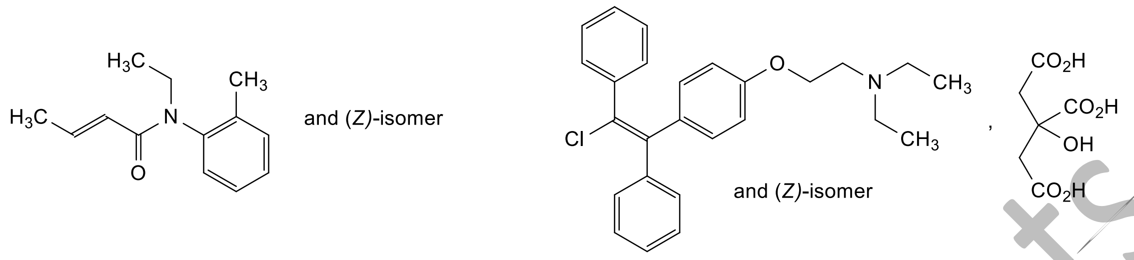
Examples of (*E*)-isomer (left) and (*Z*)-isomer (right) for *dosulepin* and *cefixime,* respectively



Mixtures of (*E*)- and (*Z*)-isomers of a compound are usually represented with depiction of the (*E*)-isomer in combination with the legend “and the (*Z*)-isomer”. This representation does not necessarily mean that the two isomers are present in the same quantities.

Fig. 1.5.2.1.1-2

Examples of mixtures of (*E*)- and (*Z*)-isomers, the compounds shown are *crotamiton* (left) and *clomifene citrate* (right)

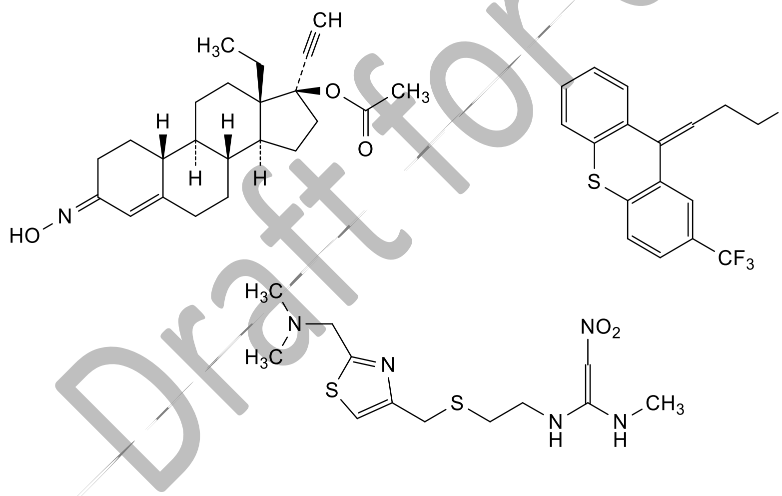


Unspecified configuration: (*E*)-and/or (*Z*)-isomers

In cases where the double bond has a single substituent on one or both ends, unspecified configuration is preferably depicted with one of those substituents aligned with the double bond.

Fig. 1.5.2.1.2-1

Examples of compounds with unspecified geometric configuration with *norgestimate* (top left), *flupentixol* (top right) and *nizatidine* (bottom)



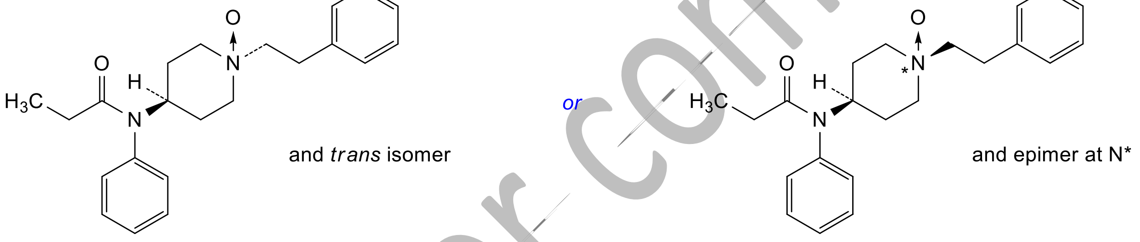
The information conveyed by this representation is considerably different from that conveyed by the representation of a mixture of both isomers in that it might concern either the (*E*)-isomer, the (*Z*)-isomer, or a mixture of both.

1.5.2.2. *Cis*- and *trans*-isomerism of rings

Graphical representations of individual isomers, as well as mixtures of *cis-trans* ring isomers, follow standard conventions of this guide and are detailed in *section 1.5.4.2.* Mixtures of *cis-trans* ring isomers, however, are sometimes depicted as the *cis*-isomer in combination with the legend 'and *trans*-isomer' added to the right of the structure. This type of representation does not necessarily mean that the two isomers are present in the same quantities.

Fig. 1.5.2.2-1

Mixtures of *cis* and *trans* ring isomers are sometimes represented as the *cis*-isomer in combination with the legend "and *trans*-isomer" as illustrated with *fentanyl N-oxide*. Standard conventions, detailed in *section 1.5.4.2.*, lead to the representation on the right and are preferred for being consistent with other rules for graphical representation.



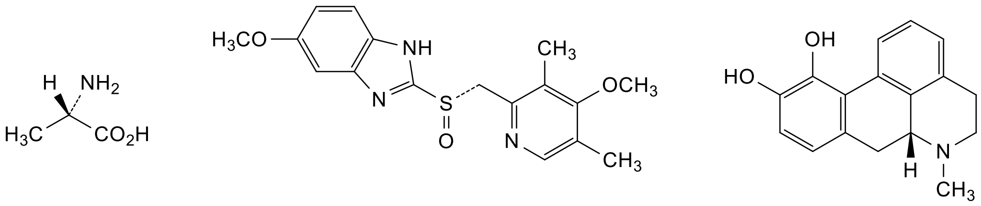
1.5.3. Compounds with one centre of chirality

1.5.3.1. Specified configuration

General conventions are applicable in order to represent stereochemistry of either acyclic or cyclic compounds with one centre of chirality. Whereas, in most cases, configuration of the stereogenic centre is indicated with both a dashed and a filled wedge, in some cases, only a single stereo bond is used, such as in structural diagrams of chiral sulfoxides and chiral fusion atoms. In the absence of a legend, a structural diagram represents the single specific molecular entity drawn.

Fig. 1.5.3.1-1

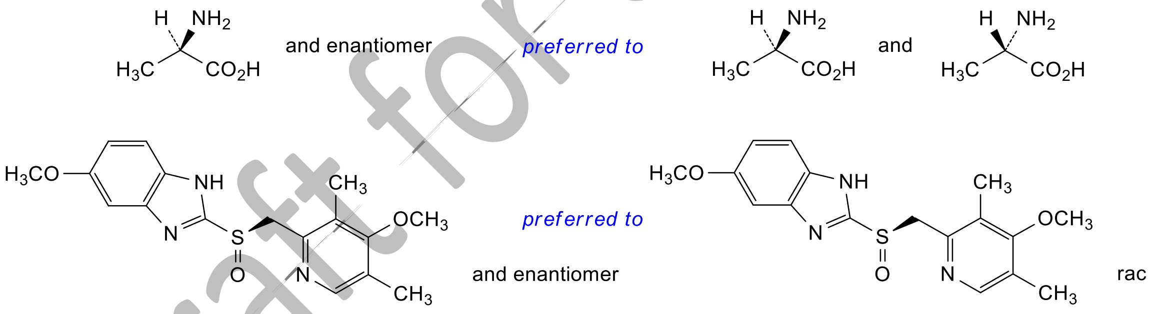
A structural diagram represents the single specific molecular entity drawn, the optical pure isomers shown are L-*alanine* (left), *esomeprazole* (middle) and *apomorphine* (right)



A mixture of enantiomers is usually represented with the sole (*R*)-isomer followed by the legend "and enantiomer" rather than depicting the isomers side by side and joined with the legend "and". Although this type of representation may represent a racemic mixture, it does not necessarily mean that the two isomers are always present in the same quantities. The more restrictive and language-neutral legend "rac" is not considered clear enough and is rarely used instead of "and enantiomer".

Fig. 1.5.3.1-2

A racemic mixture of enantiomers is preferably shown with the (*R*)-isomer in combination with the legend "and enantiomer" as illustrated with DL-*alanine* (top) and *omeprazole*(bottom)

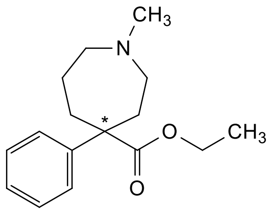


1.5.3.2. Unspecified configuration

If the chirality of the center of asymmetry is unknown or not specified, the bonds adjoining atoms or groups to the chiral atom are shown as lines of 'normal' thickness and an asterisk is used to identify this chiral centre. In the absence of a legend, chiral centres flagged with an asterisk might concern either the (*R*)-isomer, the (*S*)-isomer, or a mixture of both.

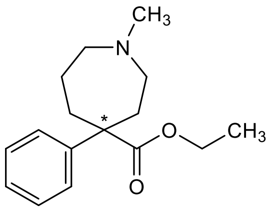
Fig. 1.5.3.2-1

Unspecified configuration is indicated with an asterisk; the example shown is *methethoheptazine*



Isomer(s) present:

(R) or (S)
or
(R) and (S)



Isomer(s) present:

pure (R) or pure (S)
or undefined mixture
of (R) and (S)

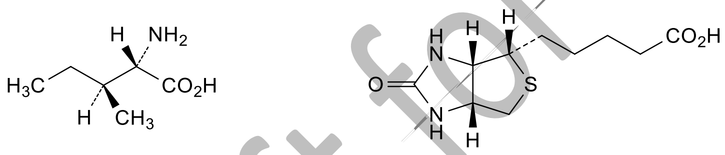
Asterisks are positioned so that they clearly identify the chiral centre which concerns either an implicit carbon atom or a heteroatom label.

1.5.4. Compounds with several centres of chirality

In compounds containing several centres of chirality, the same conventions as mentioned above apply to each of these centres.

Fig. 1.5.4-1

Structural diagrams depicted for compounds containing several chiral centres as illustrated with *isoleucine* (left) and *biotin* (right) representing the single specific molecular entity drawn



1.5.4.1. Mixtures of two diastereoisomers

A mixture of two diastereoisomers that are enantiomers are usually represented with the (*R*)-isomer at the position with lowest numerical locant and followed by the legend “and enantiomer”.

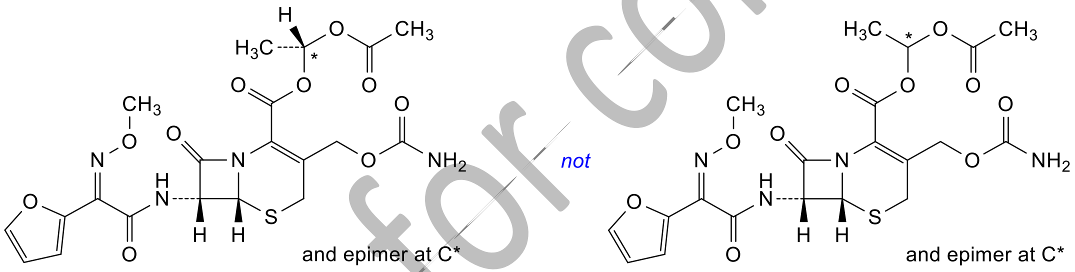
Fig. 1.5.4.1-1

Structural diagram of diastereoisomers that are enantiomers as illustrated with *phenylpropanolamine* which is preferably depicted with (1*R*,2*S*) configuration and not (1*S*,2*R*)

A mixture of two diastereoisomers that are epimers with the opposite configuration at only one of two or more tetrahedral stereogenic centers, is usually represented with the sole (*R*)-isomer with respect to the centre of asymmetry of interest. The epimeric center is marked with an asterisk and the structure is followed by the legend "and epimer at C*". If different from carbon, then legends are adequately adjusted for any heteroatoms at the epimeric centre such as in "and epimer at S*" or "and epimer at N*".

Fig. 1.5.4.1-2

Representation of a mixture of two diastereomers as illustrated with *cefuroxime axetil*

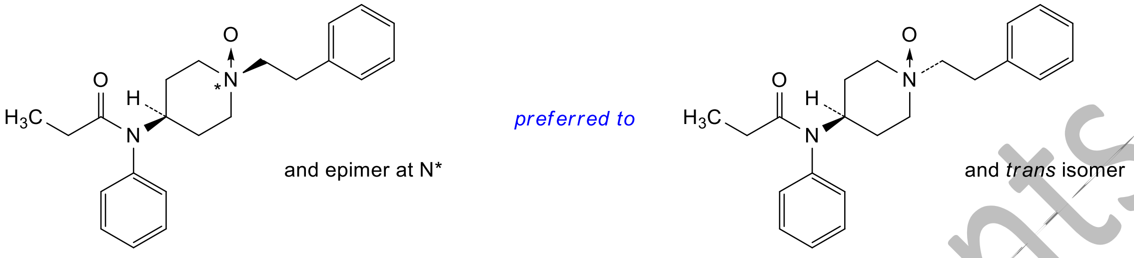


1.5.4.2. *Cis*- and *trans*-isomerism of rings

By analogy with the above, graphical representations of mixtures of *cis-trans* ring isomers may be depicted with the isomer displaying (*R*)-configuration at the position with the lowest numerical locant and followed by the legend "and epimer at C*". This type of representation does not necessarily mean that the two isomers are present in the same quantities.

Fig. 1.5.4.2-1

Standard representation of mixtures of *cis* and *trans* ring isomers are represented by flagging the epimeric centre in combination with a legend and are preferred to representation of the *cis* isomer in combination with the legend "and *trans* isomer", as illustrated with *fentanyl-N-oxide*



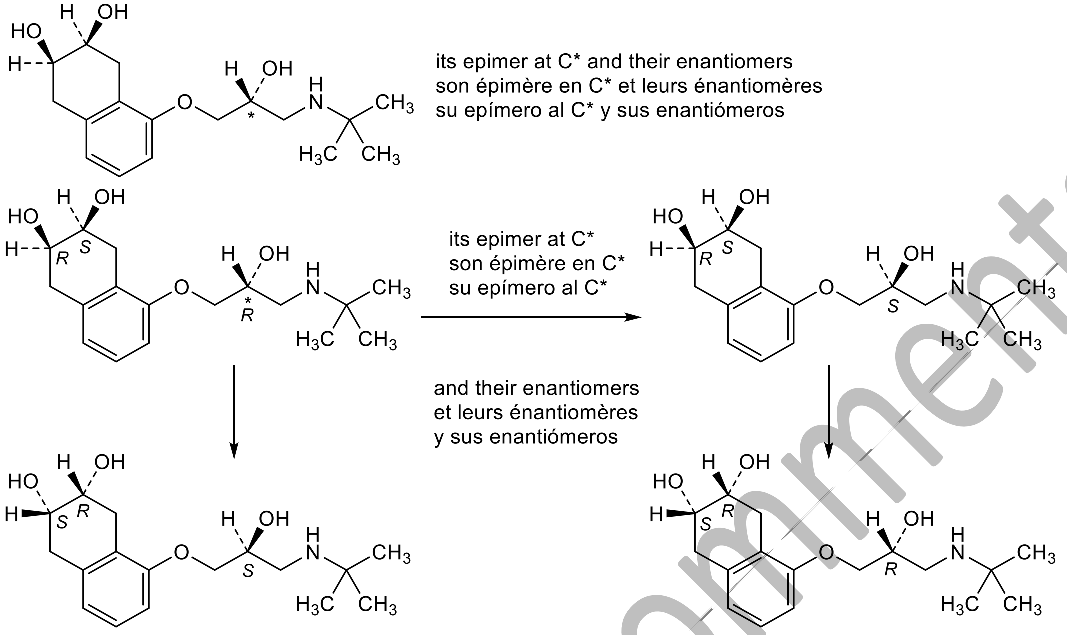
1.5.4.3. Mixtures of more than two diastereoisomers

A limited mixture of more than two diastereoisomersis is depicted as a subset of enantiomers related to a set of corresponding epimers. All centres of asymmetry that permutate together are grouped and linked to a marginal centre of asymmetry which is marked with an asterisk as the "epimeric" centre.

The "epimeric" centre is shown as the (*R*)-isomer and the legend 'and epimer at C*' is added to the structure, directly followed by the mention 'and their enantiomers'.

Fig. 1.5.4.3-1

Representation of a limited mixture of more than two diastereoisomers (top) broken down for more clarity (bottom); the example shown is *nadolol*



its epimer at C* and their enantiomers
son épimère en C* et leurs énantiomères
su epímero al C* y sus enantiómeros

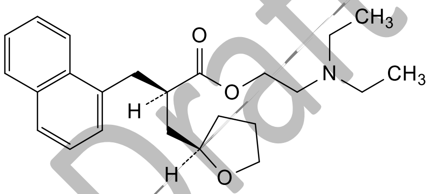
its epimer at C*
son épimère en C*
su epímero al C*

and their enantiomers
et leurs énantiomères
y sus enantiómeros

In case of a mixture containing every possible diastereoisomer, all stereogenic centres are represented with the (*R*)-configuration and the structure is followed with the legend "and stereoisomers".

Fig. 1.5.4.3-2

Representation of an unlimited mixture of diastereoisomers; the example shown is *naftidrofuryl*



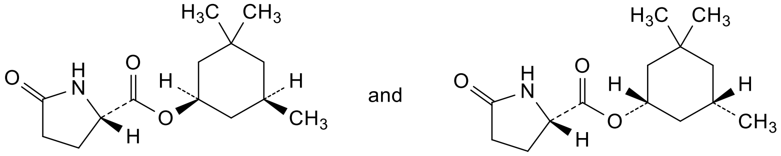
and stereoisomers

1.5.4.4. Fully depicted mixtures of stereoisomers

In more complicated cases, it is better to draw each component of the mixture to show all the peculiarities of the structure.

Fig. 1.5.4.4-1

Specification of configuration by drawing all the components of the mixture; the example shown is *crilvastatin*

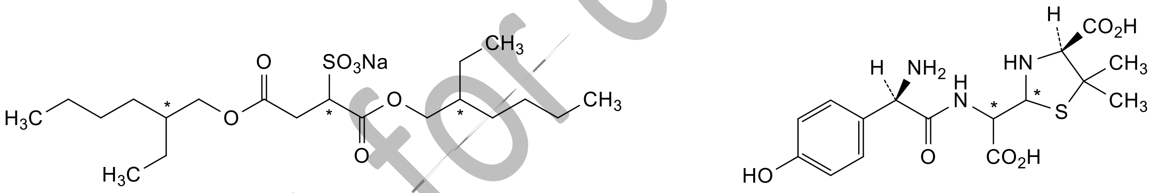


1.5.4.5. Unspecified configuration

If compounds contain several centres of asymmetry of unknown or unspecified chirality, then asterisks are used to identify all these centres. Centres flagged with an asterisk concern either (*R*)-configuration, (*S*)-configuration, or both.

Fig. 1.5.4.5-1

Unspecified configuration is indicated with an asterisk; the examples shown are *docusate sodium* (left) and an impurity of *amoxicillin* (right)

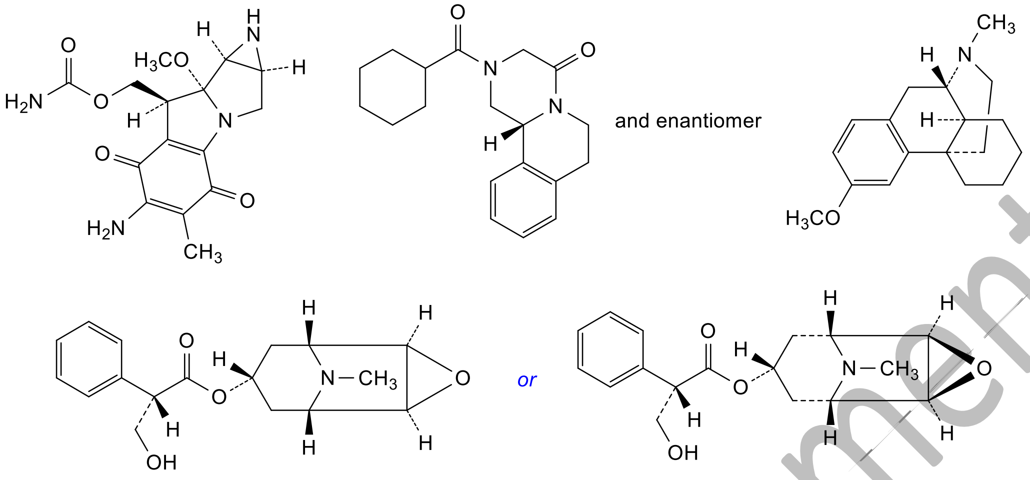


1.5.4.6. Distinctive aspects of stereogenic centres at fusion atoms and at bridgehead atoms

Substituents are linked with stereogenic fusion and bridgehead atoms through filled wedges or dashed bonds to indicate the stereochemistry resulting from the way the cycles are connected. Although configuration is mostly shown with just one stereobond associated with three normal bonds, sometimes a second stereobond is used for more clarity by replacing a normal bond of, preferably, the smaller ring.

Fig. 1.5.4.6-1

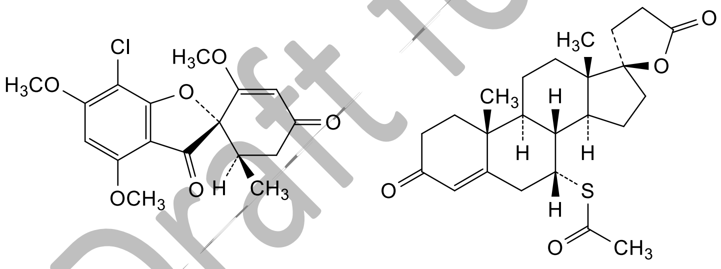
Stereogenic centres at fusion atoms and at bridge head atoms; the examples shown are *mitomycin* (top left), *praziquantel* (top middle), *dextromethorphan* (top right) and *hyoscine* (bottom)



Wherever possible, stereogenic spiro fusion atoms are displayed with two stereobonds attached to them, unless an adjacent stereogenic centre is present, in which case only one stereobond is used. Any stereobonds depicted are preferably associated with the smaller ring.

Fig. 1.5.4.6-2

Stereogenic spiro fusion atoms of *griseofulvin* (left) and *spironolactone* (right)



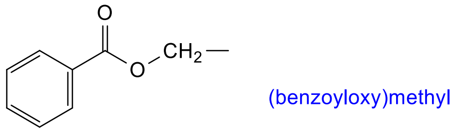
1.6. Unbound substituents or radicals

1.6.1. Attachment points

Unbound substituents or radicals are specified by drawing the structural fragment and emphasizing the location of attachment with an explicitly specified atom label for the carbon- or heteroatom bearing the free valence. Any hydrogen atoms are shown right next to the explicitly specified atoms to which they are attached. The free valence is depicted with a single bond displaying a label on one side of the bond and none on the other.

Fig. 1.6.1-1

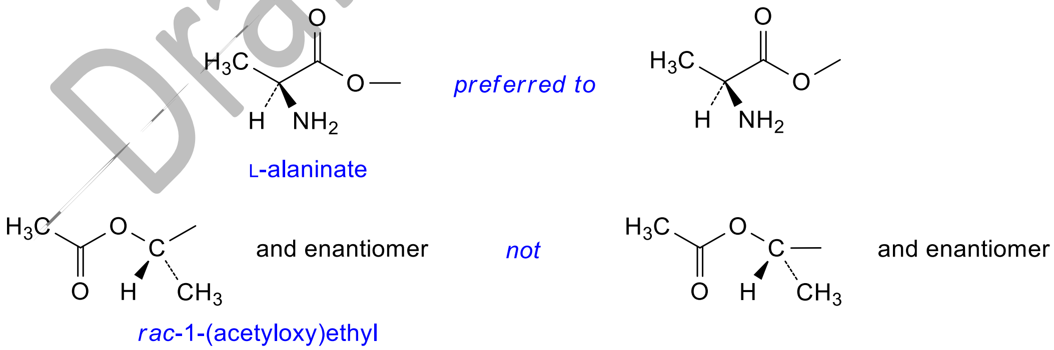
Depiction of the radical *bezomil* with an explicitly specified atom label for emphasizing the location of the free valence



For linear structures, the bond indicative for the free valence is usually oriented horizontally unless the connecting atom is congested with substituents.

Fig. 1.6.1-2

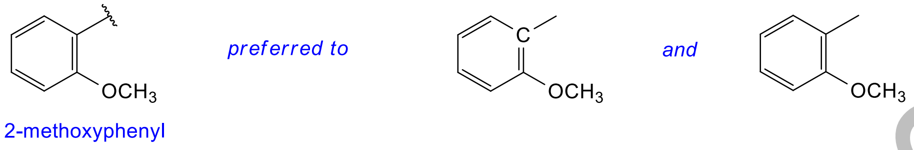
Depiction of the radicals *alaninate* (top) and *axetil* (bottom)



In case of cyclic radicals, it might be desirable for aesthetic reasons *not* to specify the carbon atom bearing the free valence which is instead indicated as a cut bond with a clearly distinguishable symbol such as a perpendicular wavy line (——⌇).

Fig. 1.6.1-3

Depiction of the radical *guacil* for which the free valence is preferably emphasized with a cut bond instead of an explicitly specified atom label

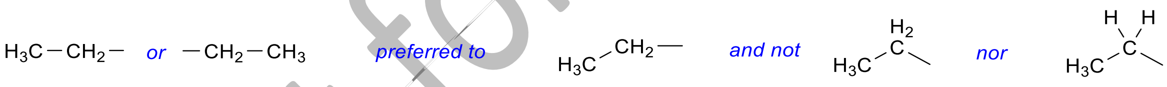


1.6.2. Orientation

Wherever possible, structures are set out horizontally with the free valence either on the left or on the right.

Fig. 1.6.2-1

Depiction of the radical *ethyl* which is preferably drawn without a bend

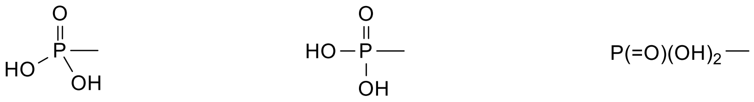


1.6.3. Depiction style

Line formulas, if concise and clear enough to describe substituent groups, are sometimes used for radicals which are intended as annotations to generic structures.

Fig. 1.6.3-1

The radical *phospono* as depicted with different drawing styles; conventional (left), as expanded line formula (middle) and as condensed line formula (right)

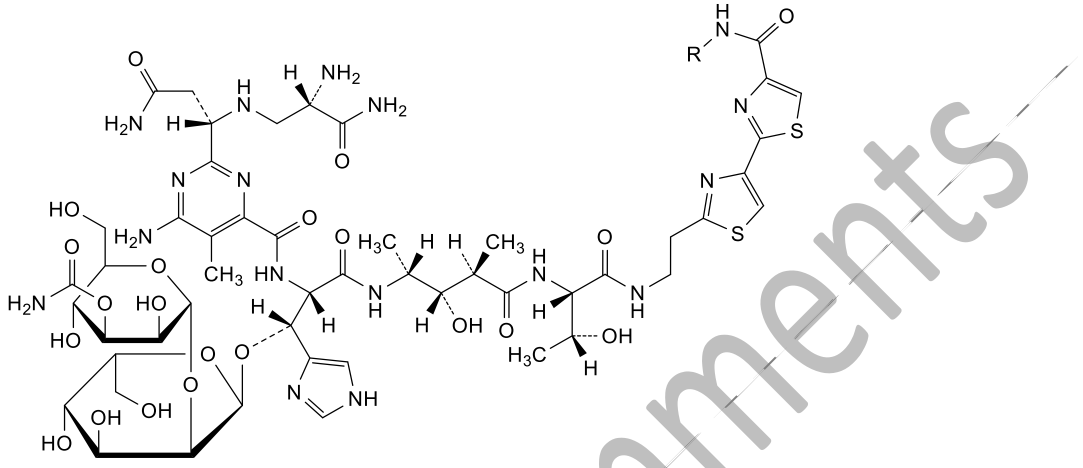


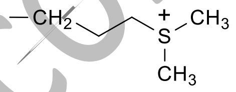
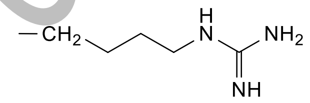
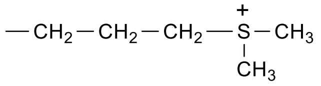
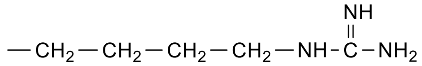
If an expanded linear drawing style is used, then all acyclic chain atoms are depicted explicitly and are connected through horizontal and vertical bonds exclusively. Polyatomic labels show hydrogen atoms right next to the atoms to which they are attached if not drawn with explicit bonds.

For a condensed linear drawing style, either no bonds or only horizontal bonds are shown. The order of the atoms suggests the connectivity and parentheses are used for groups not part of the main chain. Square brackets and associated subscripts may be used for repeating units.

Fig. 1.6.3-2

Explanatory annotation using different drawing styles associated to the generic structure of *bleomycin* (top). Among the different styles, the condensed line formula with horizontal bonds only represents a concise and clear description of the different fragments characteristic for bleomycin A₂ and B₂.



drawing style :	*annotation :*
conventional	bleomycin A$_2$ —R =  bleomycin B$_2$ —R = 
expanded line formula	bleomycin A$_2$ —R =  $-CH_2-CH_2-CH_2-\overset{+}{S}(-CH_3)-CH_3$ bleomycin B$_2$ —R =  $-CH_2-CH_2-CH_2-CH_2-NH-\overset{NH}{\overset{\Vert}{C}}-NH_2$
condensed line formula with horizontal bonds only (preferred)	bleomycin A$_2$ $-R = -[CH_2]_3-S^+(CH_3)_2$ bleomycin B$_2$ $-R = -[CH_2]_4-NH-C(=NH)-NH_2$ *or* bleomycin A$_2$ $R = [CH_2]_3\text{-}S^+(CH_3)_2$ bleomycin B$_2$ $R = [CH_2]_4\text{-}NH\text{-}C(=NH)\text{-}NH_2$
condensed line formula	bleomycin A$_2$ $R = CH_2CH_2CH_2S^+(CH_3)_2$ bleomycin B$_2$ $R = CH_2CH_2CH_2CH_2NHC(NH)NH_2$

1.7. Generic structures

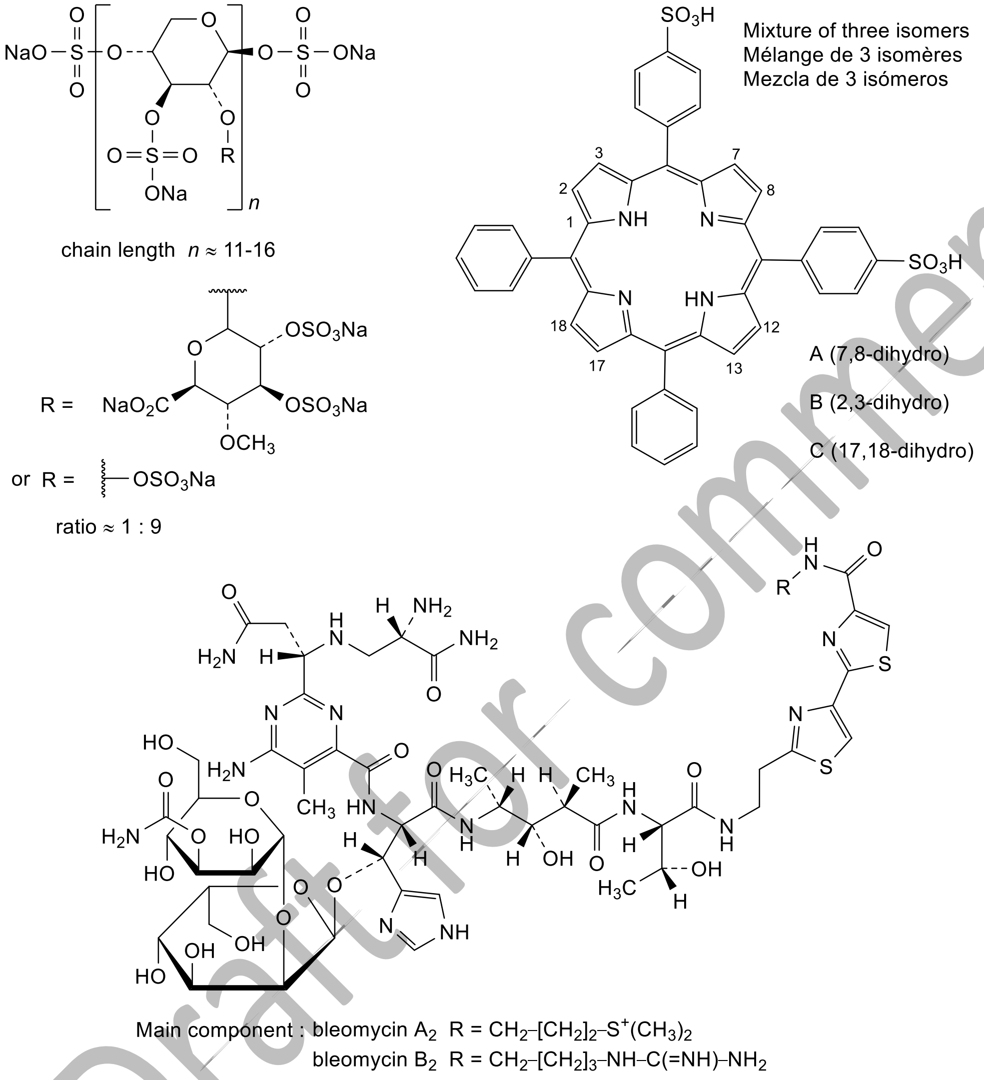
Although it is the policy of the INN programme *not* to assign names for mixtures of substances, some exceptions exist that concern either polymer, several substances in fixed proportions and natural products for which mixtures consist of structurally similar compounds that display comparable activities that are impossible to separate.

1.7.1. General aspects

Mixtures of substances are usually depicted with a generic structure shown along with information on the centres of diversity using labels or values of repeating units.

Fig. 1.7.1-1

Depictions of the generic structures of the polymer *pentosane polysulfate sodium* (top left), *fimaporfin*, a 25%, 50%, 25% mixture of three isomers A, B, C (top right) and natural product *bleomycin* (bottom)



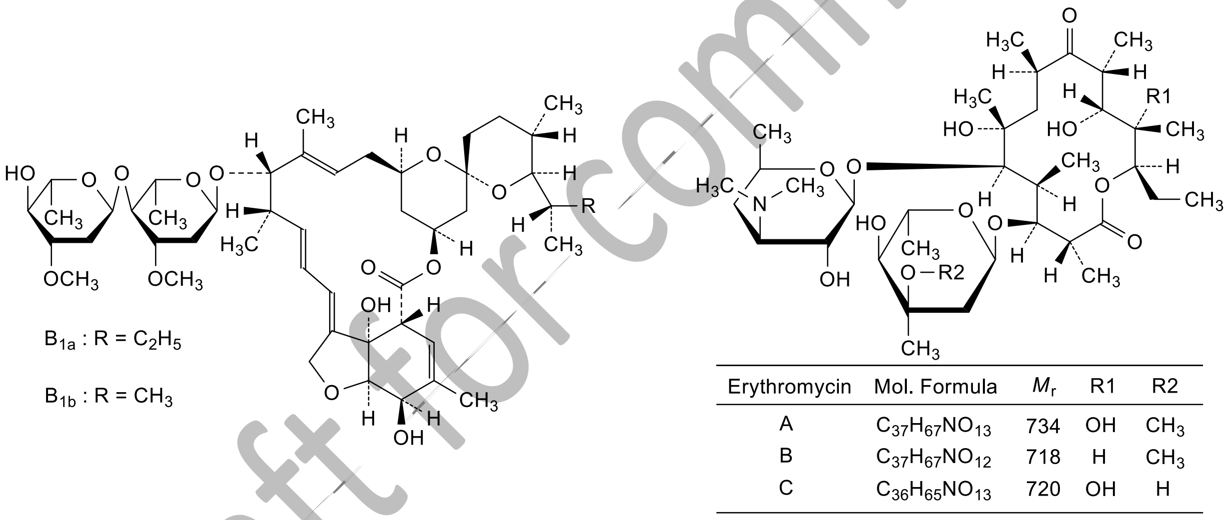
Generic structures should generally be kept simple with the variable substituent groups of the different compounds making up the mixture as small as possible. Any information shown along with a generic structure for specifying diversity should be self-explanatory and may be depicted with drawing styles (mostly linear condensed formulas) that differ from the general rules described in this paper. Compounds that make up a mixture are usually specified in tabular form with the main compound mentioned first and all others classified in order of decreasing content, along with radicals and other structural information.

1.7.2. Symbols for substituent groups and repeating units

A single centre of diversity for a given generic structure is designated R. Two centres of diversity are usually designated R and R, although they may also be designated R1 and R2, in particular in crowded structures where the prime symbol might not be noticed. When more than two centres of diversity are present, they are designated R1, R2, R3...R*n*. Although the number assigned to a label is independent of its position in the structure, in simple cases labels are preferably positioned to follow the order of numbering of the skeletal atoms, mostly in a clockwise direction for cyclic compounds or from right to left for linear ones.

Fig. 1.7.2-1

Depictions of the generic structures of *ivermectin* (left) and *erythromycin* (right) with respectively one and two centres of diversity as symbolized with different type of labels

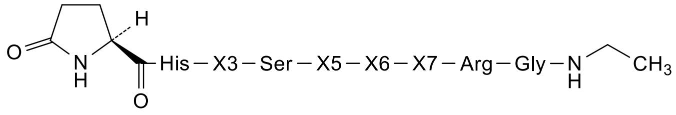


Erythromycin	Mol. Formula	M_r	R1	R2
A	$C_{37}H_{67}NO_{13}$	734	OH	CH_3
B	$C_{37}H_{67}NO_{12}$	718	H	CH_3
C	$C_{36}H_{65}NO_{13}$	720	OH	H

Specific symbols, X and Y are used in peptides (*section 2.1.*) to designate up to two varying amino acid residues. Symbols like X1, X2, X3, and so on, are used to represent three residues and beyond with the numerical indexation corresponding to the residue number.

Fig. 1.7.2-2

Leuprorelin impurities with the varying residues symbolized by "X_n_"-labels

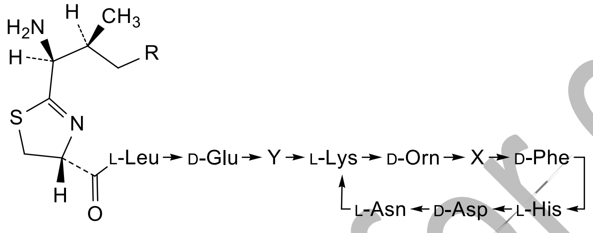


C : X3 = Trp, X5 = Tyr, X6 = X7 = Leu
E : X3 = D-Trp, X5 = Tyr, X6 = D-Leu, X7 = Leu
G : X3 = Trp, X5 = D-Tyr, X6 = D-Leu, X7 = Leu
H : X3 = Trp, X5 = Tyr, X6 = X7 = D-Leu

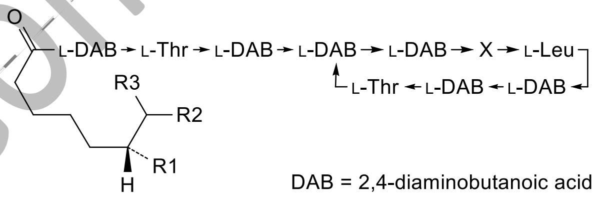
Sometimes a combination of different labels is used depending on the structural features and complexity of the mixture.

Fig. 1.7.2-3

Depiction of the generic structure of _bacitracin_ (left) and _colistin_ (right) as symbolized with a combination of different types of labels



Name	Mol. Formula	X	Y	R
Bacitracin A	$C_{66}H_{103}N_{17}O_{16}S$	L-Ile	L-Ile	CH_3
Bacitracin B1	$C_{65}H_{101}N_{17}O_{16}S$	L-Ile	L-Ile	H
Bacitracin B2	$C_{65}H_{101}N_{17}O_{16}S$	L-Val	L-Ile	CH_3
Bacitracin B3	$C_{65}H_{101}N_{17}O_{16}S$	L-Ile	L-Val	CH_3



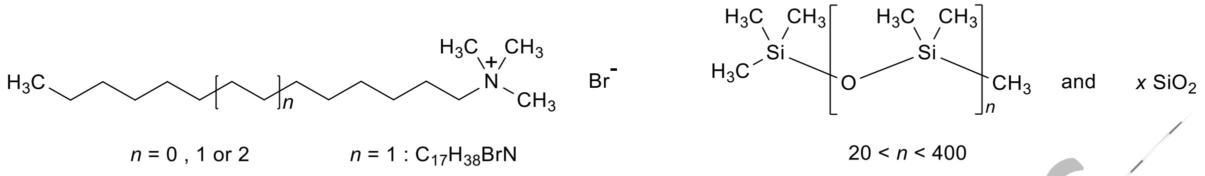
DAB = 2,4-diaminobutanoic acid

polymyxin	X	R1	R2	R3	Mol. Formula	M_r
E1	D-Leu	CH_3	CH_3	H	$C_{53}H_{100}N_{16}O_{13}$	1170
E2	D-Leu	CH_3	H	H	$C_{52}H_{98}N_{16}O_{13}$	1155
E3	D-Leu	H	CH_3	H	$C_{52}H_{98}N_{16}O_{13}$	1155
E1-I	D-Ile	CH_3	CH_3	H	$C_{53}H_{100}N_{16}O_{13}$	1170
E1-7MOA	D-Leu	H	CH_3	CH_3	$C_{53}H_{100}N_{16}O_{13}$	1170

In some cases, such as polymers, diversity can be indicated using variables for repeating units. These repeating units are placed in square brackets and their number is indicated with a variable *n* as a subscript on the right. When more than two repeating units are defined, then the subscript letters used to denote multiplicities of polymeric sequences vary in function of the type of polymeric fragment they are related to, see *section 2.4*.

Fig. 1.7.2-4

Depiction of the generic structures of *cetrimide* (left) and *simethicone* (right), a mixture of α-(trimethylsilyl)-ω-methylpoly[oxy(dimethylsilylene)] and silicondioxide

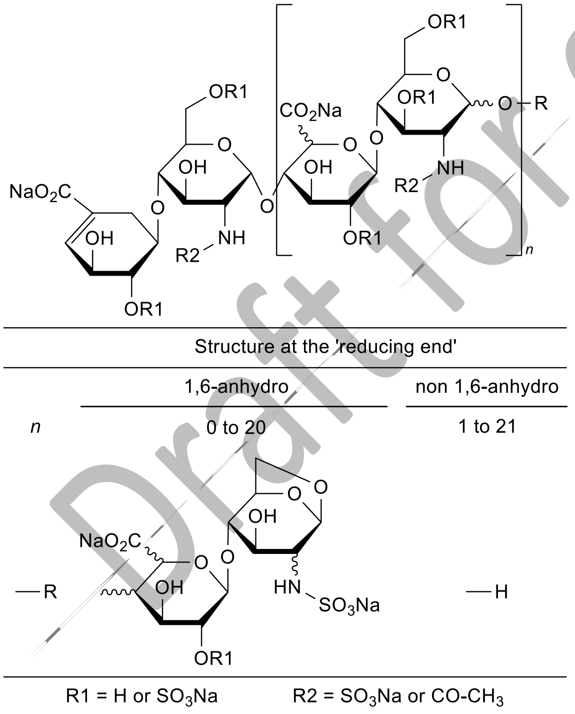


1.7.3. Mixtures of complicated structures

For certain mixtures of structures (in particular heparins), the complexity is such that it is preferable to describe relevant structural information rather than depicting it.

Fig. 1.7.3-1

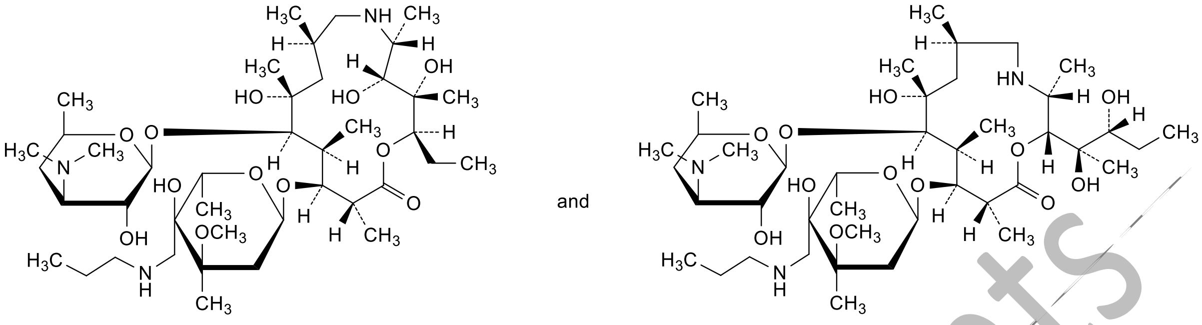
Depiction of the rather complicated generic structure of *enoxaparin sodium*



Compounds that cannot be represented with a single generic structure are drawn side by side and shown with a + (plus) sign or preferably with "and" in between.

Fig. 1.7.3-2

Depiction of *tulathromycin*, a mixture of two isomers, tulathromycin A (left) and tulathromycin B (right)

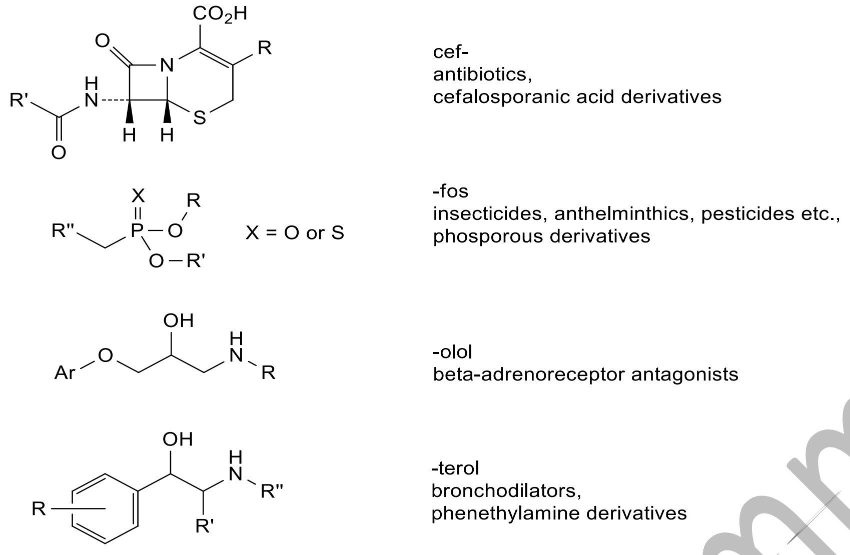


1.7.4. Structural differences versus common features

A particular application of generic structures in the INN programme concerns formulas which are intended to represent a group of pharmaceutical substances in accordance with their stem[9] definition. In this particular case, the generic structure often covers a large number of compounds and is used for emphasizing common structural features of compounds belonging to the same class and not for giving details on individual compounds. As such, there are fewer constraints on depicting molecular formulas and a larger extent of labels and symbols for defining atoms, groups and structural fragments may be used.

Fig. 1.7.4-1

Depiction of generic structures used for stem definitions, the examples shown are *cef-*, *-fos*, *-olol* and -*terol* (from top to bottom)



cef-
antibiotics,
cefalosporanic acid derivatives

-fos
insecticides, anthelminthics, pesticides etc.,
phosporous derivatives

-olol
beta-adrenoreceptor antagonists

-terol
bronchodilators,
phenethylamine derivatives

II. Main structural classes

2.1. Structure representation of amino acids and peptides[10]

2.1.1. General aspects

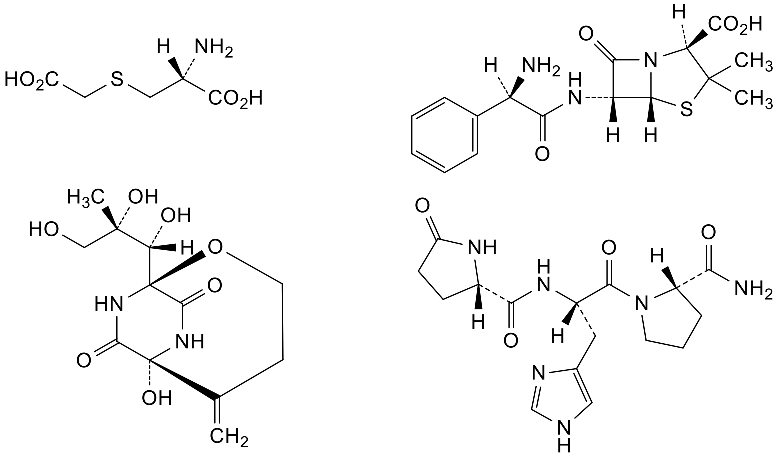
Amino acids are represented either with a full depiction or, in the case of alignments of amino acid residues, with the corresponding 3-letter or 1-letter codes. Stereochemical features and structural modifications of residues are subsequently indicated in different ways. All linear sequences of amino acid residues are shown with the amino-terminal residue on the left and the carboxy-terminal residue on the right with residue numbering beginning at the N-terminus.

2.1.2. Full depiction

Although the species made out of one or several amino acids generally possess amphoteric properties, amino acids, whether single or in condensed form, are preferably shown as neutral compounds with respect to the recommendations described under General Issues (*section 1.4.2.5.*). Single residues and small peptide derivatives, whether linear or cyclic, are either defined with systematic or semi-systematic names and are always depicted in full in line with standard conventions.

Fig. 2.1.2-1

Small amino acid derivatives are represented in full. The examples shown are *carbocisteine* (top left), *ampicillin* (top right), *bicozamycin* (bottom left) and *protirelin* (bottom right)

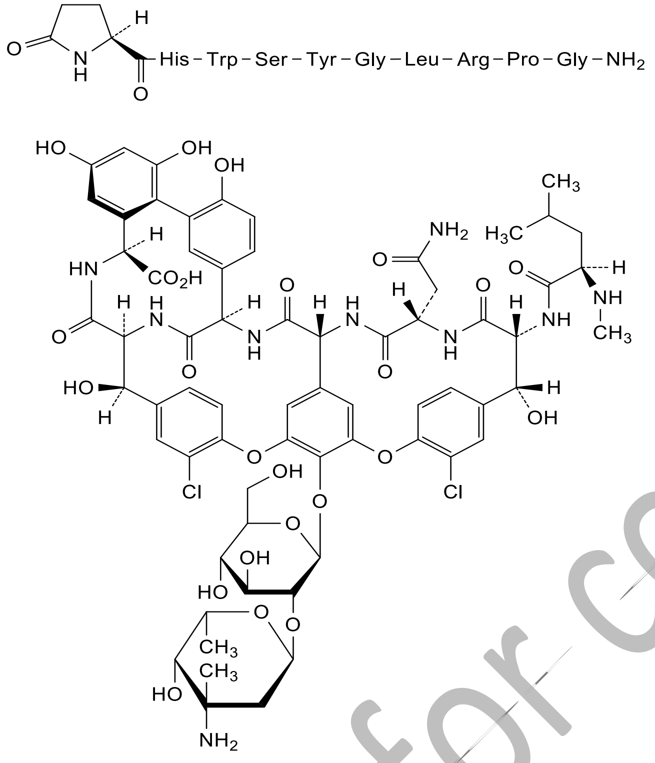


2.1.3. Standard 3-letter code

Oligopeptides and derivatives consisting of four residues and beyond are, on the other hand, mostly defined with semi-systematic names and are represented with the standard 3-letter code unless structures or structural fragments necessitate full depiction for clarity.

Fig. 2.1.3-1

Oligopeptides are generally depicted with the standard 3-letter code but, in case of derivatives with complicated structures, they are represented in full. The examples shown are *gonadorelin* (top) and *vancomycin* (bottom).



The use of the standard 3-letter code for the representation of amino acids in oligo- and polypeptides implies some characteristics as to configurational specifications and structural modifications of residues.

2.1.3.1. Stereochemical features

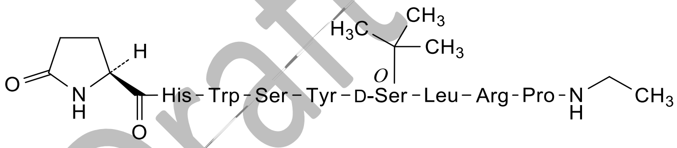
For stereochemical designation, prefixes like ”L”, ”D”, “DL” or “*ambo*” precede the 3-letter code and are joined to it by a hyphen. The prefix “ξ” (greek xi) may be used to indicate unknown configuration. The configuration of amino acid residues in polypeptides is assumed to be “L” when not indicated otherwise.

2.1.3.2. Amino acid modifications

A hyphen on the left, or either above or below the amino acid symbol, generally signifies removal of a hydrogen atom on respectively the amine group and the side chain, as well as representing the bond formed by the residue thus produced. In some cases where carboxyl groups are present in side chains, hyphens above or below may also represent the removal of a hydroxyl group, such as hyphens on the right of a residue symbol representing loss of the α-carboxylic OH group. If an amino acid residue is substituted on the N-2 atom, the symbol for the substituent is placed directly *before* the three-letter code without hyphen, whereas fully depicted substituents at the N-terminal amino group are separated with a hyphen. Modifications of the C-terminal carboxyl group are indicated with fully depicted substituents *after* the three letter-code, again separated with a hyphen. If a side-chain modification occurs, the substituent may be depicted either in full or by means of its symbol placed above or below the 3-letter code and joined to it by a vertical line centered on the middle letter. If necessary, a locant is placed beside the vertical line that represents side-chain substitution. Substituents are generally depicted in full wherever possible, especially at N- and C-termini, but symbols like “Ac” “Me”, “Et” and “Bzl” are sometimes used instead for saving place. Details on symbols for substituents can be found in the IUPAC publication on nomenclature and symbolism for amino acids and peptides.

Fig. 2.1.3.2-1

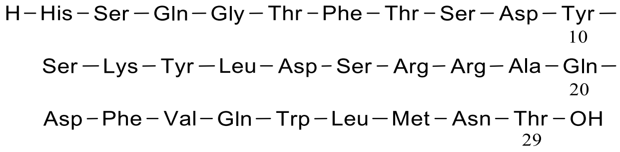
Structure of *buserelin* illustrating configurational specification and structural modification of residues represented with the standard 3-letter code



If it is desired to emphasize lack of substitution, H or OH may be added to the hyphen or vertical line that represents removal of one of these groups. If the polypeptide occupies more than one line, then a hyphen is placed at the end of each successive line until the formula has been completed. Numbers may be indicated underneath the amino acid symbols generally every tenth residue and at the end of the last line.

Fig. 2.1.3.2-2

Alignment of *human glucagon* illustrating lack of substitution at N- and C-termini through addition of H and OH at terminal residues

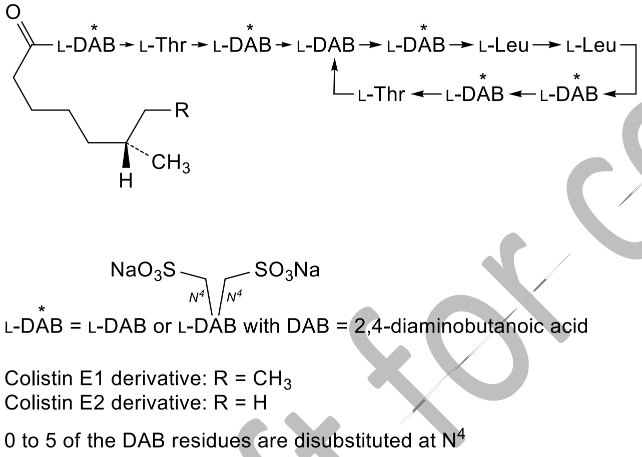


$$\begin{array}{l}\text{H}-\text{His}-\text{Ser}-\text{Gln}-\text{Gly}-\text{Thr}-\text{Phe}-\text{Thr}-\text{Ser}-\text{Asp}-\underset{10}{\text{Tyr}}-\\ \text{Ser}-\text{Lys}-\text{Tyr}-\text{Leu}-\text{Asp}-\text{Ser}-\text{Arg}-\text{Arg}-\text{Ala}-\underset{20}{\text{Gln}}-\\ \text{Asp}-\text{Phe}-\text{Val}-\text{Gln}-\text{Trp}-\text{Leu}-\text{Met}-\text{Asn}-\underset{29}{\text{Thr}}-\text{OH}\end{array}$$

If unusual residues are incorporated in sequences, then abbreviations used to represent them should appropriately be defined. An asterisk may be placed above 3-letter symbols to represent such residues.

Fig. 2.1.3.2-3

Incorporation of unusual residues as illustrated with *colistimethate sodium*



L-DAB* = L-DAB or L-DAB with DAB = 2,4-diaminobutanoic acid

Colistin E1 derivative: R = CH_3
Colistin E2 derivative: R = H

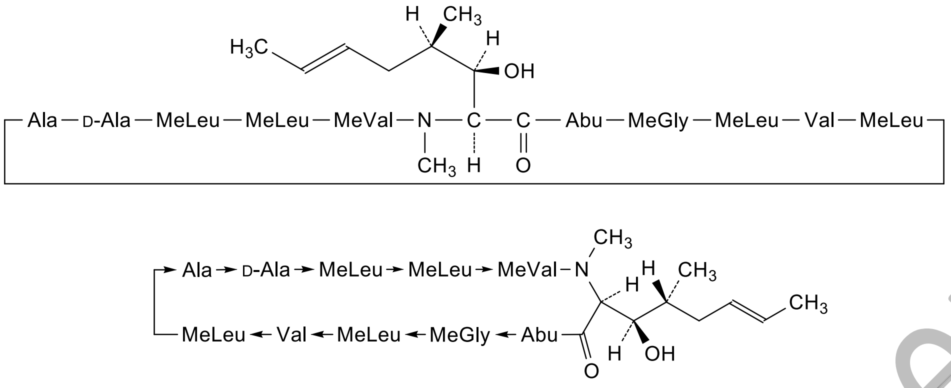
0 to 5 of the DAB residues are disubstituted at N^4

2.1.3.3. Homodetic Cyclic Peptides

Cyclic peptides in which the ring is formed through peptide linkages of amino acids are usually represented with the residues joined at each end on one line with a lengthened bond which may be placed above or below the unit chain, according to the constraints of the structure to be represented. In such linear cyclopeptides, the residues are preferably arranged in alphabetical order (“Ala” before “Leu” and “Ala” before “D-Ala”). If residues are written on two lines, as in head-to-tail cyclopeptides, the sequence is reversed on one of them, hence the CO to NH direction within the peptide bond must be indicated by arrows.

Fig. 2.1.3.3-1

***Ciclosporin* represented as a linear and a head-to-tail cyclopeptide with respectively residues on a single line (top) or two lines (bottom).**

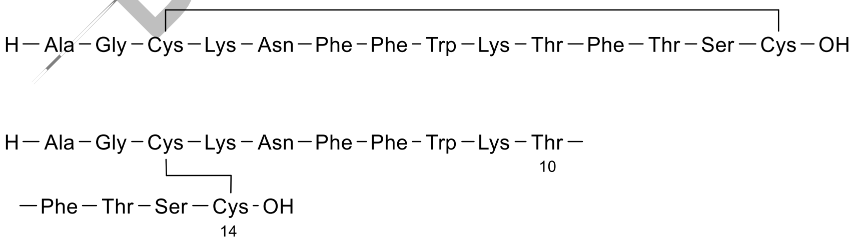


2.1.3.4. Heterodetic Cyclic Peptides

Cyclic peptides, in which the ring is formed through one or several bonds other than peptide linkages (disulfide, ester …) are represented with a combination of lines in the way substitutions of amino acids are depicted. Disulfide bridges, for example, are drawn by connecting side chains of two cysteine residues without explicitly showing any of the sulfur atoms. Sulfanyl groups on side chains are symbolized with vertical lines and are connected with a horizontal line preferably placed above the unit chain if possible. Depending on the length of the peptide sequence, bridges can be drawn in between different lines of residues.

Fig. 2.1.3.4-1

***Somatostatin*, a side chain-to-side chain cyclopeptide as represented with lines between cysteine residues as to illustrate a disulfide bridge. Several depiction styles are possible depending whether the residues or on the same line (top) or not (bottom)**



2.1.3.5. Amino acid 1-letter code

Because of size, polypeptides produced by the condensation of large number of amino acids are more easily interpreted and analyzed, whether or not computer assisted, if represented with 1-letter codes rather than with the standard 3-letter codes. The use of the amino acid 1-letter codes is not restricted to proteins as it is currently used for peptides of about 10 to 20 residues and beyond. Sequences of 1-letter codes are arranged in five sets of ten letters separated by a space and numbers of amino acids are generally placed at the end of each line. A distinction between residues of different chains can be made by using single, double and triple prime symbols. One letter codes of modified residues are underlined and defined below the sequence. All structural characteristics of the alignment such as modified residues, disulfide bridges and glycosylation sites are specified by indicating the positions of the concerned amino acids.

Fig. 2.1.3.5-1

One letter code representation of *nonacog gamma* (human coagulation factor IX (rDNA)), a 415-amino acid protein containing numerous modified residues, disulfide bridges and glycosylation sites

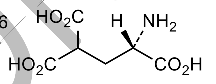
Sequence / Séquence / Secuencia

YNSGKLEEFV QGNLERECME EKCSFEEARE VFENTERTTE FWKQYVDGDQ 50
CE**S**NPCLNGG **S**CKDDINSYE CWCPFGFEGK NCELDVTCNI KNGRCEQFCK 100
NSADNKVVCS CTEGYRLAEN QKSCEPAVPF PCGRVSVSQT SKLTRAEAVF 150
PDVDYV**N**STE AETILD**N**I**T**Q S**T**QSFNDF**T**R VVGGEDAKPG QFPWQVVLNG 200
KVDAFCGGSI VNEKWIVTAA HCVETGVKIT VVAGEHNIEE TEHTEQKRNV 250
IRIIPHHNYN AAINKYNHDI ALLELDEPLV LNSYVTPICI ADKEYTNIFL 300
KFGSGYVSGW GRVFHKGRSA LVLQYLRVPL VDRATCLRST KFTIYNNMFC 350
AGFHEGGRDS CQGDSGGPHV TEVEGTSFLT GIISWGEECA MKGKYGIYTK 400
VSRYVNWIKE KTKLT 415

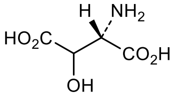
Disulfide bridges location / Position des ponts disulfure / Posiciones de los puentes disulfuro
18-23 51-62 56-71 73-82 88-99 95-109
111-124 132-289 206-222 336-350 361-389

Modified residues / Résidus modifiés / Restos modificados

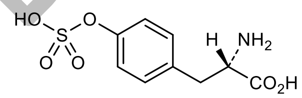
E
7-8-15-17-20-21-26
27-30-33-36-40
4-carboxyGlu



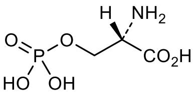
D
64
3-hydroxyAsp



Y
155
O-sulfoTyr



S
158
O-phosphonoSer



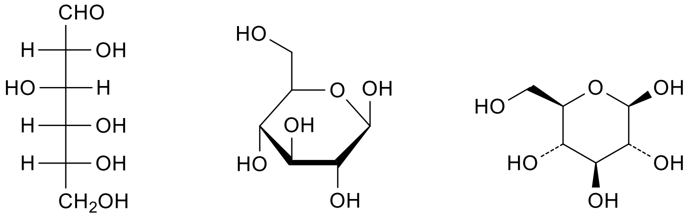
Glycosylation sites / Sites de glycosylation / Posiciones de glicosilación (**N, S*, T***)
Ser-53* Ser-61* Asn-157 Thr-159* Asn-167 Thr-169* Thr-172* Thr-179*
* potential sites / sites potentiels / posiciones posibles

2.2. Structure representation of carbohydrates& derivatives[11]

The simplest carbohydrates, monosaccharides, which can exist in an open chain (acyclic) form or in closed chain (cyclic) form, are generally represented in a variety of ways. The most commonly used are Fischer projections and Haworth and Mills depictions.

Fig. 2.2-1

Depiction of *D-glucose* with a Fischer projection and Haworth and Mills depictions (from left to right)

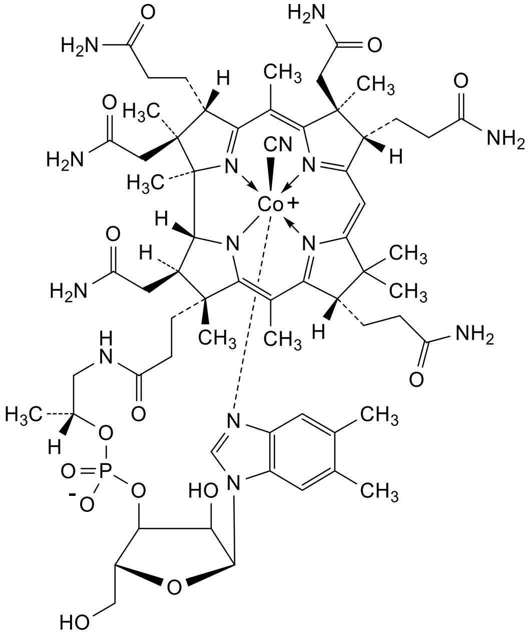


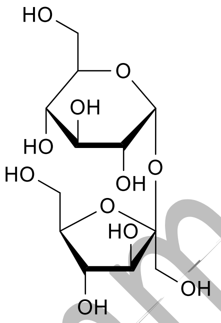
2.2.1. Haworth depictions

Within the INN programme, monosaccharides are preferably illustrated in their cyclic form with simplified Haworth depictions. Pyranose ring hemiacetals are usually represented with the oxygen atom in the ring drawn in the upper right-hand position of the hexagon, with the hemiacetal carbon to the right of this position and thus clockwise numbering of the ring atoms. Furanose rings are drawn with the oxygen at the top in the centre. Carbon atoms are implicit and hydrogens may be omitted for clarity. The lower side of the ring, assumed to be nearer to the observer, is usually thickened to suggest perspective. Only in exceptional cases, when there are bulky substituents to be represented, one may consider drawing Haworth depictions with a different ring orientation.

Fig. 2.2.1-1

Depiction of *cyanocobalamin*(left) with an unconventional Haworth representation due to the presence of bulky substituents as opposed to a conventional representation of a furanose ring in *sucrose* (right)



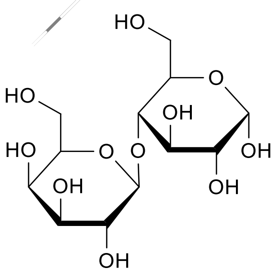


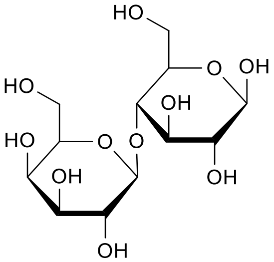
2.2.1.1. Stereochemical considerations

Substituents attached to the ring carbons are drawn using vertical lines to indicate whether a group is oriented to the top-face or to the bottom-face of the ring. The orientation of substituents thereby specifies the configuration at stereogenic centres.

Fig. 2.2.1.1-1

Specification of configuration at the anomeric carbon atom with the alcohol group pointed downwards for *α-lactose* (left) and upwards for *β-lactose* (right)

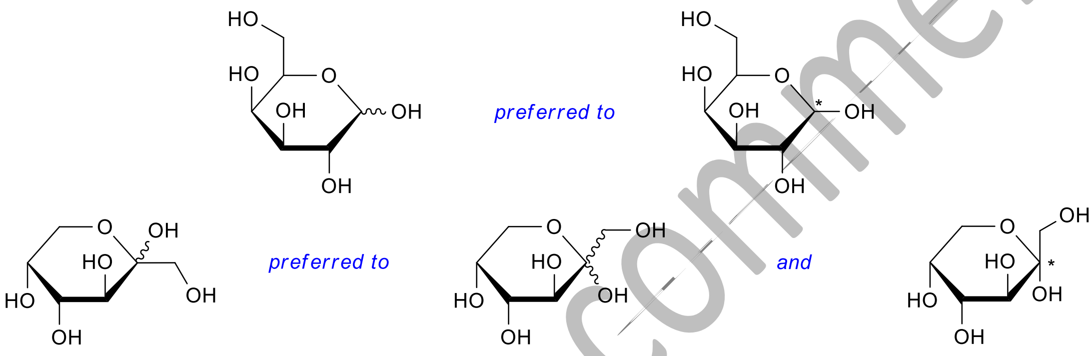




Unknown or unspecified configuration of the anomeric carbon is preferably indicated with a wavy bond rather than by using the general conventions which imply an asterisk flagging the chiral centre in association with lines of normal thickness (*section 1.5.3.2.*). Wavy bonds of monosubstituted anomeric carbon atoms are oriented horizontally whereas those of disubstituted ones are inclined as to differentiate with depictions of known configuration.

Fig. 2.2.1.1-2

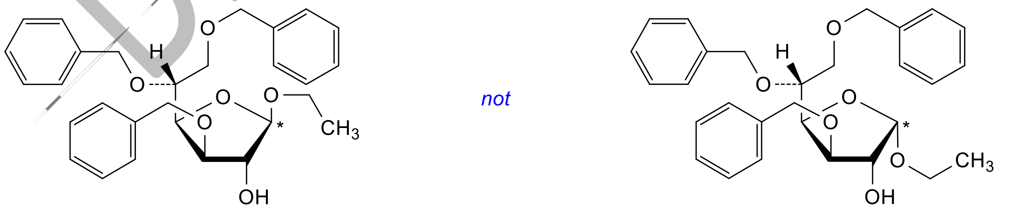
Depiction of unspecified configuration of the anomeric carbon atom of *galactose* (top) and *fructose* (bottom). Drawings represent the α-anomer, the β-anomer, or a mixture of both.



Mixtures of two anomers are depicted by drawing the (*R*)-isomer with respect to the centre of asymmetry of interest marked with an asterisk in association with the legend ‘and anomer at C*’ or ‘and epimer at C*’ joined to the structure.

Fig. 2.2.1.1-3

Depiction of a mixture of anomers; the example shown is *tribenoside* with the (*R*)-isomer of the anomeric carbon drawn on the left which is preferred to the (*S*)-isomer (right)



and anomer at C*

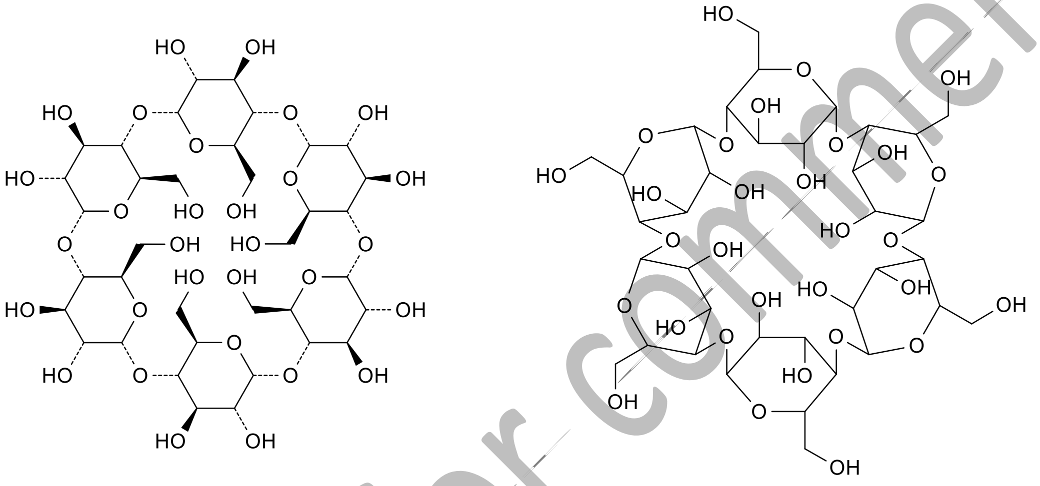
and anomer at C*

2.2.2. Mills depictions

In cases of congested structures in particular, the Mills depiction offers an alternative to Haworth projections to preserve clarity. The main hemiacetal ring is drawn in the plane of the paper and filled wedges and dashed bonds are used to denote the orientation of respectively substituents above and below the plane. Hydrogen atoms on ring carbon atoms are omitted and, if necessary, ambiguity in configuration at stereogenic centres can be shown by using either a wavy bond or an asterisk.

Fig. 2.2.2-1

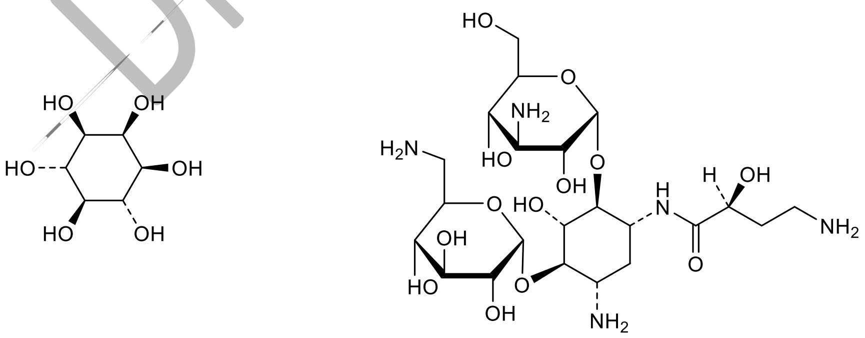
Alfadex as represented with the Mills depiction (left) and the Haworth depiction (right)



Mills depictions are also commonly used for depicting congested derivatives of cyclitols such as streptamine derived aminoglycosides.

Fig. 2.2.2-2

Mills depictions of *myo-inositol* (left) and *amikacin* (right)

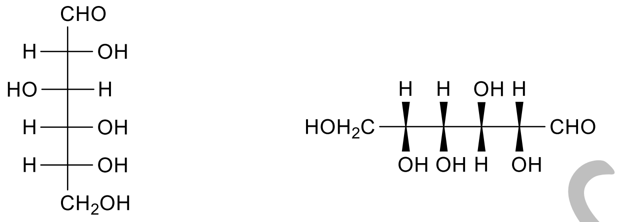


2.2.3. Conformations of acyclic chains: Fischer projections

Fischer projections allow easy classification as either the D-enantiomer or the L-enantiomer and, as such, generally offer a practical way of representing the acyclic structure of a sugar. In true Fischer projections, all of the horizontal bonds are oriented out of the plane and all of the vertical bonds are oriented into the plane as to show the configuration at each stereogenic centre. Modified Fischer projections are set out horizontally with the carbon C-1 on the right and with all vertical bonds shown as filled wedges to convey information on configuration without ambiguity. In the latter case, all horizontal bonds represented with lines of normal thickness point away from the viewer.

Fig. 2.2.3-1

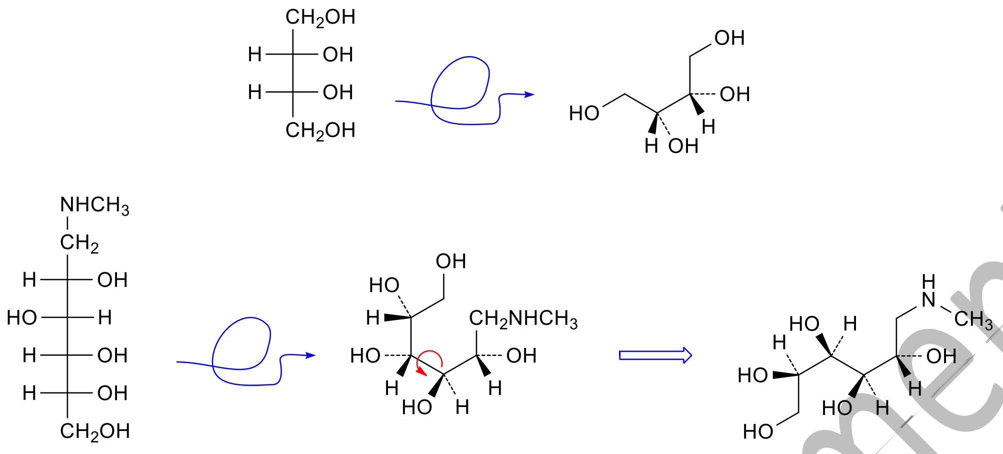
Depiction of *D-glucose* with a true Fischer projection (left) and with a modified Fischer projection (right)



Although true and modified Fischer projections have been used within the INN programme, a more conventional depiction style is preferred in order not to mix different ways of suggesting perspective. The Fischer projection is therefore converted to show the complete backbone in the plane of depiction with the use of filled wedges and dashed bonds to orient substituents. Twisting the backbone may be necessary for long chains to avoid structure overlap.

Fig. 2.2.3-2

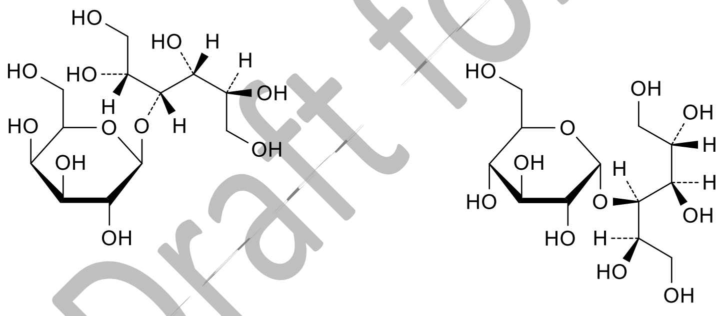
Depiction of *erythritol* (top) and *meglumine* (bottom) as converted from true Fischer projections by drawing backbones in the plane of depiction



Converted Fischer projections are also preferred to depict non-cyclic fragments that consist of sugar alcohols that are part of saccharides.

Fig. 2.2.3-3

Depiction of *lactitol* (left) and *maltitol* (right) with non-cyclic fragments depicted as converted Fischer projections by drawing backbones in the plane of depiction

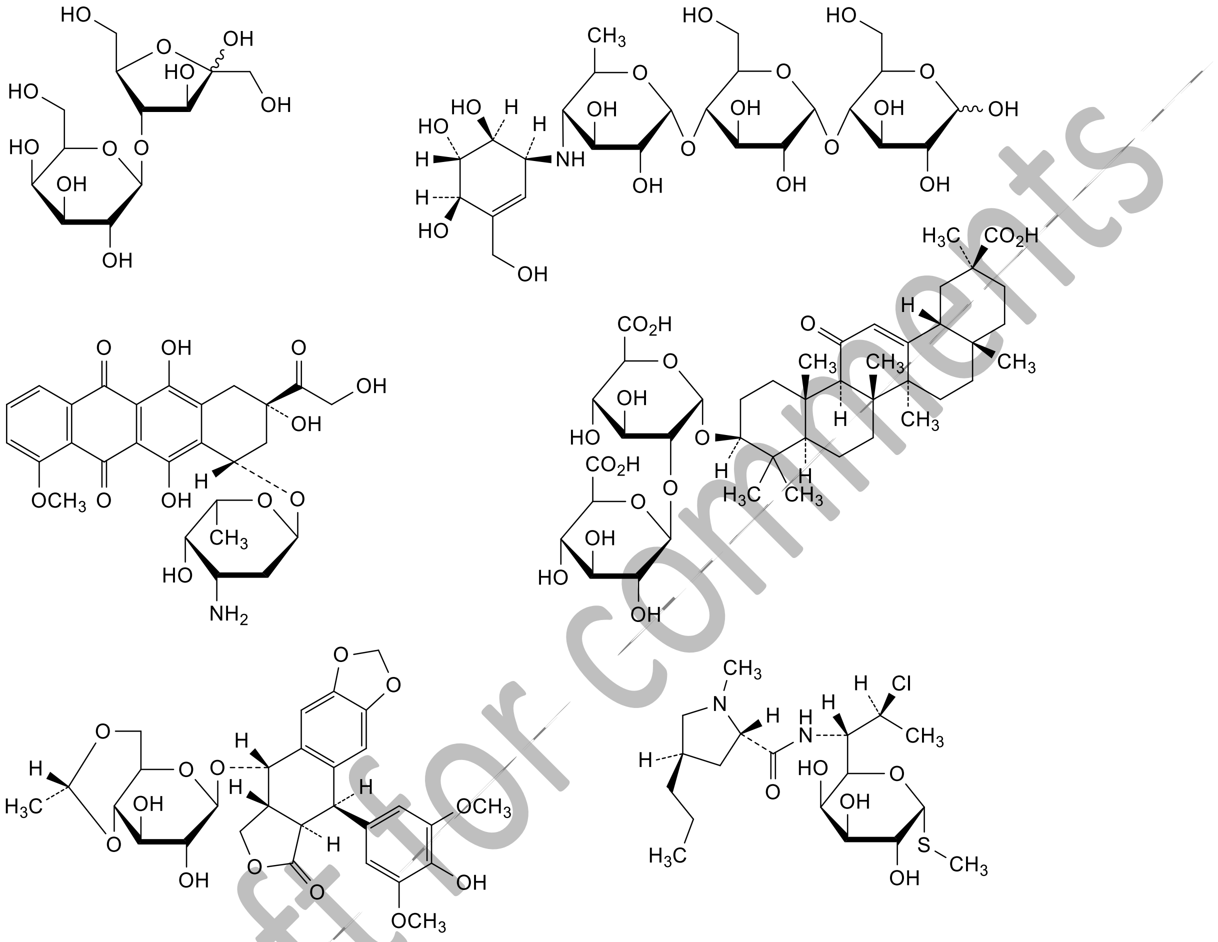


2.2.4. Oligosaccharides and saccharide derivatives

Oligo- and polysaccharides are depicted in line with the above-mentioned recommendations. For structures that are only partly saccharide, only the fragment corresponding to a carbohydrate is depicted as such, the remaining of the molecule being represented according to the general conventions or recommendations for specific classes of compounds, were applicable.

Fig. 2.2.4-1

Different classes of compounds containing carbohydrate moieties as illustrated with *actulose* (top left), *acarbose* (top right), *doxorubicin* (middle left), *glycyrrhizin* (middle right), *etoposide* (bottom left) and *clindamycin* (bottom right)



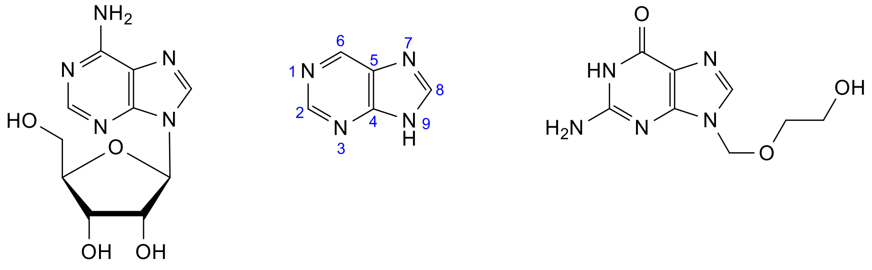
2.3. Structure representation of nucleosides and polynucleotides[12]

2.3.1. Nucleosides and small derivatives

Nucleosides are, although characterized in the case of purine-derivatives with a non-systematic numbering of the main skeleton, represented following the standard conventions as described under General Issues.

Fig. 2.3.1-1

Orientation of *adenosine* (left) and *aciclovir* (right) as compared to the purine skeleton (middle)



2.3.2. Nucleic acids and derivatives

Alignments containing nucleoside residues, whether modified or not, are depicted by a sequence of single capital letters separated with hyphens to represent phosphate residues. A diesterified phosphate residue, represented by a hyphen, is considered to be attached to the oxygen atom of the 3' carbon on its left and to that of the 5' carbon on its right, unless other numerals are used. When left-to-right direction is *not* the case, this must be indicated by appropriate locants preceding the chain or by an arrow to indicate the 3'→5' direction. The standard symbols for nucleosides in case of modified residues may be prefixed by a symbol of lower-case letter(s) placed immediately before the single capital letter. Additional modifications may be indicated with lower-case letter symbols after the capital letter. Base-modifying prefixes may also precede a whole chain if applicable to each residue.

Fig. 2.3.2-1

Structure representation of *prexigebersenum* as to illustrate the use of a modifying prefix "d" (for 2'-deoxy) at the beginning of the chain applicable to each residue

(3'-5') d(A−T−A−T−T−T−G−G−C−G−A−T−G−G−C−T−T−C)

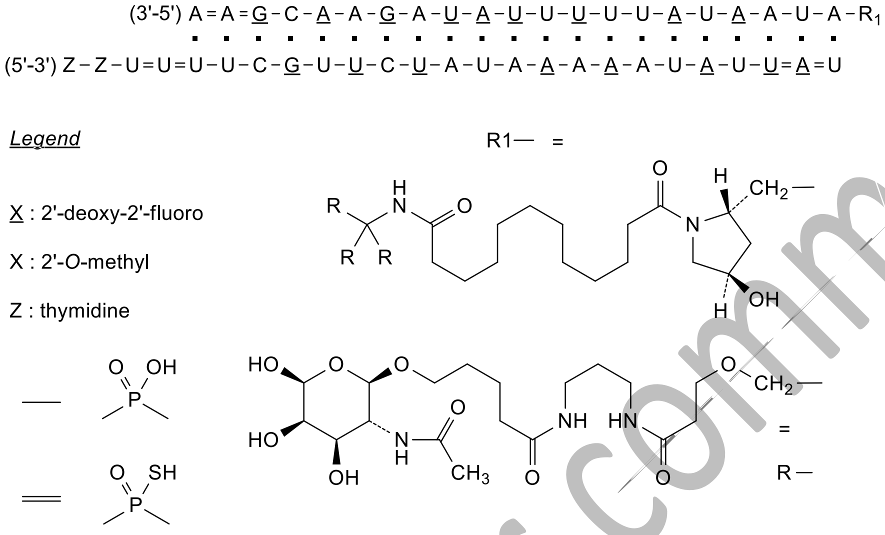
If the polynucleotide occupies more than one line, then a hyphen is placed at the end of each successive line until the formula has been completed.

A noncovalent association between polynucleotide chains, such as that ascribed to hydrogen-bonding, is indicated by a centered dot between nucleotide units of different chains.

For complex modifications, capital letters representing residues may be underlined once (or twice or shown in bold if a further going distinction is required) and defined below the alignment. Modification of a phosphate residue may be symbolized with a double hyphen instead of a single one.

Fig. 2.3.2-2

Structure representation of *cemdisiran* as to illustrate the use and definition of modified symbols to designate structural complex residues



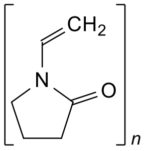
2.4. Structure representation of polymers[13]

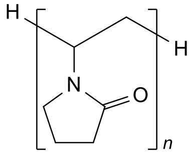
2.4.1. General aspects

The pictorial representation of polymers is either based on the structure of monomers from which they are derived or the preferred constitutional repeating units (CRUs) and, as such, although numerous exceptions exist, it parallels polymer nomenclature which is either source-based or structure-based.

Fig. 2.4.1-1

Source-based depiction (left) as opposed to structure-based depiction (right); the example shown is *povidone*

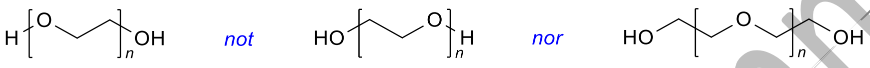




Whereas formula representation of monomers is in accordance with the recommendations described under General Issues, the CRUs are depicted as multivalent radicals which are preferably oriented as such to show descending order of seniority of subunits from left to right wherever possible.

Fig. 2.4.1-2

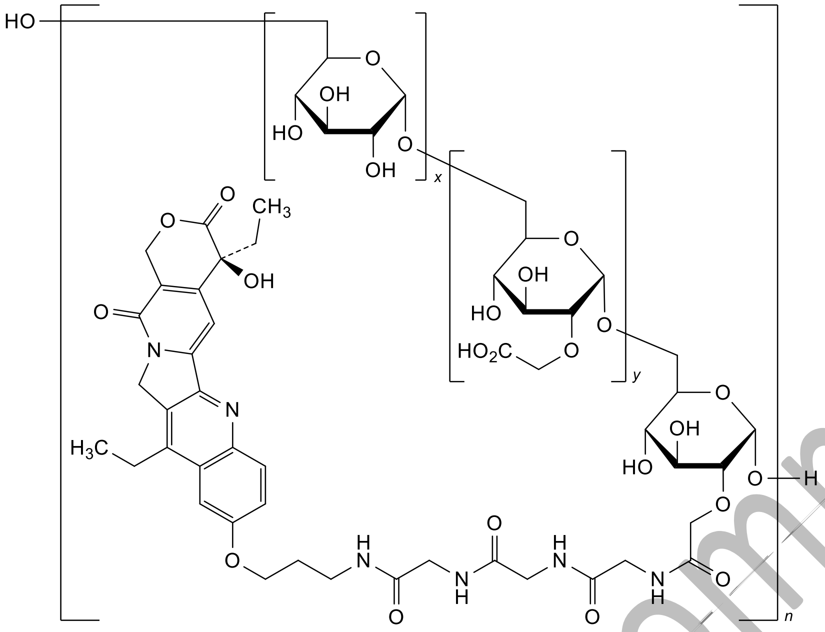
The preferred depiction of a CRU is characterized with the smallest structural repeating unit showing senior elements first from left to right; the example shown is *macrogol*



Square brackets together with italic subscript letters denote multiplicity of both monomers and CRUs. The subscript letters *m*, *n, p*, *q,* etc. denote multiplicities of polymeric sequences, whereas the subscript letters *a, b, c* or *x, y, z,* and so on, designate repetition units to characterize linear arrangements of specified polymer fragments and blocks derived from a typical species of monomer. Dashes indicating free valences of CRUs are usually drawn through the brackets. The formulas of end groups, if known, are attached to the bonds at the ends of the constitutional units, but placed outside the brackets.

Fig. 2.4.1-3

The use of square brackets and italic subscript letters to denote multiplicity as illustrated with *delimotecan*



The graphic representation of polymers depends on what is known about the sequence arrangement of the constituent monomers and the way they are connected.

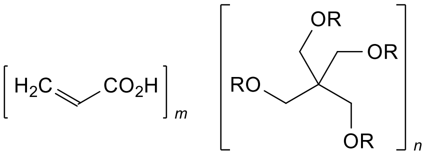
2.4.2. Unspecified topology

Source-based depictions are employed when it is difficult to specify sequence arrangements of monomers. Thus, the constituent species from which the polymer is derived are depicted as such.

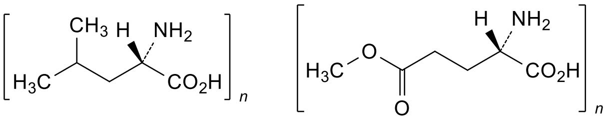
Fig. 2.4.2-1

Source-based representation of *carbomer* (top), which consists of homopolymers of acrylic acid crosslinked with pentaerythritol allyl ethers. *Leuciglumer* (bottom) is an example of a copolymer of leucine and methylglutamic acid in equimolar proportions but with undefined sequential arrangement of the different monomers.

carbomer



$-R = -H, -CH_2-CH=CH_2$



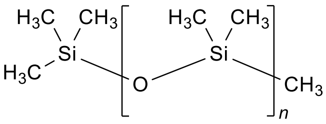
leuciglumer

2.4.3. Specified topology

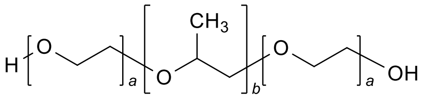
If the sequence arrangement of monomeric units of a polymer is defined, then structure-based depictions are preferred.

Fig. 2.4.3-1

Structure-based representation of *dimeticone* (top) and *poloxamer* (bottom left) which are respectively defined as a homopolymer and a block copolymer of known topology. The latter is composed of a central chain of polyoxypropylene flanked by two chains of polyoxyethylene, as such its structure-based depiction is preferred to its source-based depiction (bottom right).

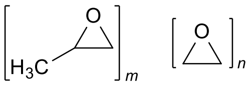


dimeticone



polaxamer

preferred to



III. Structure representation of specific chemical classes

3.1. General considerations

Displaying structures as accurately as possible to current best practices implies taking into account IUPAC nomenclature recommendations, all the more that one of the basic principles is about structures reflecting their chemical name as closely as possible. In structure representation, the orientation of the skeleton is generally related to the numbering pattern established for nomenclature purposes. Substances defined with systematic names are, as such, easily depicted with a consistent style as described under General Issues. Natural products[14] and derivatives however are sometimes preferably defined with semi-sytematic names and can be characterized with special (non-systematic) numbering patterns. Such structures may therefore not always be oriented consistently with respect to their systematic names. In these cases, priority is given to standardize the orientation of specific skeletons of closely related substances to favor easy structural comparison. This section details the special conventions for such compounds.

3.2. Alkaloids

Alkaloids are characterized by a great structural diversity and, therefore, may be represented in many different ways. Similarity of main skeletons, however, has led to several structurally related compounds being favored.

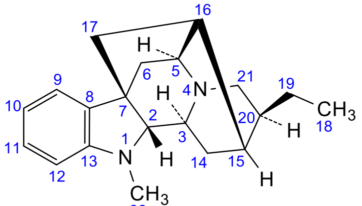
3.2.1. Preferred orientation based on a stereoparent

Structural complex alkaloids are considered through a fundamental stereoparent structure with a specific numbering pattern and orientation of the skeleton. Such structural diagrams are used to reflect alkaloids which are defined with (non-systematic) stereoparent names. Preferred orientations for structures of stereoparents are listed below together with some relevant examples.

ajmalan derivatives

(stem: *-aj-*)

stereoparent: ajmalan

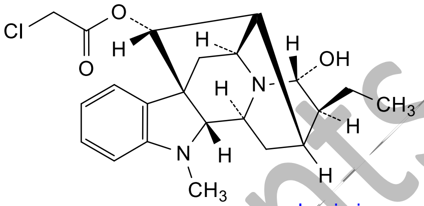


ajmalan

example *lorajmine* :

(17*R*)-21α-hydroxyajmalan-17-yl

chloroacetate

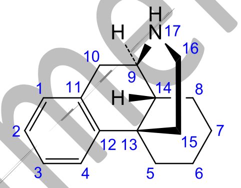


lorajmine

morphinan derivates

(stems: *nal-*, *orphan*)

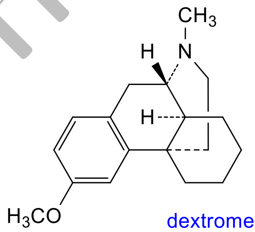
stereoparent: morphinan



morphinan

example *dextromethorphan* :

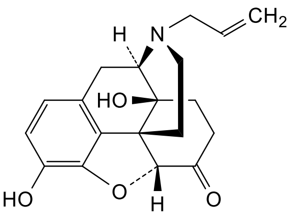
ent-3-methoxy-17-methylmorphinan



dextromethorphan

example *naloxone* :

4,5α-epoxy-3,14-dihydroxy-17-(prop-2-en-1-yl)morphinan-6-one

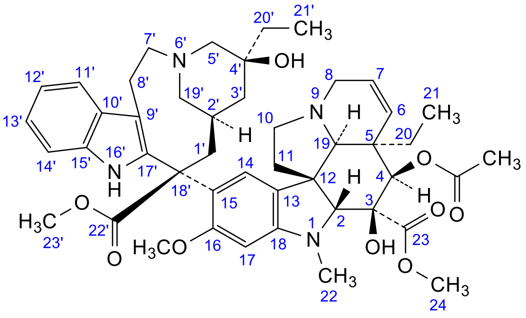


naloxone

vincaleucoblastine derivatives

(stem: *-vin-)*

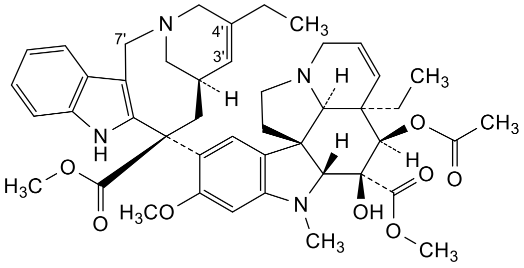
stereoparent: vincaleucoblastine



vincaleucoblastine

example *vinorelbine* :

4'-deoxy-3',4'-didehydro-8'-norvincaleukoblastine

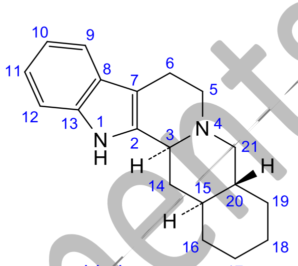


vinorelbine

yohimban derivatives

(stem: *-serpine*)

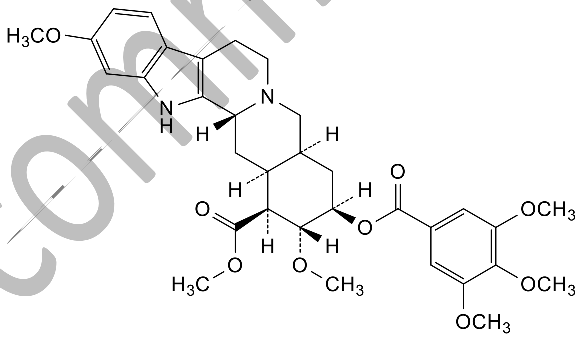
stereoparent: yohimban



yohimban

example *reserpine*:

methyl 11,17α-dimethoxy-18β-[(3,4,5-trimethoxybenzoyl)oxy]-3β,20α-yohimban-16β-carboxylate



reserpine

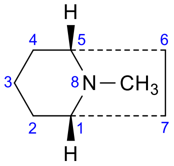
3.2.2. Preferred orientation based on a core structure, a parent hydride or a stereoparent

Less complex alkaloids are defined with a systematic name and may be depicted with a preferred orientation of the main skeleton to make clear that (all) the compounds belonging to the same chemical class are structurally analogous. The core structure can, but need not be, a fundamental stereoparent or parent hydride.

atropine derivatives

(stem: *trop*)

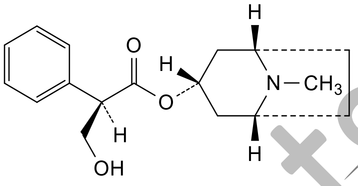
stereoparent: tropane



tropane

example *atropine*:

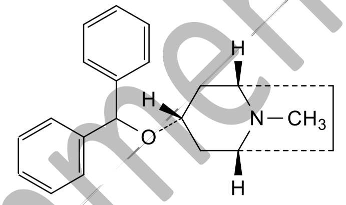
(1*R*,3*r*,5*S*)-8-methyl-8-azabicyclo[3.2.1]octan-3-yl (2*RS*)-3-hydroxy-2-phenylpropanoate



atropine

example *benzatropine*:

(1*R*,3*r*,5*S*)-3-(diphenylmethoxy)-8-methyl-8-azabicyclo[3.2.1]octane

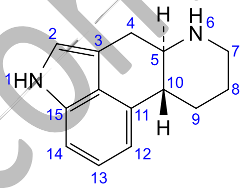


benzatropine

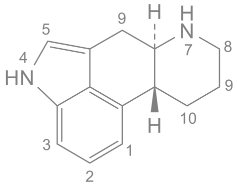
ergoline derivatives

(stems: *erg, ergolide*)

stereoparent: ergoline



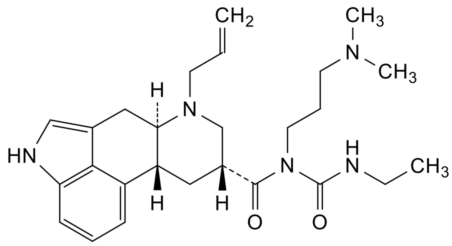
ergoline



systematic numbering

example *cabergoline*:

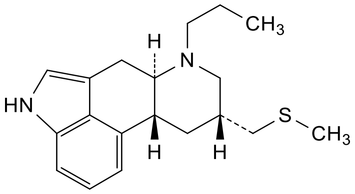
(6a*R*,9*R*,10a*R*)-*N*-[3-(dimethylamino)propyl (ethylcarbamoyl)-7-(prop-2-en-1-yl)-4,6,6a,7,8,9,10,10a-octahydroindolo[4,3-*fg*]quinoline-9-carboxamide



cabergoline

example *pergolide*:

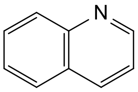
(6a*R*,9*R*,10a*R*)-9-[(methylsulfanyl)methyl]-7-propyl-4,6,6a,7,8,9,10,10a-octahydroindolo[4,3-*fg*]quinoline



pergolide

quinolineError! Bookmark not defined. **derivatives**

(stem: *-quine*)

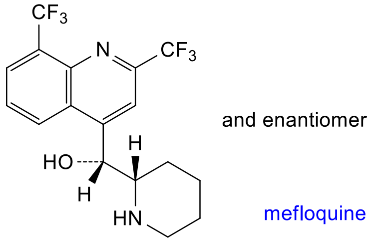


quinoline

parent hydride: quinoline

example *mefloquine*:

(*RS*)-[2,8-bis(trifluoromethyl)-quinolin-4-yl][(2*SR*)-piperidin-2-yl]methanol



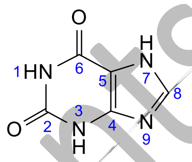
mefloquine

xanthine derivatives

(stem: -*fylline*)

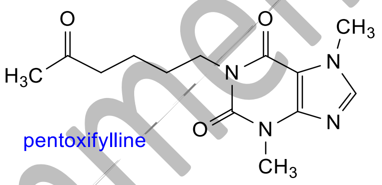
core structure: xanthine

xanthine



example *pentoxifylline*:

3,7-dimethyl-1-(5-oxohexyl)-3,7-dihydro-1*H*-purine-2,6-dione

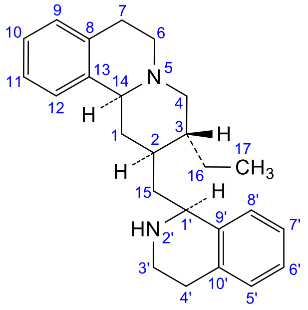


pentoxifylline

3.2.3. Orientation based on standard graphical conventions

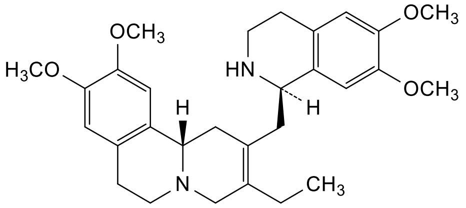
If no, or only a few, compounds share similarity, then most alkaloids of relatively simple structure are preferably defined with a systematic name and are depicted following the standard conventions described under General Issues. As an example, the systematic name and a standard conventional depiction are preferred for *dehydroemitine*, as shown below.

stereoparent: emetan



emetan

example *dehydroemetine*:



dehydroemetine

systematic name:	(11b*S*)-2-{[(1*R*)-6,7-dimethoxy-1,2,3,4-tetrahydroisoquinolin-1-yl]methyl}-3-ethyl-9,10-dimethoxy-4,6,7,11b-tetrahydro-1*H*-pyrido[2,1-*a*]isoquinoline
semi systematic name:	2,3-didehydro-6',7',10,11-tetramethoxyemetan

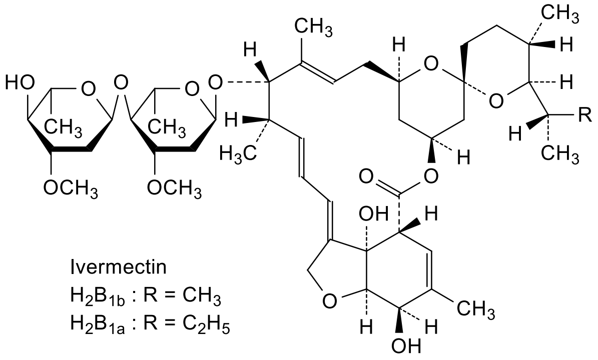
3.3. Polyketides

Antibiotics include numerous substances of relatively large and complicated structures which can be characterized as polyketides of different types among which are marcolides, polyenes and tetracyclines. These substances are defined with a systematic name and depicted in a uniform way in line with the recommendations described under General Issues. Graphic representations are based on that of a parent structure.

avermectin derivatives

(stem: *-ectin*)

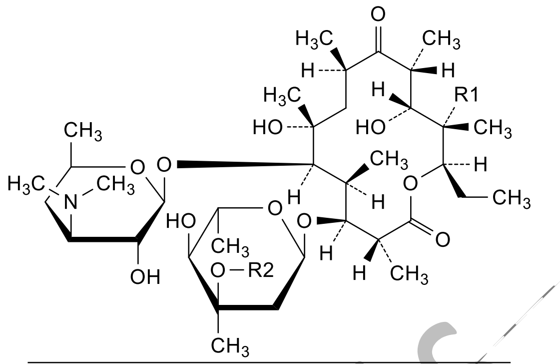
example *ivermectin*



erythromycin derivatives

(stem: -*mycin*)

example *erythromycin*

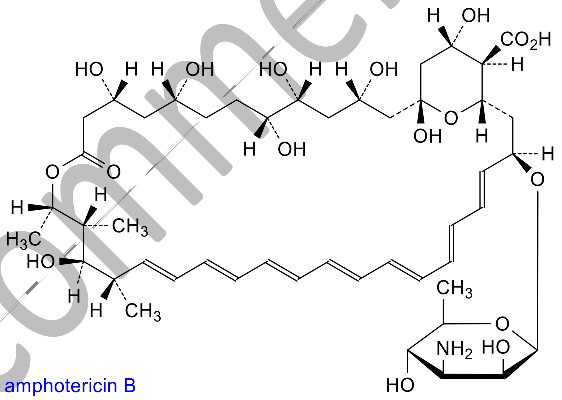


Erythromycin	Mol. Formula	M_r	R1	R2
A	$C_{37}H_{67}NO_{13}$	734	OH	CH_3
B	$C_{37}H_{67}NO_{12}$	718	H	CH_3
C	$C_{36}H_{65}NO_{13}$	720	OH	H

partricin derivatives

(stem: -*tricin*)

example *amphotericin B*

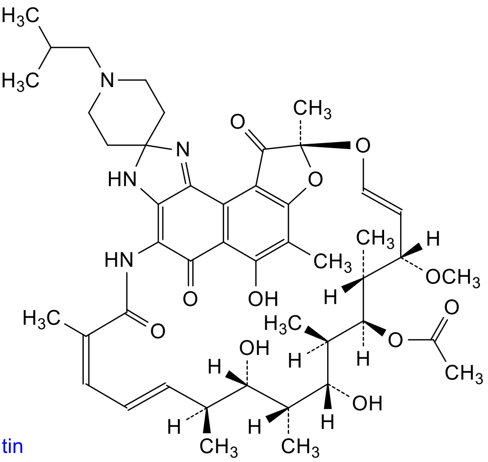


amphotericin B

rifamycin derivatives

(stem: *rifa*-)

example *rifabutin*

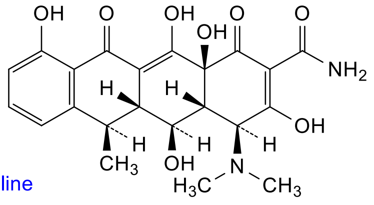


rifabutin

tetracycline derivatives

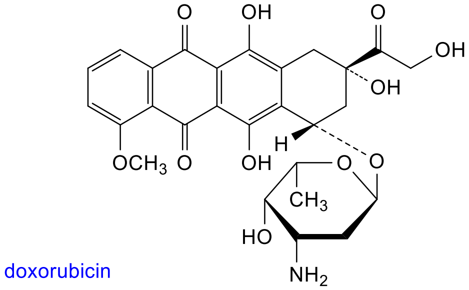
(stems: -*cycline*, -*rubicin*)

example *doxocycline*



doxycycline

example *doxorubicin*

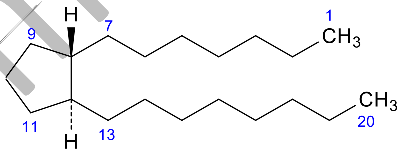

doxorubicin

3.4. Prostanoids

Prostanoids include prostaglandins and prostacyclins and constitute a homogenous series of compounds with limited structural diversity closely related to prostane. They are defined with systematic names and are depicted as the stereoparent, which is in line with the standard conventions described under General Issues.

prostanoids

stereoparent: prostane

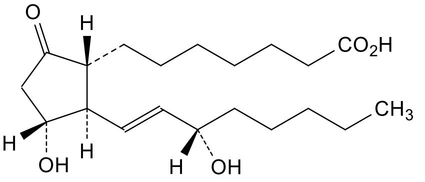


prostane

prostaglandins (stem: *prost*)

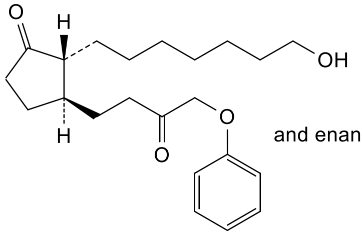
example *alprostadil*:

7-{(1*R*,2*R*,3*R*)-3-hydroxy-2-[(1*E*,3*S*)-3-hydroxyoct-1-en-1-yl]-5-oxocyclopentyl}heptanoic acid


alprostadil

example *oxoprostol*:

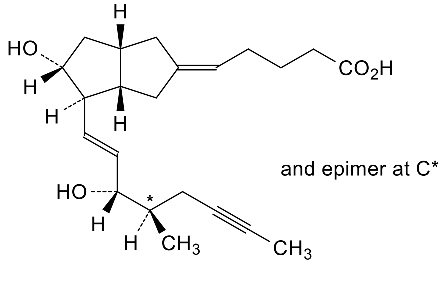
(2*R*,3*R*)-2-(7-hydroxyheptyl)-3-[3-oxo-4-(phenoxy)butyl]-cyclopentan-1-one


and enantiomer

oxoprostol

prostacyclins (stem: *prost*)

example *iloprost*:
(5*E*)-5-[(3a*S*,4*R*,5*R*,6a*S*)-5-hydroxy-4-[(1*E*,3*S*,4*RS*)-3-hydroxy-4-methyloct-1-en-6-yn-1-yl]hexahydropentalen-2(1*H*)-ylidene]pentanoic acid



iloprost

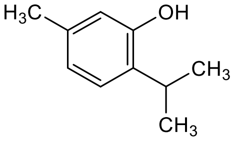
3.5. Terpenoids

Terpenoids present a large and structurally diverse class of compounds. Basic carbon skeletons within this class of compounds can be distinguished according to the number of terpene/isoprene units necessary to constitute the core structure and the way they are assembled.

3.5.1. Mono- and diterpenoids

The smaller terpenoids of relatively simple structure, essentially monoterpenoids together with some diterpenoids such as retinol derivatives, are usually defined with their systematic name and are depicted following the standard conventions described under General Issues.

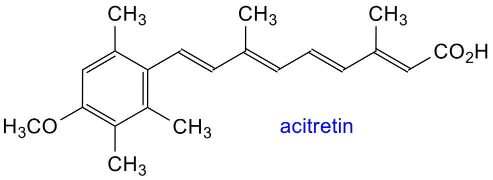
example *thymol*
5-methyl-2-(methylethyl)phenol



thymol

retinol derivatives (stem: *retin*)

example *acitretin*:
(all-*E*)-9-(4-methoxy-2,3,6-trimethylphenyl)-3,7-dimethylnona-2,4,6,8-tetraenoic acid

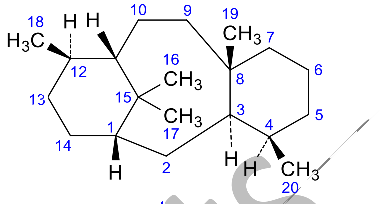


acitretin

Diterpenoids of more complicated structure, such as taxane derivatives, are defined with a semi-systematic name based on a fundamental stereoparent with specific numbering pattern and orientation of the skeleton. If no stereoparent name has become established for a class of compounds, for example *pleuromulin* derivatives, then substances are defined by their systematic name and are depicted in a uniform way as to show structural similarity between compounds belonging to the same chemical class.

taxane derivatives (stem: -*taxel*)

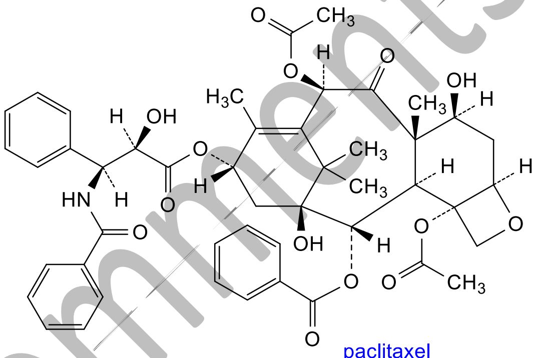
stereoparent: taxane



taxane

example *paclitaxel*:

5β,20-epoxy-1,7β-dihydroxy-9-oxotax-11-ene-2α,4,10β,13α-tetrayl 4,10-diacetate 2-benzoate 13-[(2*R*,3*S*)-3-benzamido-2-hydroxy-3-phenylpropanoate]

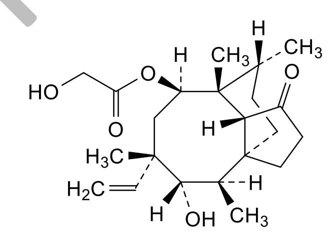


paclitaxel

pleuromulin derivatives

(stem: -*mulin*)

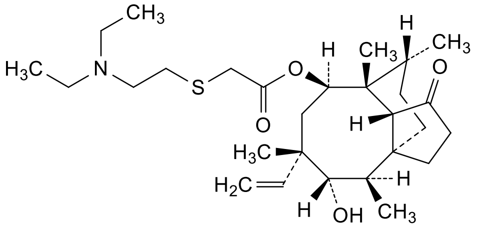
core structure: pleuromulin



pleuromulin

example *tiamulin*:

(3a*S*,4*R*,5*S*,6*S*,8*R*,9*R*,9a*R*,10*R*)-6-eth-1-en-1-yl-5-hydroxy-4,6,9,10-tetramethyl-1-oxodecahydro-3a,9-propanocyclopenta[8]annulen-8-yl {[(diethylamino)ethyl]sulfanyl}acetate



tiamulin

3.5.2. Triterpenoids

Larger terpenoids concern in some cases substances sharing structural similarity with a considerable number of compounds. Among such triterpenoids, the cyclic ones containing four rings and beyond are defined with semi systematic names and are depicted with a specific orientation of the main skeleton.

Examples are steroids and glycyrrhetic acid derivatives. Secosteroids which bear only three rings in the core structure are derived by ring opening from steroids and are considered and depicted as such.

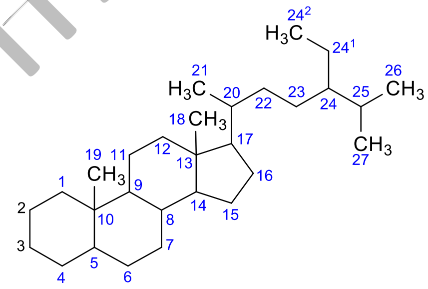
Structures are denoted as projections of the carbon framework onto the plane of depiction. Solid wedges and broken lines are used at stereogenic centers to orient atoms or groups which are designated alpha (α) if they lay below or beta (β) if they lay above the plane of depiction. Use of this system requires strict orientation of structures which are normally represented having position 3 at the bottom left with counterclockwise numbering for the first ten skeletal atoms. Unknown configuration of a stereogenic centre of the ring system is indicated with a wavy bond and designated xi (ξ). In the case of a racemic compound, the enantiomeric structure drawn should be the one that shows the lowest numbered chiral center in the α-configuration. Graphical specifications of chirality concerning side chains of the ring system follow standard conventions described under General Issues.

steroids –

various stereoparents[15]

(stems: *andr*, *-betasol*, *bol*, *cort*, *estr*, *gest*, *-gestr-*, *-met(h)asone*, *-olone*, *pred*, *-pris-*, *-renone*, *-ster-*, *-terone*)

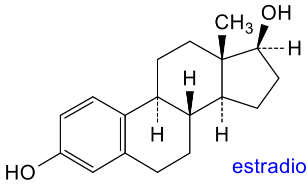
general core structure and numbering pattern



example *estradiol*:

estra-1,3,5(10)-triene-3,17β-diol

(stereoparent: estrane)

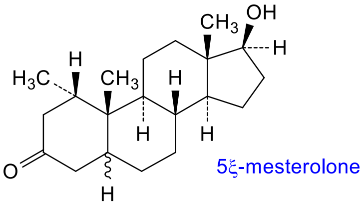


estradiol

example *5ξ-mesterolone*:

17β-hydroxy-1α-methyl-5ξ-androstan-3-one

(stereoparent: androstane)

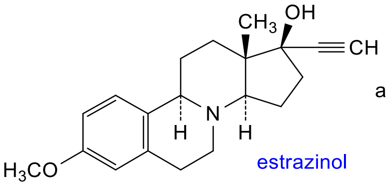


5ξ-mesterolone

example *estrazinol*:

rac-3-methoxy-8-aza-19-nor-17α-pregna-1,3,5(10)-triene-20-yn-17-ol

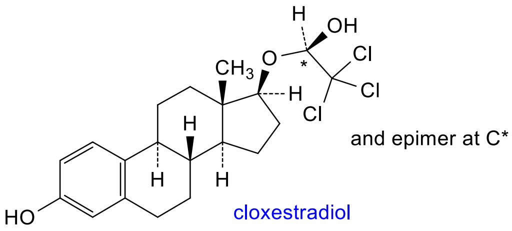
(stereoparent: pregnane)



and enantiomer

estrazinol

example *cloxestradiol*: 17β-[(1*RS*)-2,2,2-trichloro-1-hydroxyethoxy]estra-1,3,5(10)-trien-3-ol(stereoparent: estrange)

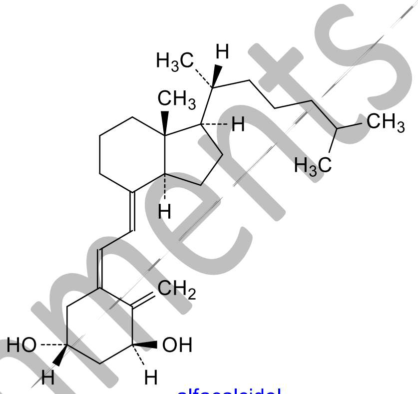

and epimer at C*

cloxestradiol

<u>secosteroids</u>

(stem: *calci*)

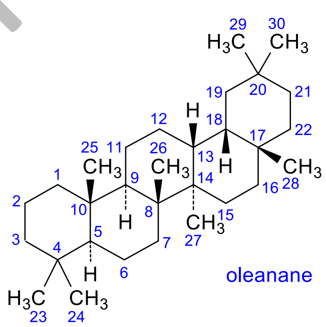
example *alfacalcidol*: (1*S*,3*S*,5*Z*,7*E*)-9,10-secocholesta-5,7,10(19)-triene-1,3-diol (stereoparent: cholestane


alfacalcidol

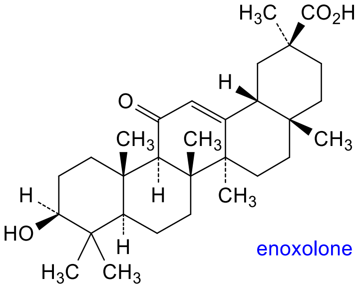
glycyrrhetic acid derivatives

(stem: -*olone*)

stereoparent: oleanane


oleanane

example *enoxolone*: (20β)-3β-hydroxy-11-oxoolean-12-en-29-oic acid.


enoxolone

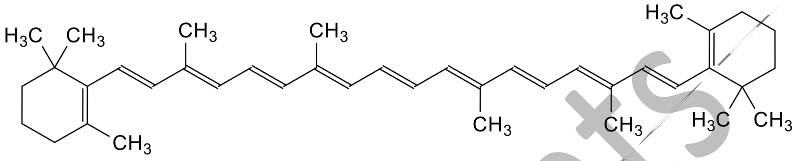
3.5.3. Tetraterpenoids

Tetraterpenoids or carotenoids constitute large molecules of relatively simple structure. Although fundamental parent hydrides with specific numbering have been established for carotenoids, they are commonly defined with a systematic name and depicted in line with the standard conventions described under General Issues.

carotinoid derivatives

(stem: -*carotene*)

example *betacarotene*



betacarotene

systematic name	(all-*E*)-3,7,12,16-tetramethyl-1,18-bis(2,6,6-trimethylcyclohex-1-en-1-yl)octadeca-1,3,5,7,9,11,13,15,17-nonaene
parent hydride	β,β-carotene (trivial name)

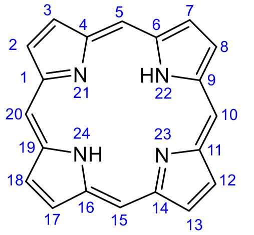
3.6. Tetrapyrroles

Tetrapyrroles or porphyrins are a structurally limited class of natural products and derivatives related to porph(yr)in. Both their nomenclature and representation is based on the parent structure porphyrin which special clockwise numbering pattern consists of the carbon atoms numbered 1 to 20 and the nitrogen atoms numbered 21 to 24.

porphyrin derivatives

(stem: -*porfin*)

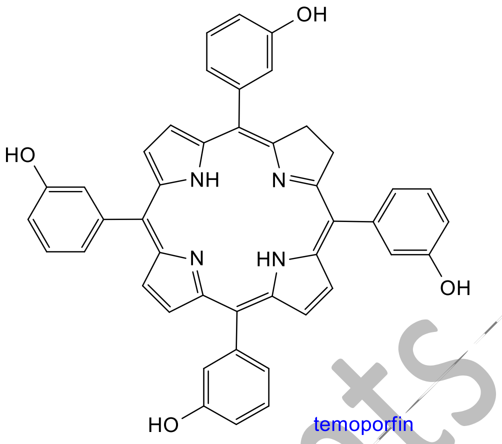
parent hydride : porphyrin



porphyrin

example *temoporfin*:

3,3',3'',3'''-(2,3-dihydroporphyrin-5,10,15,20-tetrayl)tetraphenol



1842

1843

IV. References

1. **http://www.who.int/medicines/services/inn/en/**
2. **http://apps.who.int/medicinedocs/pdf/h1807e/h1807e.pdf**
3. **GRAPHICAL REPRESENTATION STANDARDS FOR CHEMICAL STRUCTURE DIAGRAMS(IUPAC Recommendations 2008)***Pure Appl. Chem.*, Vol. 80, No. 2, pp. 277–410, 2008.
4. **GRAPHICAL REPRESENTATION OF STEREOCHEMICAL CONFIGURATION (IUPAC Recommendations 2006)***Pure Appl. Chem.*, Vol. 78, No. 10, pp. 1897–1970, 2006.
5. **NOMENCLATURE OF ORGANIC CHEMISTRY (IUPAC Recommendations and Preferred Names 2013),** Published by*The Royal Society of Chemistry,*Cambridge (UK),ISBN: 978-0-85404-182-4, 2014.
6. **https://mednet-communities.net/inn**
7. **NOMENCLATURE OF INORGANIC CHEMISTRY (IUPAC Recommendations 2005),** Published by*The Royal Society of Chemistry,*Cambridge (UK), ISBN 0-85404-438-8, 2005.
8. **NOMENCLATURE OF ORGANIC CHEMISTRY. SECTION H: ISOTOPICALLY MODIFIED COMPOUNDS,** *Pure Appl. Chem.*, Vol. 51, pp. 353-380, 1979.
9. **http://www.who.int/medicines/services/inn/StemBook_2013_Final.pdf**
10. **NOMENCLATURE AND SYMBOLISM FOR AMINO ACIDS AND PEPTIDES (IUPAC-IUB Recommendations 1983)** *Pure Appl. Chem.*, Vol. 56, No. 5, pp. 595-624, 1984.
11. **NOMENCLATURE OF CARBOHYDRATES (IUPAC Recommendations 1996)***Pure Appl. Chem.*, Vol. 68, No. 10, pp. 1919-2008, 1996.
12. **ABBREVIATIONS AND SYMBOLS FOR NUCLEIC ACIDS, POLYNUCLEOTIDES AND THEIR CONSTITUENTS (IUPAC-IUB Recommendations 1970)** *Pure Appl. Chem.*, Vol. 40,, pp. 277-290, 1974.
13. **NOMENCLATURE OF REGULAR SINGLE-STRAND ORGANIC POLYMERS (IUPAC RECOMMENDATIONS 2002)** *Pure Appl. Chem.*, Vol. 74, No. 10, pp. 1921-1956, 2002.
14. **REVISED SECTION F: NATURAL PRODUCTS AND RELATED COMPOUNDS (IUPAC Recommendations 1999)***Pure Appl. Chem.*, Vol. 71, No. 4, pp. 587-643, 1999.
15. **NOMENCLATURE OF STEROIDS (IUPAC Recommendations 1989)** *Pure Appl. Chem.*, Vol. 61, No. 10, pp. 1783-1822, 1989.

V. Structure index

***