

***Nomenclature of Organic Chemistry. IUPAC Recommendations and Preferred Names 2013.***

Prepared for publication by Henri A. Favre and Warren H. Powell,  
Royal Society of Chemistry, ISBN 978-0-85404-182-4

---

**Chapter P-10 PARENT STRUCTURES FOR NATURAL PRODUCTS  
AND RELATED COMPOUNDS**

P-100 Introduction

P-101 Nomenclature for natural products based on parent hydrides (alkaloids, steroids, terpenes, carotenes, corrinoids, tetrapyrroles, and similar compounds)

P-102 Carbohydrate nomenclature

P-103 Amino acids and peptides

P-104 Cyclitols

P-105 Nucleosides

P-106 Nucleotides

P-107 Lipids

**P-100 INTRODUCTION**

In the field of natural products, three levels of nomenclature are recognized. A new compound, isolated from a natural source, is generally given a 'trivial' name. By common usage, these trivial names are commonly related to the biological origin of the material, but frequently not in a rational way, since the available structure is not known with great detail. These trivial names are considered to be ephemeral and replaced for chemical purposes by names describing the skeleton, the characteristic groups, and the organyl substituent groups.

When the full structure is known, a 'systematic name' may be generated in accordance with Rules described in Chapters P-1 through P-9 of these recommendations. However, this name may be too cumbersome to be continually inserted into the text of a scientific paper. To overcome this difficulty and show the close similarity to related compounds, a 'semisystematic name' can be formed. Preferred IUPAC names (PINs) are not identified for the compounds in this Chapter. The choice between a semisystematic name and a systematic name will be made in cooperation with the IUPAC-IUBMB Joint Commission on Biochemical Nomenclature and will appear in a future publication.

Semisystematic names are based on specific parent structures, generally including the configuration, that can later on be used to describe a compound fully by using the rules of systematic nomenclature. There are two general types of semisystematic parent structures used for naming natural products and related compounds:

(a) parent hydrides, i.e., structures that do not have terminal heteroatoms or functional groups and therefore consist only of skeletal atoms and hydrogen, for example, in steroid (ref.16), terpene, carotene (ref. 40), corrinoid (ref. 45), tetrapyrrole (ref. 17), lignan and neolignan (ref. 46), and alkaloid nomenclature. This type of semisystematic parent is analogous to the parents described in Chapter P-2 and is treated in the same manner to generate complete names;

(b) functional parents, which are analogous to the functional parents described in Section P-34, and used in amino acid and peptide (ref. 18), carbohydrate (ref. 27), cyclitol (ref. 39), nucleoside and nucleotide (ref. 47), and lipid (ref. 48) nomenclature; they have characteristic groups implied in their name, and can be modified by specific rules and by methods used in systematic nomenclature.

Section P-101 describes the rules to form trivial names and semisystematic names used as parent hydrides, and those related to their skeletal transformation and functionalization for naming alkaloids, steroids, terpenes, and some related compounds. Section P-102 describes the rules for naming carbohydrates, P-103 deals with the nomenclature of amino acids and peptides, P-104 describes the nomenclature for cyclitols, P-105 and P-106 deal with nucleosides and nucleotides, and finally P-107 discusses the nomenclature of lipids. If difficulties are encountered, consultation of the full publications may be necessary, as indicated in each Section.

**P-101 NOMENCLATURE FOR NATURAL PRODUCTS BASED ON PARENT HYDRIDES,  
(ALKALOIDS, STEROIDS, TERPENES, CAROTENES, CORRINOIDS, TETRAPYRROLES, AND SIMILAR  
COMPOUNDS)**

This Section is based on the recent publication 'Revised Section F: Natural Products and Related Compounds, IUPAC Recommendations 1999' and the additional document 'Corrections and Modifications 2004' (ref. 9).

- P-101.1 Biologically based trivial names
- P-101.2 Semisystematic nomenclature for natural products (stereoparent hydrides)
- P-101.3 Skeletal modifications of parent structures
- P-101.4 Replacement of skeletal atoms
- P-101.5 Addition of rings and ring systems
- P-101.6 Modification of the degree of hydrogenation of parent structures
- P-101.7 Derivatives of parent structures
- P-101.8 Further aspects of configurational specification

## **P-101.1 BIOLOGICALLY BASED TRIVIAL NAMES**

**P-101.1.1** When a compound is isolated from a natural source and a trivial name is required, the name should be based, whenever possible, on the family, genus, or species name of the biological material from which it has been isolated. If appropriate, the class or order might also be used for the name of a compound that occurs in a number of related families.

**P-101.1.2** The ending 'une' or, for euphonic reasons, 'iune' is used to indicate that the trivial name it terminates describes a compound of unknown structure.

## **P-101.2 SEMISYSTEMATIC NOMENCLATURE FOR NATURAL PRODUCTS (stereoparent hydrides)**

- P-101.2.0 Introduction
- P-101.2.1 General guidelines for choosing a parent structure
- P-101.2.2 Structural features allowed for parent structures
- P-101.2.3 Numbering of parent structures
- P-101.2.4 Identification of individual rings
- P-101.2.5 Atomic connector, terminal segment and bond connector
- P-101.2.6 Stereochemical configuration of parent structures
- P-101.2.7 Semisystematic names of recommended fundamental parent structures

### **P-101.2.0 Introduction**

Many naturally occurring compounds belong to well defined structural classes, each of which can be characterized by a set of parent structures that are closely related structurally, that is, each can be derived from a fundamental parent structure by one or more defined operations used in systematic substitutive nomenclature (see P-13).

As soon as the structure of a simple new natural product has been fully determined, the trivial name should be abandoned in favor of a systematic name formed by the Rules prescribed in Chapters P-1 through P-9 for systematic nomenclature of organic compounds. For a more complicated structure, an existing semisystematic name listed in P-101.2.7 is used to fully name the compound. If a previously known parent structure cannot be found, a new parent structure is formed and numbered as follows. To form and number a new parent structure, the procedure described in the following subsections is followed.

### **P-101.2.1 General guidelines for choosing a parent structure**

**P-101.2.1.1** A fundamental parent structure should reflect the basic skeleton (including nonterminal heteroatoms and hetero groups) that is common to most compounds of the class.

**P-101.2.1.2** Fundamental parent structures should be chosen so that as many natural products as possible can be derived from each by well defined operations and rules of the nomenclature of organic compounds.

**P-101.2.1.3** A fundamental parent structure should include as much configuration as possible that is common to the relevant class of natural products. Such parent structures are called 'stereoparents'.

### **P-101.2.2 Structural features allowed for parent structures**

The following rules are applicable to new parent structures. Existing parent structure names are considered as retained names if they do not follow the new rules (see Table 10.1).

**P-101.2.2.1** A fundamental parent structure should exceptionally include rings that are part of a characteristic group, such as a lactone or cyclic acetal.

**P-101.2.2.2** A fundamental parent structure should not contain terminal heteroatoms or characteristic groups (see P-101.2.1.1).

**P-101.2.2.3** A fundamental parent structure should contain acyclic hydrocarbon groups that occur in most of the compounds in the natural product class.

**P-101.2.2.4** A fundamental parent structure should be as nearly fully saturated or fully unsaturated in terms of maximum number of noncumulative double bonds (mancude rings), while still representing the level of saturation (or unsaturation) of as many related compounds as possible.

**P-101.2.2.5** A semisystematic name for a fundamental parent structure should be derived, as far as possible, from a trivial name formed in accordance with P-101. The endings to be used in place of 'une' or 'iune' must be assigned as follows:

- (a) 'ane', if the entire stereoparent hydride is fully saturated;
- (b) 'ene', if the cyclic or the main chain of the acyclic part contains the maximum number of noncumulative double bonds;
- (c) 'arane', if, in an otherwise fully saturated parent structure, one or more individual mancude rings is present.

Existing names of parent structures in which endings are different from those indicated above, for example morphinan and ibogamine, are exceptions and treated as retained names.

**P-101.2.2.6** Indicated hydrogen, as defined in P-14.7, P-25.7 and P-58.2, is used to describe isomers of fundamental parent structures.

### **P-101.2.3** Numbering of parent structures

**P-101.2.3.1** A numbering pattern established among a group of structurally related natural products is used for numbering the skeletal atoms of the fundamental parent structure, providing that all skeletal atoms have been included in the numbering system.

**P-101.2.3.2** If no numbering pattern has been become established among the members of a group of structurally related natural products, the fundamental parent structure is numbered according to the following guidelines:

- (a) examine the skeleton to identify the senior ring or ring system, according to P-44. The locant '1' is assigned to the atom of the senior ring system whose locant would be '1' according to systematic numbering for that particular ring or ring system;
- (b) assign all skeletal atoms of the senior ring system consecutive Arabic numbers, including atoms of fusion positions in fused ring systems, beginning with the locant '1', and following the path prescribed for that particular type of ring or ring system;
- (c) number acyclic substituents to skeletal atoms of ring components or connecting acyclic structures each in its entirety, including branches, in order of the increasing value of the locant of the skeletal atom to which it is attached;
- (d) number skeletal atoms of acyclic connections to other ring or ring systems, if any, consecutively beginning with the atom next to the senior ring system, followed by the skeletal atoms of the other rings or ring systems as prescribed in (b) above; if two or more acyclic connections to other rings or ring systems are present, the one attached to the senior ring or ring system at the lowest numbered position is numbered first, then the ring attached to it, followed by the acyclic connector at the next lower position of the senior ring or ring system, etc.;
- (e) number the larger group, in terms of the number of skeletal atoms, between two groups at a geminal disubstituted position first; if there is still a choice, alphanumerical order is followed (Rule P-14.5). If the two groups are then identical and attached to a stereoparent structure properly drawn (see Appendix 3), the group that is stereochemically ' $\alpha$ ' (according to P-101.2.6) is numbered first; if the two groups are identical and attached to an acyclic terminal double bond, the group '*trans*' to the main chain is numbered first, as described in the carotenoid recommendations (Rule 12.4 in ref. 40).

### **P-101.2.4** Identification of individual rings

Since locants are used to describe skeletal modifications, as indicated in P-101.3, the identification of individual rings by letters A, B, C, etc., used in the past is no longer recommended, except for the rather special case of the removal of a terminal ring (see P-101.3.6). Nevertheless, to provide continuity with the use of this system, names using letters to identify rings are given where appropriate, but are no longer recommended.

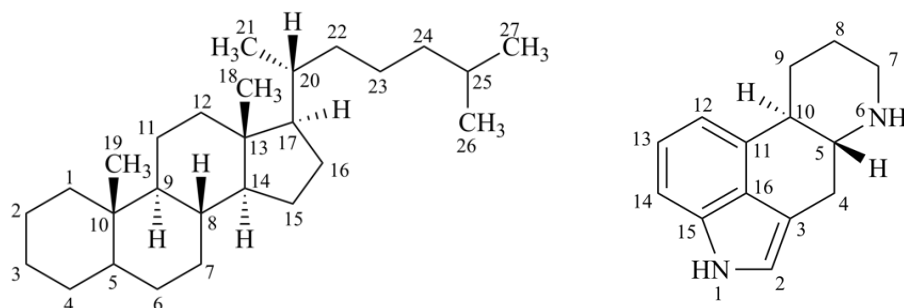
### **P-101.2.5** Atomic connector, terminal segment, and bond connector

For nomenclature purposes, the fundamental parent structures are described by specific arrangements of atoms or groups of atoms called 'atomic connectors', 'terminal segments' and 'bond connectors', that must be taken into consideration in accordance with the additive or subtractive operations modifying a fundamental parent structure.

An 'atomic connector' is a chain of homogeneous skeletal atoms of the same element connecting any combination of bridgehead or ring junction atoms, rings, or ring systems (i.e. ring assemblies), substituted skeletal atoms in parent

structure, or heteroatoms. A 'terminal segment' of a skeletal structure is an acyclic portion of homogeneous skeletal atoms connected at only one end by the features of structure that terminate atomic connectors. A 'bond connector' is a connection between any combination of bridgehead or ring junction atoms, rings, or ring systems (i.e. ring assemblies), substituted skeletal atoms, or heteroatoms. The structures below illustrate atomic connectors, bond connectors, and terminal segments. The use of these terms is further illustrated in P-101.3.1 in relation to the removal of skeletal atoms denoted by the prefix 'nor'.

Examples:



cholestane

ergoline

atomic connectors:

terminal portions:

in cholestane: 1-4, 6-7, 11-12, 15-16, and 22-24

in cholestane: 18, 19, 21, 26, and 27

in ergoline: 2, 4, 7-9, and 12-14

in ergoline: none

bond connectors:

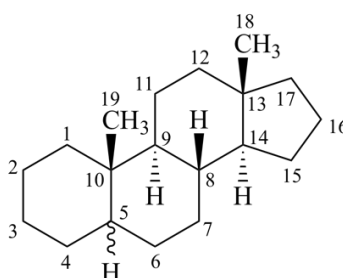
In cholestane: 5-10, 8-9, 8-14, 9-10, 13-14, 13-17, and 17-20

In ergoline: 1-15, 3-16, 5-6, 5-10, 10-11, 11-16, and 15-16

#### P-101.2.6 Stereochemical configuration of parent structures

The name of a fundamental parent structure usually implies the absolute configuration of all chirality centers and the configuration of double bonds, when applicable, without further specification. All chirality must be defined so that for example with a steroid the stereochemistry at 'C-5', when relevant, is indicated by  $\alpha$ ,  $\beta$  or  $\xi$ . When a planar or quasi planar system of rings is denoted as a projection, as in this recommendation, an atom or group attached to the ring is called ' $\alpha$ ' if it lies below or ' $\beta$ ' if it lies above the plane of the paper. Use of this system requires the orientations of structure as given in the examples used to exemplify the various rules and in Appendix 3. In the example below, the implied configurations shown define the attached hydrogen atoms and methyl groups at positions '8', '10', and '13' as ' $\beta$ ', and at positions '9' and '14' as ' $\alpha$ '; here, the configuration of the hydrogen atom at position 5 is not known and thus the orientation is ' $\xi$ ' ( $\xi$ ), denoted by a wavy line in the formula. The stereodescriptors ' $\alpha$ ', ' $\beta$ ', and ' $\xi$ ' used to describe implicit or indicated configuration are cited before the name of the fundamental parent structure without parentheses.

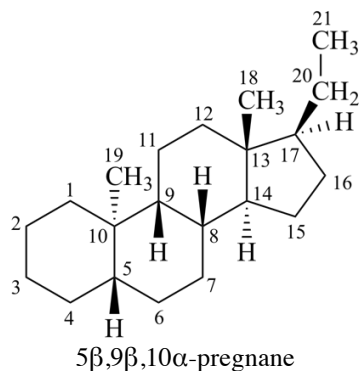
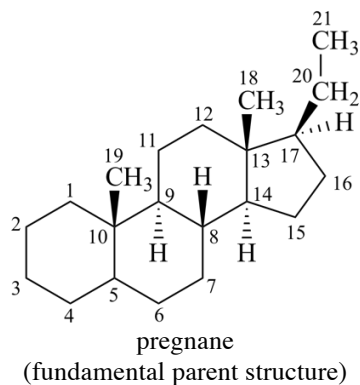
The ' $\alpha/\beta$ ' symbolism is used as defined above and extended in the following way to express different aspects of the configuration of modified fundamental parent structures.



#### P-101.2.6.1 Configurations that are different from those in the parent structure

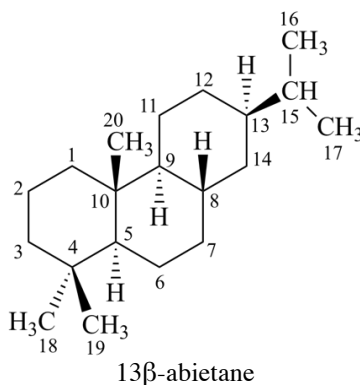
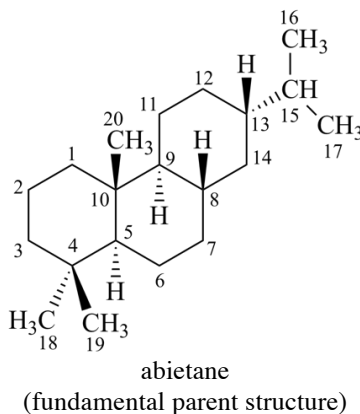
**P-101.2.6.1.1** At chirality centers, the ' $\alpha/\beta$ ' system is used as described in IUPAC- IUBMB recommendations for the nomenclature of steroids (ref. 16). Each chirality center is described by the stereodescriptor ' $\alpha$ ', ' $\beta$ ', or ' $\xi$ ' to indicate a configuration that must be specified and those that are inverted. The symbols ' $\alpha$ ', ' $\beta$ ', or ' $\xi$ ', preceded by the appropriate locants, are placed immediately at the beginning of the name of the fundamental parent structure. In the following examples, configuration at 'C-5' must be specified; configurations at bridgeheads 'C-9' and 'C-10' are inverted when compared with those of the fundamental parent structure. This method is preferred to the alternatives described in P-101.2.6.1.2.

Example:



A change in configuration of a nonbridgeheaded side chain that is part of the parent is denoted by the method specified for 'C-17' of steroids (see 3S-5.2, ref. 16), where ' $\alpha$ ' or ' $\beta$ ' refers to the side chain itself and not to the hydrogen atom in the same position.

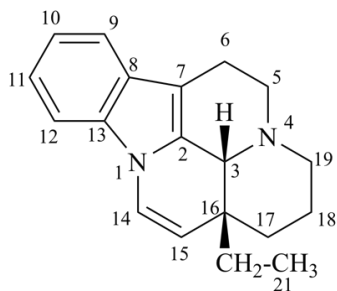
Example:



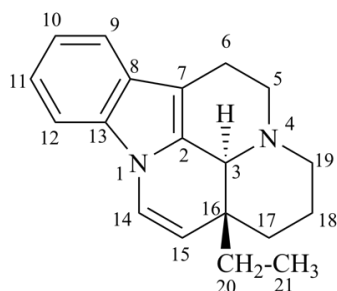
**P-101.2.6.1.2** Configurational inversion at one of stereogenic centers whose configuration is implied or stated in the name of the fundamental parent structure can be indicated by the italicized prefix '*epi*' (derived from 'epimer') placed at the front of the name of the parent structure and prefixed by the locant of the affected atom.

The name 13 $\beta$ -abietane, described above in P-101.2.6.1.1, can also be named 13-*epi*-abietane.

Example:



eburnamenine  
(fundamental parent structure)



3-*epi*-eburnamenine  
3 $\alpha$ -eburnamenine

#### P-101.2.6.1.3 The stereodescriptors ‘*R*’ and ‘*S*’

The stereodescriptors ‘*R*’ and ‘*S*’ are used to describe the absolute configuration not specified by the ‘ $\alpha/\beta$ ’ system described above, in accord with the CIP priority system and the rules and conventions described in Chapter 9. The stereodescriptors ‘*R*’ and ‘*S*’ are also used when a ring is opened and two chirality centers are created, one of which may rotate, as described for vitamin D in P-101.8.4.

P-101.2.7 Semisystematic names of recommended parent structures are listed in Table 10.1. Structures are shown in Appendix 3.

**Table 10.1 Names of fundamental stereoparent structures (nonlimiting)**

#### (a) alkaloids

aconitane	emetan	oxyacanthan
ajmalan	ergoline	pancracine
akuammilan	ergotaman	rheadan
alstophyllan	erythrinan	rodiasine
aporphine	evonimine	samandarine
aspidofractinine	evonine	sarpagan
aspidospermidine	formosanan	senecionan
atidane	galanthamine	solanidane
atisine	galanthan	sparteine
berbaman	hasubanan	spirosolane
berbine	hetisan	strychnidine
cephalotaxine	ibogamine	tazettine
cevane	kopsan	tropane
chelidonine	lunarine	tubocuraran
cinchonan	lycopoladane	tubulosan
conanine	lycorenan	veratraman
corynan	lythran	vincaleukoblastine
corynoxan	lythranidine	vincane
crinan	matridine	vobasan
curan	morphinan	vobtusine
daphnane	nupharidine	yohimban
dendrobane	ormosanine	
eburnamenine	18-oxayohimban	

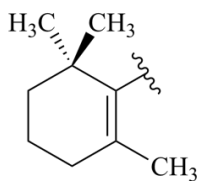
(b) **steroids**

androstane	cholestane	gorgostane
bufanolide	ergostane	poriferastane
campestande	estrane	pregnane
cardanolide	furostan	spirostan
cholane	gonane	stigmastane

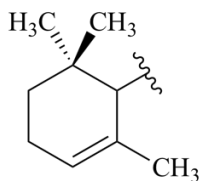
(c) **terpenes** (all are parent hydrides except retinal)

abietane	drimane	menthane ( <i>p</i> -isomer)
ambrosane	eremophilane	oleanane
aristolane	eudesmane	ophiobolane
atisane	fenchane	picrasane
beyerane	gammacerane	pimarane
bisabolane	germacrane	pinane
bornane	gibbane	podocarpane
cadinane	grayanotoxane	protostane
carane	guaiane	retinal
$\beta$ , $\phi$ -carotene*	himachalane	rosane
$\beta$ , $\psi$ -carotene*	hopane	taxane
$\epsilon$ , $\kappa$ -carotene*	humulane	thujane
$\epsilon$ , $\chi$ -carotene*	kaurane	trichothecane
caryophyllane	labdane	ursane
cedrane	lanostane	
dammarane	lupane	

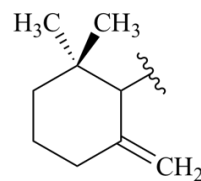
\* Four different carotenes are exemplified; there are 28 carotene parent structures derived from all permutations of the seven following end groups:



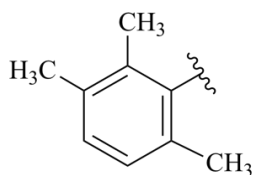
$\beta$  (beta)



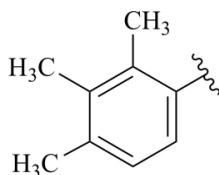
$\epsilon$  (epsilon)



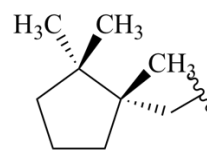
$\gamma$  (gamma)



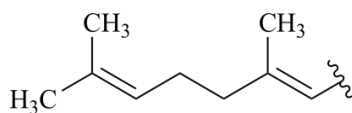
$\phi$  (phi)



$\chi$  (chi)



$\kappa$  (kappa)



$\psi$  (psi)

(d) **Miscellaneous** (all are parent hydrides except cepham and penam)

21 <i>H</i> -biline	isoflavan	penam
cepham	lignane	porphyrin
corrin	neoflavan	prostane
flavan	neolignane	thromboxane

## P-101.3 SKELETAL MODIFICATIONS OF PARENT STRUCTURES

### P-101.3.0 Introduction

P-101.3.1 Removal of skeletal atoms without affecting the number of rings

P-101.3.2 Addition of skeletal atoms without affecting the number of rings

P-101.3.3 Bond formation

P-101.3.4 Bond cleavage

P-101.3.5 Bond migration

P-101.3.6 Removal of a terminal ring

P-101.3.7 Combination of the prefixes 'cyclo', 'seco', 'apo', 'homo', and 'nor'

### P-101.3.0 Introduction

The skeleton of parent structures can be modified in many ways, contracted, expanded, or rearranged by using operations described in P-13. These operations are denoted by specific nondetachable prefixes that are added to the name of the parent structure. Changes affecting the configuration must be shown as indicated in P-101.2.6. In natural product nomenclature, the number of operations is not subject to limitations.

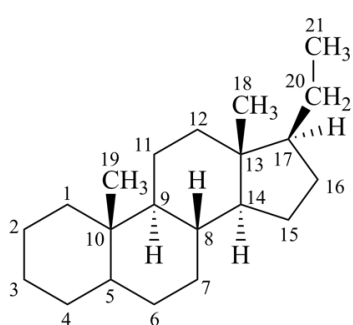
This Section supersedes the Section F Rules (ref. 9) and Rules A-71 through A-75 related to terpene hydrocarbons as prescribed in the 1979 Recommendations (ref. 1).

### P-101.3.1 Removal of skeletal atoms without affecting the number of rings

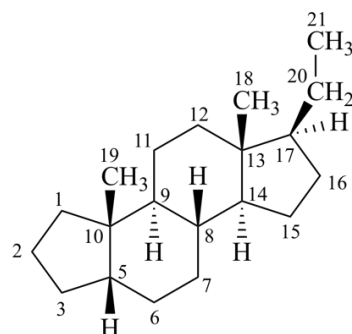
**P-101.3.1.1** The removal of an unsubstituted skeletal atom, saturated or unsaturated, from a ring or of an unsubstituted skeletal atom from a saturated acyclic portion of a fundamental parent structure with its attached hydrogen atom(s) is described by the nondetachable prefix 'nor'; the loss of two or more such skeletal atoms is indicated by the usual numerical multiplicative prefixes 'di', 'tri', etc. added before 'nor'.

The position of the skeletal atom that is removed is denoted in all cases by its locants in the numbering of the fundamental parent structure. Although, because the locant of each skeletal atom removed is cited, an unambiguous name can be generated by the removal of any skeletal atom, carbon atom or heteroatom, it is traditional to remove skeletal atoms with the highest possible locant in an atomic connector in a cyclic portion of the skeletal structure. In carotenoids, as an exception, the locant attached to 'nor' is the lowest possible (see Rule Carotenoid 5.1, ref. 40).

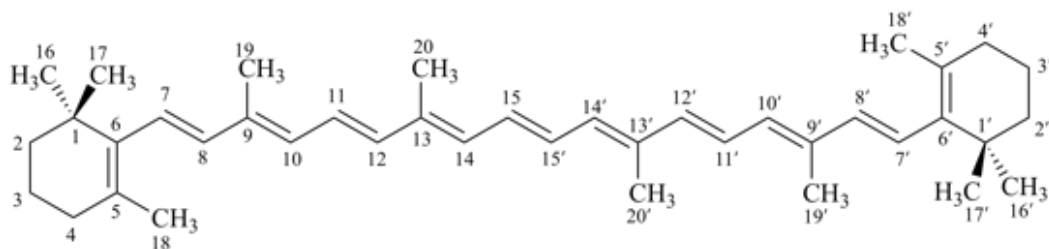
Examples:



pregnane  
(fundamental parent structure)

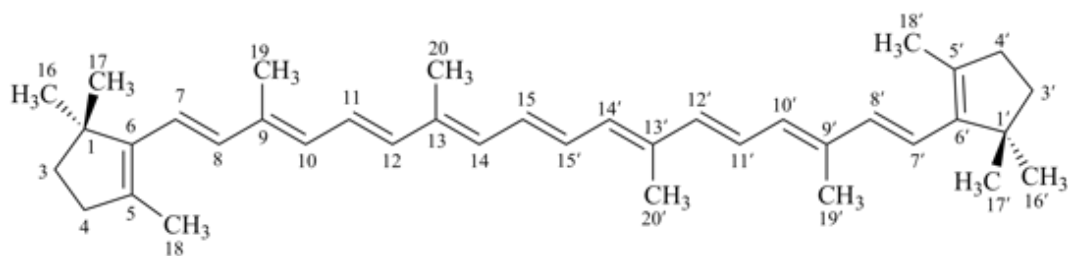


4-nor-5 $\beta$ -pregnane



$\beta,\beta$ -carotene  
(fundamental parent structure)

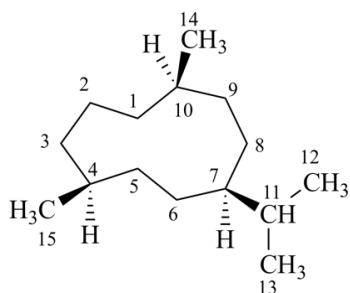




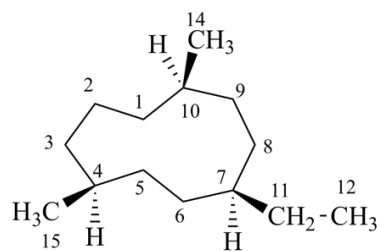
2,2'-dinor- $\beta,\beta$ -carotene

In an acyclic portion of a skeletal structure, the skeletal atom removed preferably is the one of an acyclic atomic connector or a terminal segment nearest to the free end of this acyclic portion (this is done in order to maintain as far as possible the numbering of structural features of the compound and of compounds derived from it).

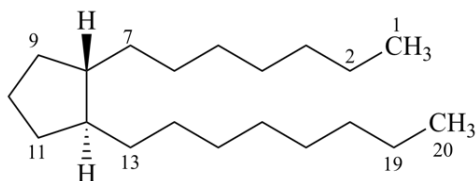
Examples:



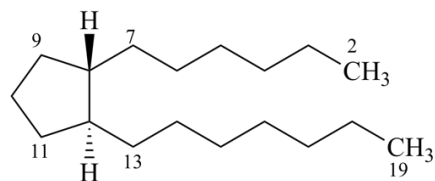
germacrane  
(fundamental parent structure)



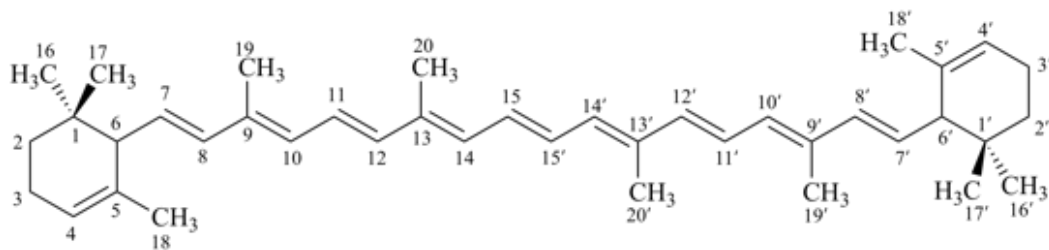
13-norgermacrane  
(1*R*,4*s*,7*S*)-4-ethyl-1,7-dimethylcyclodecane  
(position 1 is position 10 for germacrane)



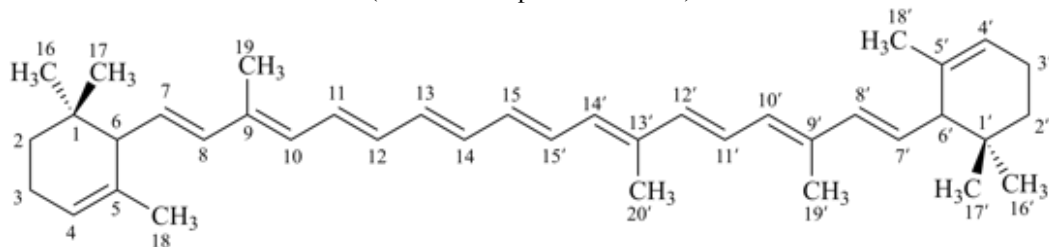
prostan  
(fundamental parent structure)



1,20-dinorprostan  
(1*S*,2*S*)-1-heptyl-2-hexylcyclopentane



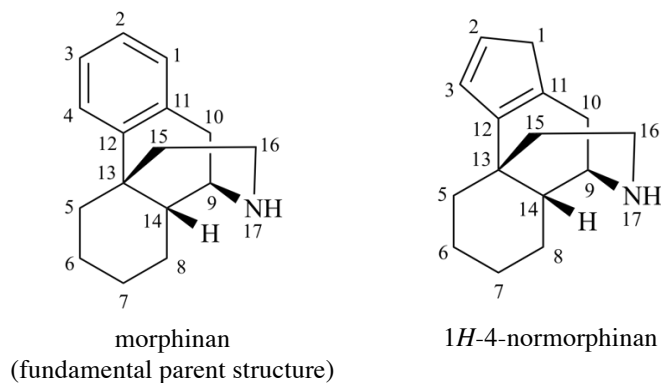
$\epsilon,\epsilon$ -carotene  
(fundamental parent structure)



20-nor- $\epsilon,\epsilon$ -carotene  
(see P-14.3.5 for use of unprimed and primed locants)

**P-101.3.1.2** When the removal of an unsaturated skeletal atom from a mancude ring (containing the maximum number of noncumulative double bonds) in the fundamental parent structure results in the creation of a saturated ring position, this position is described by indicated hydrogen (see P-14.7). In names, the symbol *H*, denoted by the appropriate locant, is cited at the beginning of the name modified by the nondetachable prefix.

Example:



**P-101.3.2** Addition of skeletal atoms without affecting the number of rings

**P-101.3.2.1** The addition of a methylene ( $-\text{CH}_2-$ ) group between two skeletal atoms of a fundamental parent structure is described by the nondetachable prefix ‘homo’; the addition of two or more methylene groups is indicated by the numerical multiplicative prefixes ‘di’, ‘tri’, etc. Positions of the inserted methylene groups in the modified fundamental parent structure are indicated by the locants of the added methylene groups cited in front of the prefix ‘homo’, preceded by multiplicative prefixes when required.

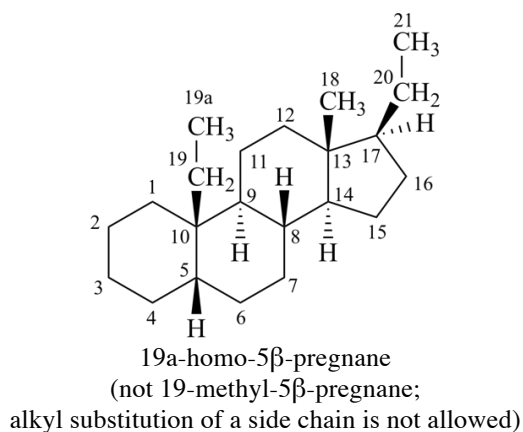
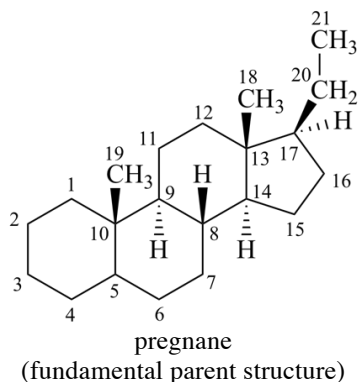
The assignment of the locants to an added methylene group depends on whether it is considered to be inserted into an atomic connector or terminal acyclic portion or into a bond connector.

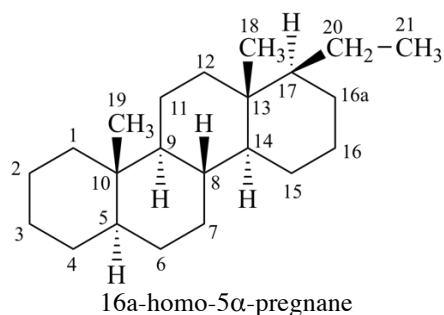
**P-101.3.2.2** Numbering of additional skeletal atoms

**P-101.3.2.2.1** Methylene groups inserted into an atomic connector or into a terminal segment are identified by adding a letter ‘a’, ‘b’, etc., to the locant of the highest numbered skeletal atom of the atomic connector or terminal portion consistent with the location of double bonds remaining in the structure. If there are equivalent atomic connectors, the highest atomic connector is chosen, and the methylene group is inserted after the highest numbered skeletal atom in that connector.

Addition of acyclic side chains or extension of terminal portions of a side chain already attached to the stereoparent hydride may also be done by principles of substitutive nomenclature. The added substituent(s) are numbered as described above for ‘homo’ atoms.

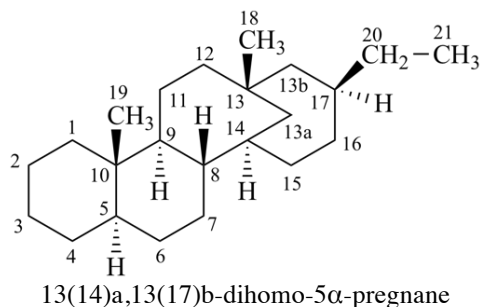
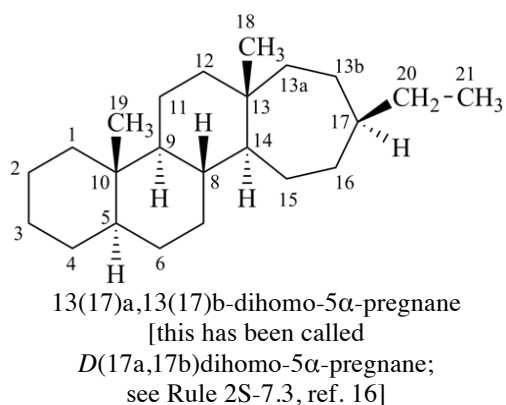
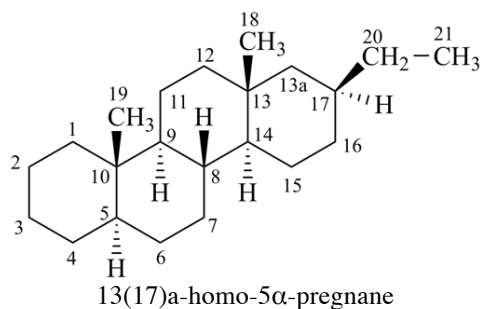
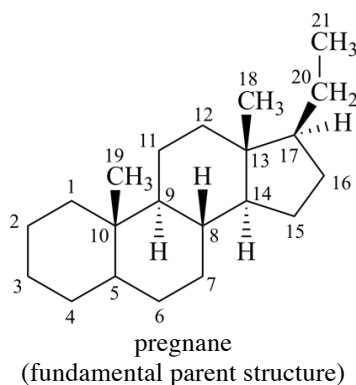
Examples:





**P-101.3.2.2.2** Methylene groups inserted into a bond connector are identified by citing both locants of the skeletal atoms terminating the bond connector enclosing the second (higher) number in parentheses, followed by a letter 'a', 'b', etc. according to the number of methylene groups

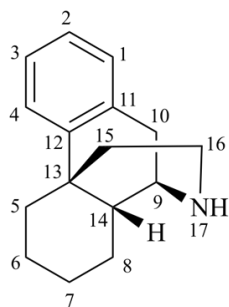
Examples:



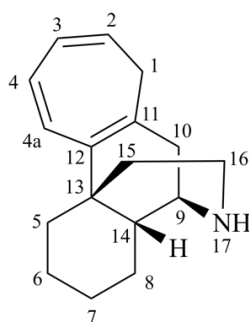
**P-101.3.2.2.3** The insertion of a methylene group into a mancude ring or ring system (that contains the maximum number of noncumulative double bonds) or into a system of conjugated double bonds may create a saturated ring position that is described by 'indicated hydrogen' (see P-14.7 and P-58.2). The position of the methylene group is

prescribed by P-101.3.2.2.2, even though the saturated ring position may be elsewhere in the unsaturated ring system as denoted by the appropriate locant for the indicated hydrogen; this is a change for names of the homoporphyrins (see ref. 17, Rule TP-5.1). Two tautomeric forms, (A) and (B), are represented below and specifically numbered and named.

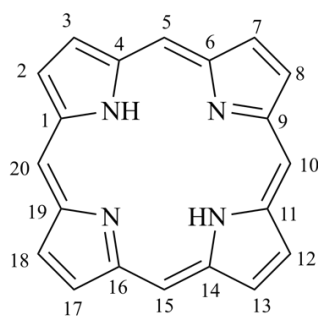
Examples:



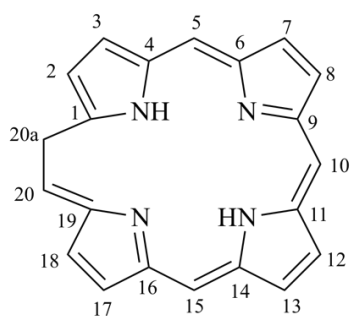
morphinan  
(fundamental parent structure)



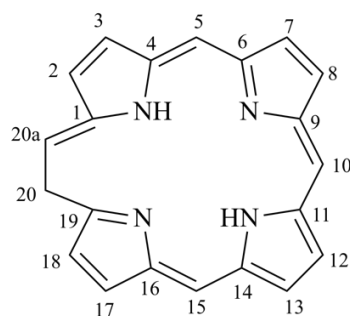
1H-4a-homomorphinan



porphyrin  
(fundamental parent structure)



(A) 20aH-20a-homoporphyrin



(B) 20H-20a-homoporphyrin

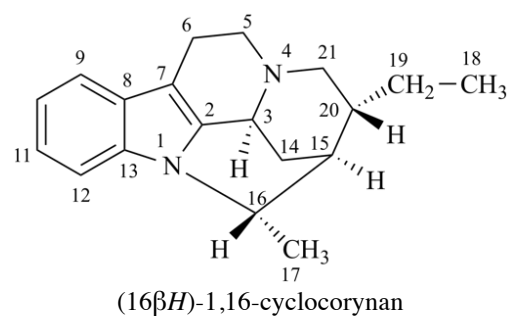
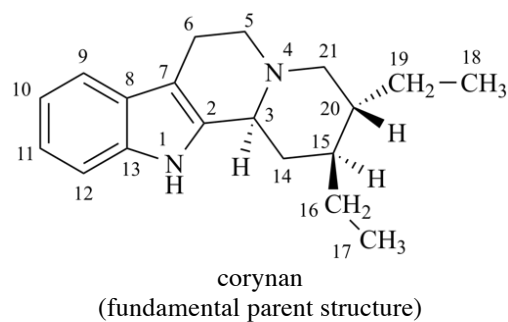
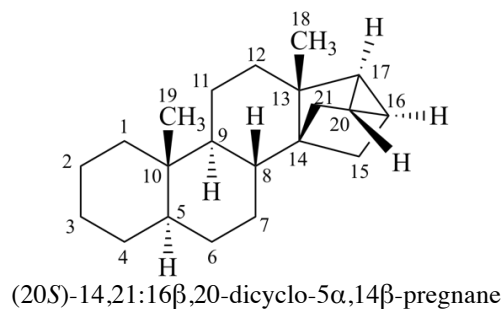
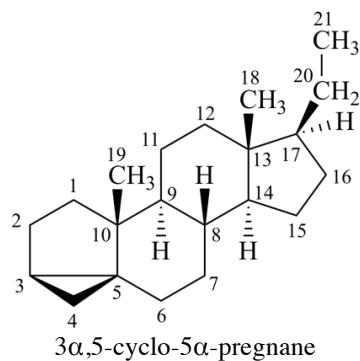
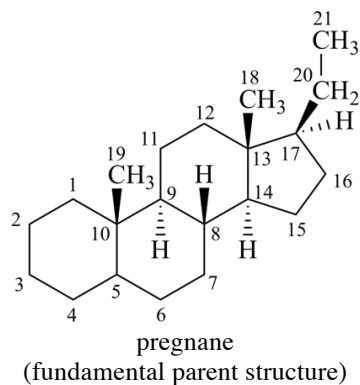
### P-101.3.3 Bond formation

The creation of an additional ring by a conjunctive operation (see P-13.5.3) by means of a direct link between any two atoms of a parent structure is described by the nondetachable prefix 'cyclo' (not italicized) preceded by the locants of the skeletal atoms so connected. When necessary, the configuration created by the new bond is denoted by  $\alpha$ ,  $\beta$ , or  $\xi$  descriptors in accord with P-101.2.6, or by describing the configuration of the hydrogen atom in accordance with P-101.2.6.1.3.

Configurations of the fundamental parent structure are retained. New configurations of the ring atoms having one hydrogen atom still present are indicated by the ' $\alpha/\beta$ ' stereodescriptors as described in P-101.2.6, or, if necessary, by the sequence rule method (*R/S*). The projection of the hydrogen atom below, ' $\alpha$ ', or above, ' $\beta$ ', the plane of the ring system is indicated by the appropriate symbol and a capital italic letter *H* following the locant of the ring atom in the

structure, all enclosed in parentheses, and cited before the appropriate prefix, in this case 'cyclo' (see P-101.3.5.1 for the prefix 'abeo'). This method of citation differs from that used in the Steroid Rules (Rule 3S-7.5, ref. 16).

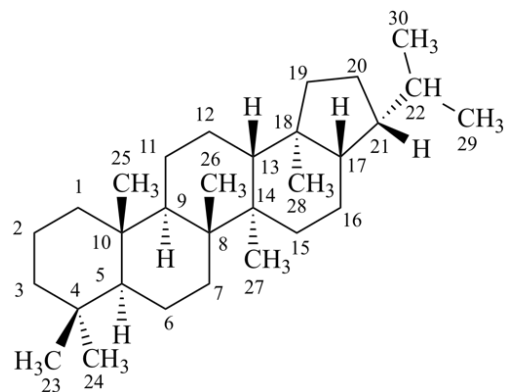
Examples:



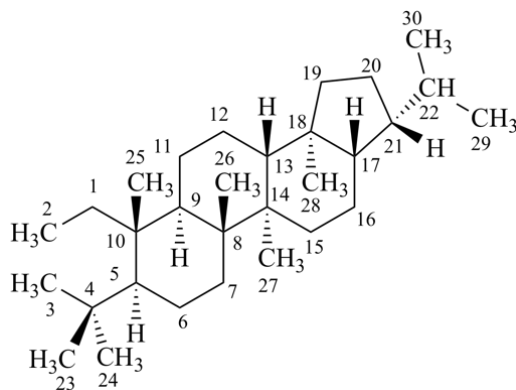
#### P-101.3.4 Bond cleavage

**P-101.3.4.1** Cleavage of a ring bond (saturated or unsaturated) with the addition of the appropriate number of hydrogen atoms at each new terminal group thus created is indicated by the prefix 'seco' (not italicized) and the locants of the cleaved bond. The original numbering is retained.

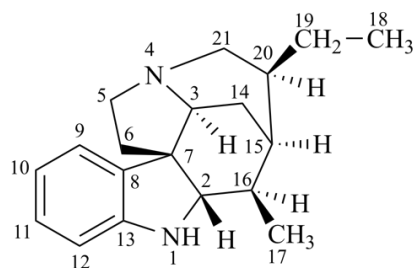
Examples:



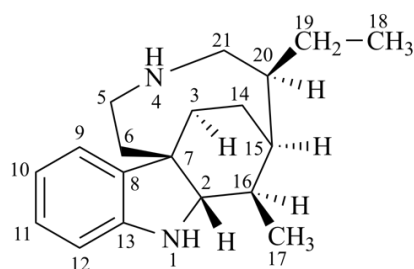
hopane  
(fundamental parent structure)



2,3-secohopane



curan  
(fundamental parent structure)



3,4-secocuran

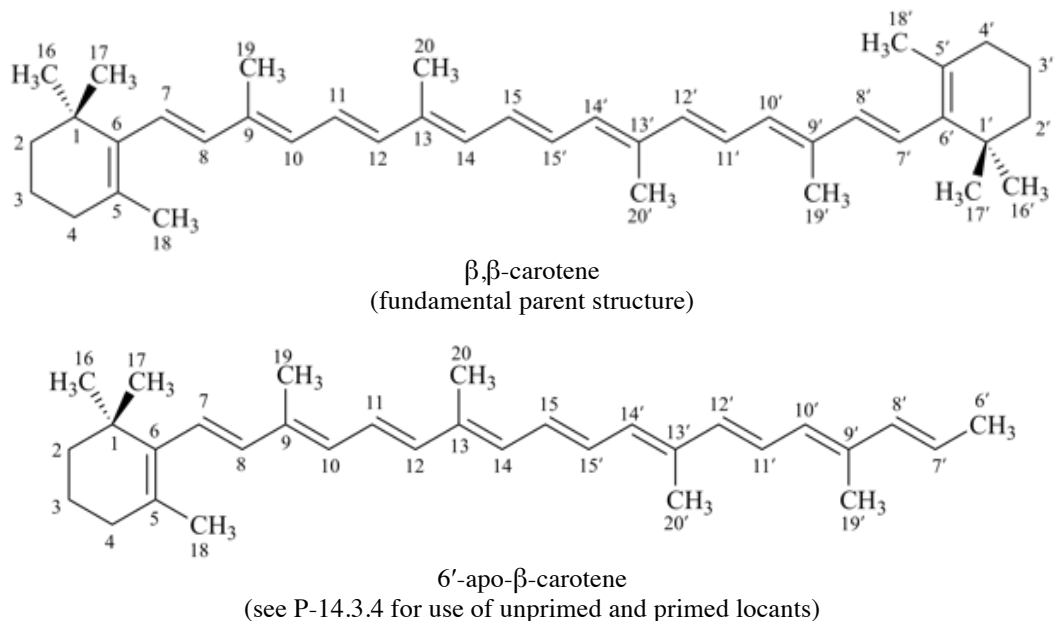
#### P-101.3.4.2 The prefix 'apo'

The unitalicized prefix 'apo' preceded by a locant is used to indicate removal of all of a side chain of a fundamental parent structure beyond the skeletal atom corresponding to that locant. Removal of two or more side chains is indicated by the prefixes 'diapo', 'triapo', etc., preceded by the required locants. Numbering of the skeletal atoms of the fundamental parent structure is retained in the resulting fragment.

The following procedure is used only in carotenoid nomenclature (see ref. 40, Rule Carotenoid 10). The unitalicized prefix 'apo', preceded by a locant, is used to indicate that all of the molecule beyond the carbon atom corresponding to that locant has been replaced by a hydrogen atom. A side-chain methyl group is not considered to be 'beyond' the carbon atom to which it is attached. Removal of fragments from both ends of a molecule is indicated by using the numerical multiplying prefix 'di', preceded by two locants. Numbering of the skeletal atoms in the parent structure is retained in the resulting fragment.

The prefix and its locant immediately precede the parent name unless the locant associated with the prefix 'apo' is greater than 5, in which case there is no need to give a Greek letter end-group designation for that end of the molecule.

Example:



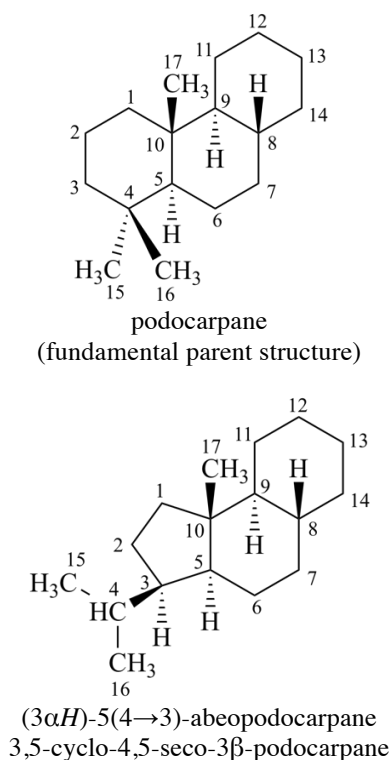
### P-101.3.5 Bond migration

Parent structures that are not simple derivatives of accepted fundamental parents, but may be considered to arise from such parents by bond migration of one or more bonds, may be named by the following method.

**P-101.3.5.1** The nondetachable prefix '*x(y→z)*-abeo' designates the migration from one end of a single bond from its original position in a fundamental parent structure to another position. In the prefix, '*x*' is the locant of the stationary, i.e. unchanged, end of the migrating bond; '*y*' is the locant of the position of the moving end of the migrating bond in the parent structure; and '*z*' is the locant of the position of the moving end in the final structure. The numbering of the initial fundamental parent structure is retained.

Previously the prefix 'abeo' was italicized (Rule F-4.9, ref. 1; Rule R-1.2.7.1, ref. 2). For consistency with the other modifying prefixes it is now recommended that a regular roman font be used.

Example:

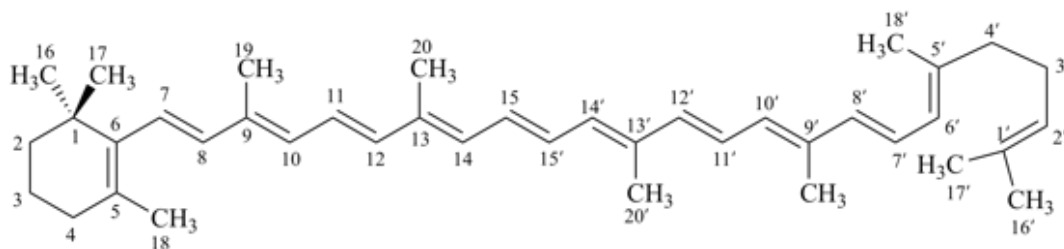


The 'abeo' nomenclature described in this rule is permissive, not compulsory. It is most suitable for use in discussions on reaction mechanisms and biogenesis.

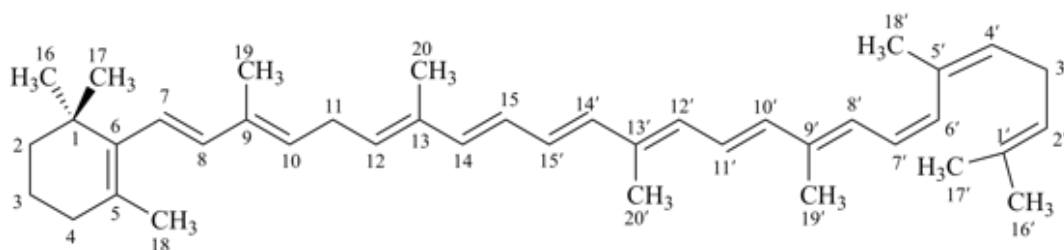
**P-101.3.5.2** The italic prefix '*retro*' preceded by a pair of locants is used to indicate a shift, by one position, of all single and double bonds of a conjugated polyene system delineated by the pair of locants; the conjugated polyene system cannot be part of a system of maximum number of noncumulative double bonds in a ring or ring system. The first locant is the skeletal atom that has lost a hydrogen atom and the second locant the one that has gained a hydrogen atom.

The descriptor '*retro*' is used in this manner only in carotenoid nomenclature (see Carotenoid Rule 9, ref.9).

Example:



$\beta,\psi$ -carotene (fundamental parent structure)

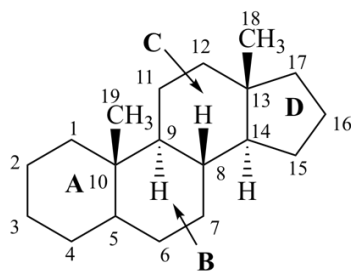


4',11-*retro*- $\beta,\psi$ -carotene  
(see P-16.9 for use of primed and unprimed locants)

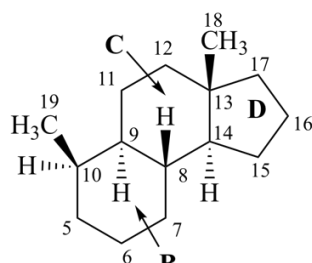
**P-101.3.6** Removal of a terminal ring.

The removal of a terminal ring from a parent structure with the addition of an appropriate number of hydrogen atoms at each junction with the adjacent ring is indicated by the nondetachable prefix '*des*' followed by the capital italic letter of the ring removed (see P-103.3.5.4 for use of '*des*' in peptide nomenclature). This is the only time that the capital letters are now used to identify rings in a parent structure. Stereochemistry implied by the name of the stereoparent structure remains the same, unless otherwise specified. Numbering of skeletal atoms of the parent structure is retained in the modified structure. This use of '*des*' is restricted to steroids.

Example:

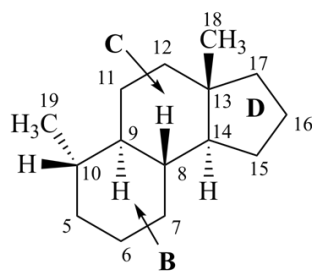


androstane (fundamental parent structure)



des-A-androstane





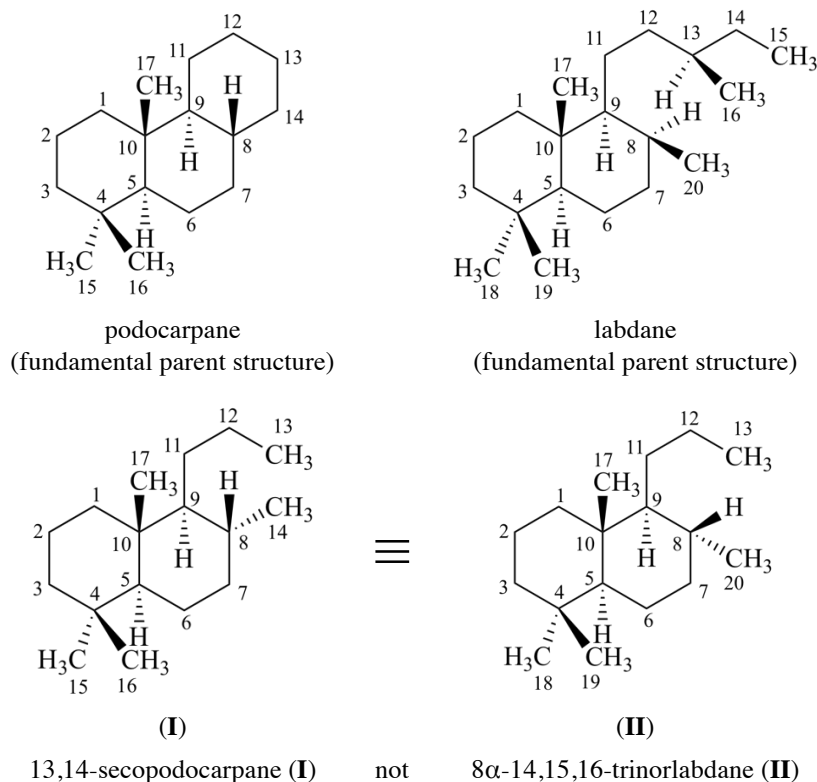
des-A-10 $\alpha$ -androstane

**P-101.3.7** Combination of the prefixes ‘cyclo’, ‘seco’, ‘apo’, ‘homo’, and ‘nor’

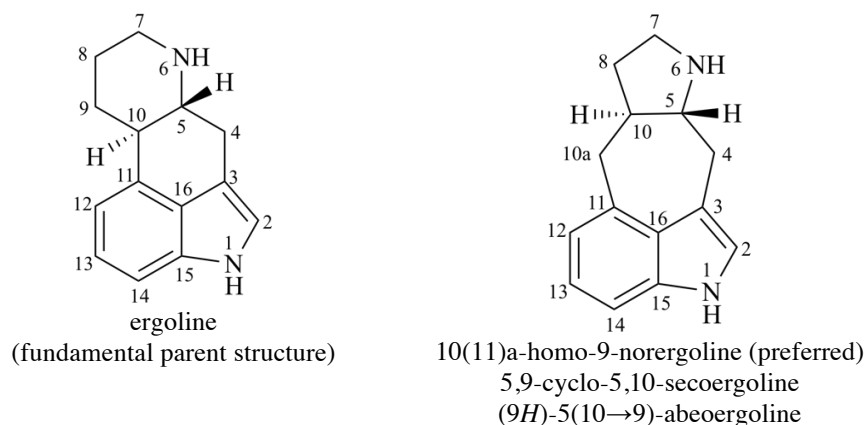
Modifications to a fundamental parent structure prescribed by the prefixes in the preceding recommendations (P-101.3.1 through P-101.3.4) may be combined to effect even more drastic changes in structure. The operation indicated by each prefix ‘cyclo’, ‘seco’, ‘apo’, ‘homo’, and ‘nor’ is applied to the fundamental parent structure sequentially as one ‘advances backward’, i.e. moves from right to left from the name of the fundamental parent structure.

**P-101.3.7.1** When different combinations of prefixes ‘cyclo’, ‘seco’, ‘apo’, ‘homo’, and ‘nor’ can be used to effect the same transformation in fundamental parent structure, the combination of choice must express the fewest number of operations. Both detachable (e.g. alkyl) and nondetachable (e.g. homo or nor) prefixes are considered as modifications but detachable prefixes are preferred. Dihomo, dinor, etc., are counted as two modifications each (see ref. 16, Rule 3S-6.3). When the number of operations is the same, the combination of homo/nor is preferred to cyclo/seco; choice between other combinations expressing the same number of operations is based on earlier alphabetical order of the prefixes.

Examples:



**Explanation:** Podocarpane may be used to generate the ‘seco’ compound with one operation; the same compound may be obtained from labdane but in three operations.



**P-101.3.7.2** The order of citation of combinations of structure modifying prefixes must avoid improper use of the prefixes as defined above or impossible situations when the corresponding operations are carried out in the manner prescribed above.

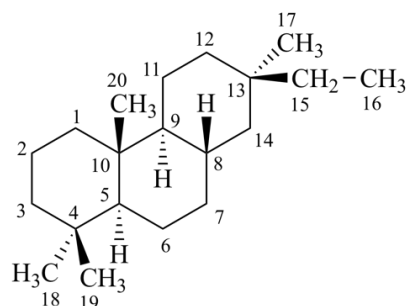
After satisfying P-101.3.7.1 and P-101.3.7.2, the nondetachable prefixes that indicate bond rearrangements ('cyclo' and 'seco') are cited, followed by those that indicate addition or removal of skeletal atoms ('homo' and 'nor'). If more than one of any of these operations is needed, they are cited in alphabetical order before the name of the fundamental parent structure. Multiplying prefixes denoting multiple operations of the same kind do not affect the order.

The preferred semisystematic name results from modifications by only two operations involving the prefixes 'cyclo', 'seco', 'apo', 'homo', and 'nor'. In general nomenclature, more than two operations are allowed. Names are formed by citing the bond rearrangement prefixes 'cyclo' and 'seco', first, (farthest from the parent structure) in that order from left to right, followed by the removal/addition prefixes 'homo' and 'nor', in that order from left to right, at the front of the name of the parent structure. Schematically this order is as follows:

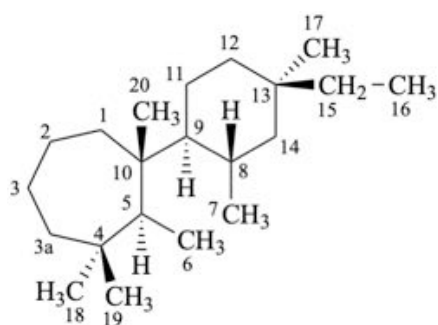
operation	bond rearrangement	addition/removal of skeletal atoms	parent structure
	cyclo, seco	apo, homo, nor	

Names in which the order of prefixes is cyclo/seco/apo/homo/nor are preferred to those denoted by the alphabetical order apo/cyclo/homo/nor/seco for the five prefixes.

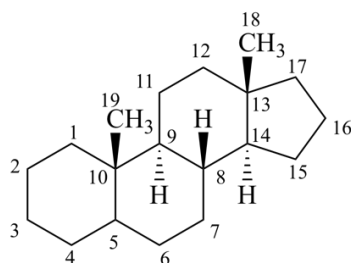
Examples:



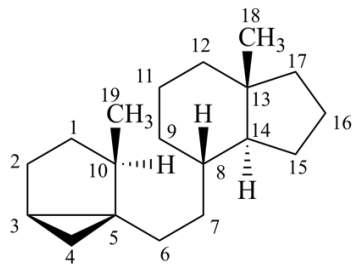
pimarane  
(fundamental parent structure)



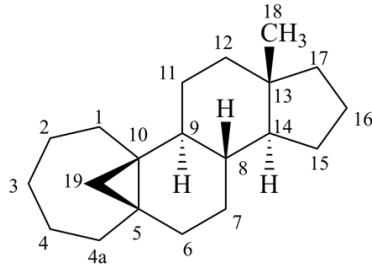
6,7-seco-3a-homopimarane (preferred)  
3a-homo-6,7-secopimarane



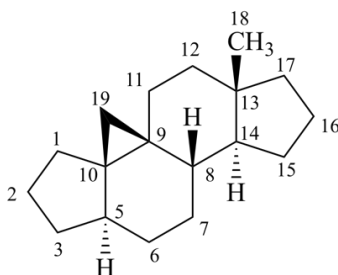
androstane  
(fundamental parent structure)



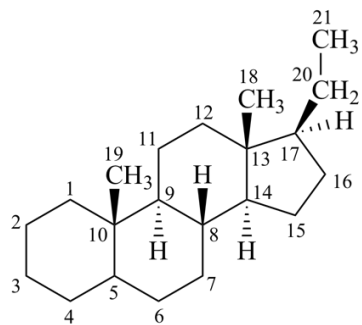
3 $\alpha$ ,5 $\alpha$ -cyclo-9,10-seco-5 $\alpha$ -androstane



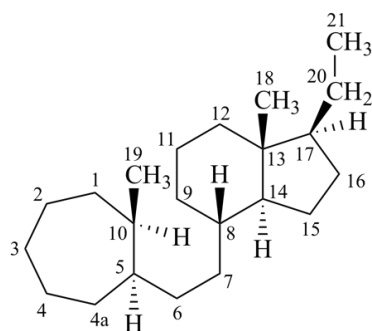
5 $\beta$ ,19-cyclo-4a-homo-5 $\beta$ -androstane



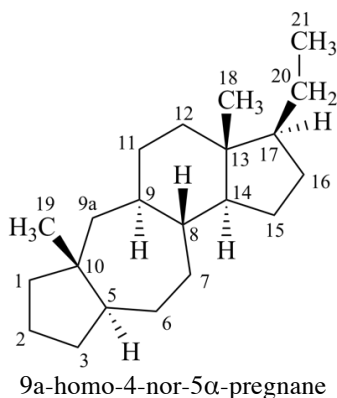
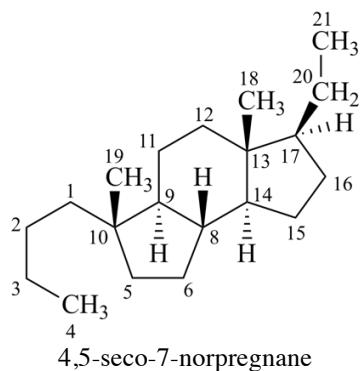
9 $\beta$ ,19-cyclo-4-nor-5 $\alpha$ ,9 $\beta$ -androstane



pregnane  
(fundamental parent structure)



9,10-seco-4a-homo-5 $\alpha$ -pregnane



## P-101.4 REPLACEMENT OF SKELETAL ATOMS

P-101.4.1 General methodology

P-101.4.2 Skeletal replacement of carbon atoms by heteroatoms

P-101.4.3 Skeletal replacement of heteroatoms by carbon atoms

P-101.4.4. Skeletal replacement of heteroatoms by other heteroatoms

P-101.4.5 Indicated hydrogen

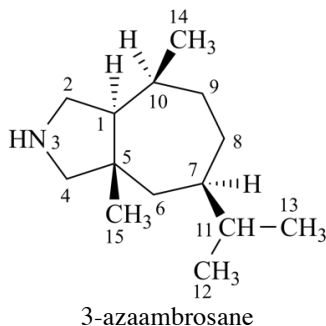
### P-101.4.1 General methodology

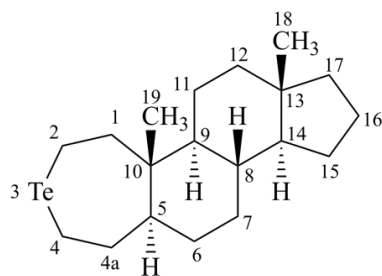
The principles of skeletal replacement ('a') nomenclature, as described in P-15.4 and P-51.4 to modify parent structures, are applied to replace carbon skeletal atoms by heteroatoms, such as O, S, N. Contrary to the recommended alphabetical order for citation in names in the Revised Section F (ref. 9), the seniority order of the 'a' prefixes prescribed in P-15.4 is recommended for skeletal replacement. In addition to the methodology used to generate systematic names, skeletal replacement ('a') nomenclature is also used to replace heteroatoms in parent structures by carbon atoms and by other heteroatoms.

### P-101.4.2 Skeletal replacement of carbon atoms by heteroatoms

Heteroatoms are denoted by 'a' prefixes that are cited before nondetachable prefixes expressing skeletal modifications in fundamental parent structures, each with a locant to indicate its position; the fixed numbering of the parent structure is maintained. Skeletal modifications, if any, must be completed before skeletal replacement ('a') nomenclature can be applied.

Examples:



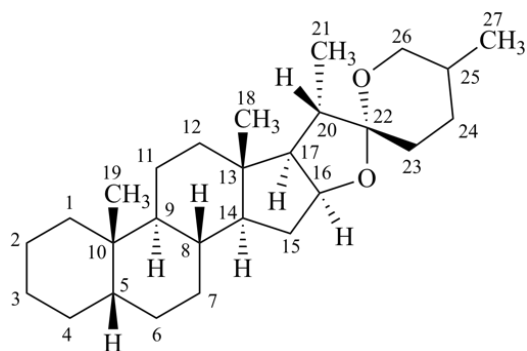


3-tellura-4a-homo-5 $\alpha$ -androstane

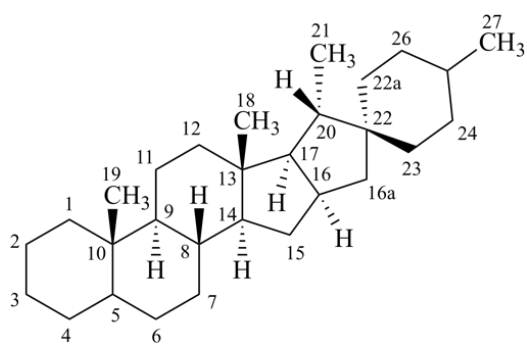
#### P-101.4.3 Skeletal replacement of heteroatoms by carbon atoms

Replacement of a heteroatom in a parent structure by a carbon atom is indicated by the replacement ('a') prefix 'carba'. The original numbering is maintained. If the heteroatom is not numbered, the replacing carbon atom is numbered by affixing the letter 'a' to the locant of the immediately adjacent lower numbered skeletal atom. If the immediately adjacent lower numbered skeletal atom is a 'homo' atom, the letter 'b', 'c', etc., as appropriate, is used. Configuration at the new carbon skeletal atom is described by methods for specifying additional configuration (see P-101.2.6.1).

Examples:



spirostan  
(fundamental parent structure)

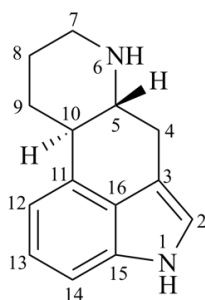


16a,22a-dicarba-5 $\beta$ -spirostan

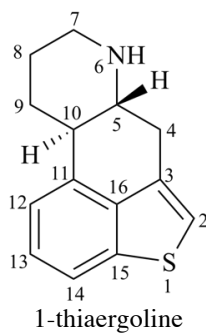
#### P-101.4.4 Skeletal replacement of heteroatoms by other heteroatoms

Replacement of a heteroatom in a stereoparent hydride by another heteroatom is denoted by the appropriate skeletal replacement ('a') prefix and locant.

Example:



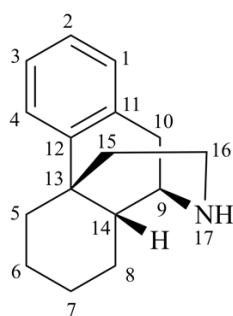
ergoline  
(fundamental parent structure)



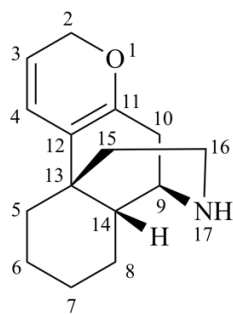
**P-101.4.5** Indicated hydrogen

When the replacement of a skeletal atom in a portion of a parent structure that is mancude (contains the maximum number of noncumulative double bonds) or an extended conjugated system of double bonds results in the creation of a saturated skeletal position, that position is indicated by the symbolism of indicated hydrogen (see P-14.7 and P-58.2).

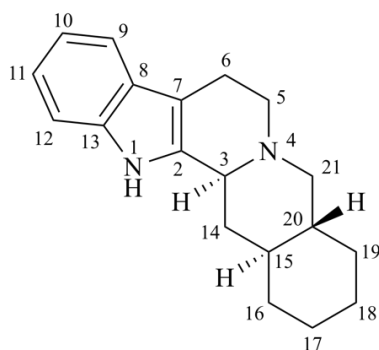
Examples:



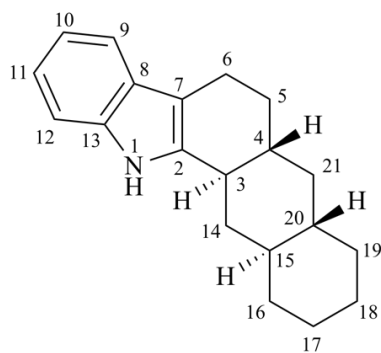
morphinan  
(fundamental parent structure)



2H-1-oxamorphinan



yohimban  
(fundamental parent structure)



(4 $\beta$ H)-4-carbayohimban

## P-101.5 ADDITION OF RINGS AND RING SYSTEMS

Three types of rings and ring systems can be incorporated into parent structures:

- P-101.5.1 Mancude rings and ring systems incorporated by fusion nomenclature
- P-101.5.2 Rings and ring systems incorporated by bridged fused ring nomenclature
- P-101.5.3 Rings and ring systems incorporated by spiro nomenclature

The methods, in certain cases adapted to parent structures, used for the construction of systematic names and described in Chapters P-1 through P-8 above, are applied.

### P-101.5.1 Mancude rings and ring systems incorporated by fusion nomenclature

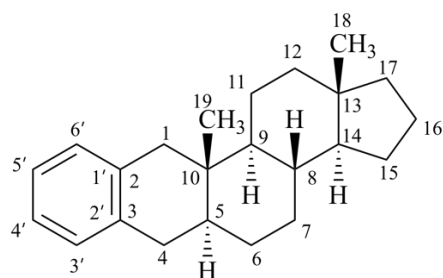
The fundamental parent structure as a component is used in fusion nomenclature in its normal state of hydrogenation. Accordingly, a double bond is not cited at the fusion site just because the other component contains the maximum of noncumulative double bonds. Furthermore, contrary to the rules prescribed in P-25, a fundamental parent structure is always chosen as the principal component and the attached component must be a mancude ring or ring system.

**P-101.5.1.1** A ring or ring system considered as a mancude parent hydride in accordance with the rules prescribed in Chapter P-2, carbocyclic or heterocyclic, fused to a parent structure is described by its fusion prefix name (see P-25) preceding the name of the fundamental parent structure. The skeletal atoms of the parent structure involved in the fusion are identified by plain (unprimed) locants and not by italicized letters 'a', 'b', etc.; the skeletal atoms of the mancude component, involved in the fusion, are identified by primed locant numbers. The position of the fusion is indicated by a fusion descriptor, including two sets of locants; the first cited set is that of the attached component, the second set relates to the principal component, the fundamental parent structure; the two sets are separated by a colon, enclosed in brackets, and cited between the two components. Where there is a choice, the locants for the mancude attached component are as low as possible and are cited in the same direction of numbering as for the parent structure.

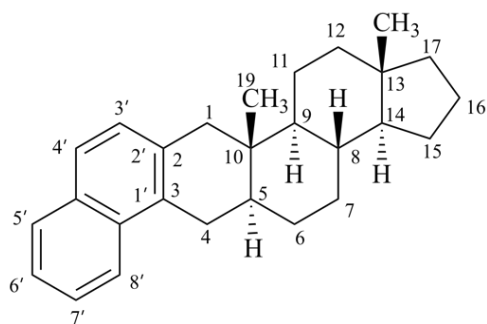
Terminal vowels, 'o' or 'a', in the name of the prefix are not elided when followed by a vowel, as prescribed for normal fusion nomenclature in P-25.3.1.3.

No elision of vowels is a change from previous recommendations.

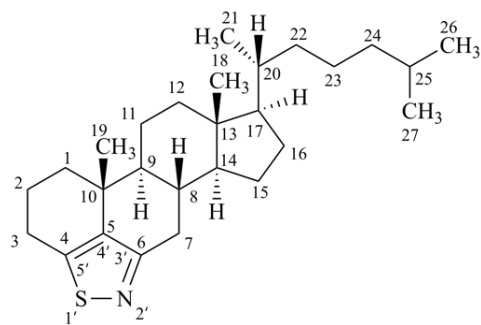
Examples:



benzo[2,3]-5 $\alpha$ -androstane  
(locants 1',2' are omitted)



naphtho[2',1':2,3]-5 $\alpha$ -androstane



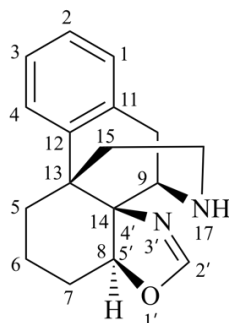
[1,2]thiazolo[5',4',3':4,5,6]cholestane

(not isothiazolo[5',4',3':4,5,6]cholestane;

the name isothiazole is no longer recommended as a fusion component, see P-25.3.2.1.2)

**P-101.5.1.2** The attached component fused to a parent structure is a mancude compound (it contains the maximum number of noncumulative double bonds). Saturated positions on such components, including the fusion sites, that have at least one hydrogen atom are specified by an indicated hydrogen. They are also specified by a descriptor composed of the locant, followed by the configuration descriptor ' $\alpha$ ' or ' $\beta$ ' and finally by the indicated hydrogen symbolism (see P-14.7), placed in parentheses at the front of the name as are stereodescriptors. Locants of the attached component are used to identify the position of the indicated hydrogen, but locants (unprimed) of the stereoparent hydride are used, if there is a choice between primed and unprimed locants.

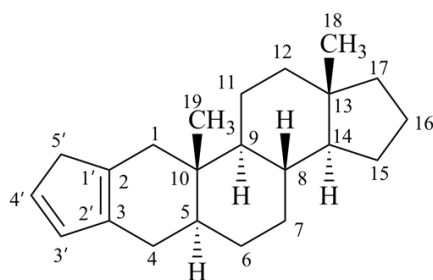
Examples:



(8 $\alpha$ H)-[1,3]oxazolo[5',4':8,14]morphinan

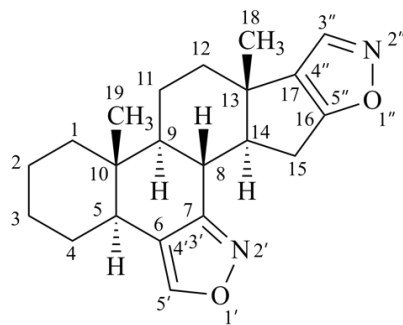
[not 8 $\alpha$ H-oxazolo[5',4':8,14]morphinan;

the name oxazole without heteroatom locants is no longer recommended as a fusion component;  
an 'indicated hydrogen atom' denoted by the stereodescriptor (8 $\alpha$ H) is needed to complete the name]

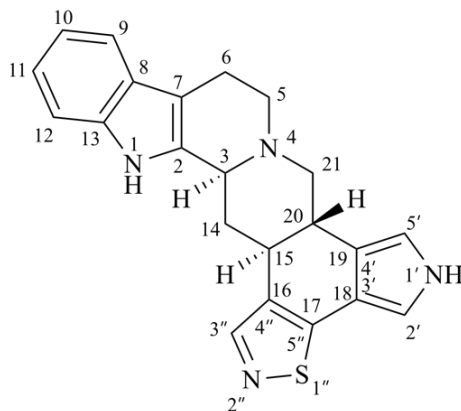


5'H-cyclopenta[2,3]-5 $\alpha$ -androstane

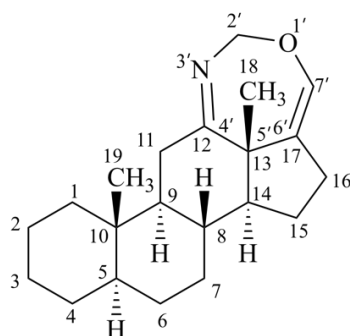




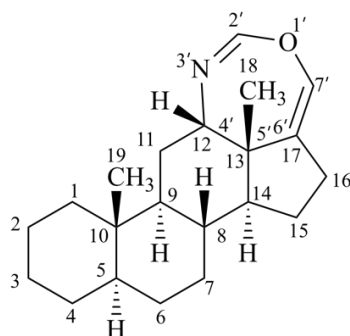
bis[1,2]oxazolo[4',3':6,7;5'',4'':16,17]-5 $\alpha$ -androstane



1'*H*-pyrrolo[3',4':18,19][1,2]thiazolo[4'',5'':16,17]yohimban



2'*H*-[1,3]oxazepino[4',5',6':12,13,17]-5 $\alpha$ -androstane

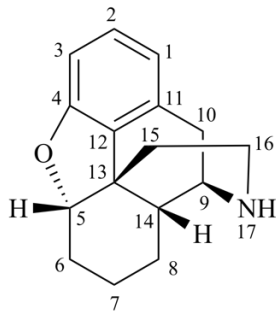


(12 $\beta$ H)-12*H*-[1,3]oxazepino[4',5',6':12,13,17]-5 $\alpha$ -androstane

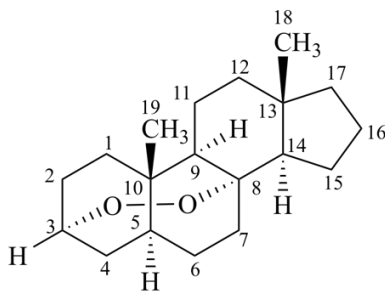
### P-101.5.2 Rings and ring systems incorporated by bridged fused ring nomenclature

Atomic bridges added to fundamental parent structures may be described by the methods used in fusion nomenclature for bridged fused ring systems. The names of the bridges are those prescribed in P-25.4. This method is often used with heteroatom bridges. In fact, this method is often more useful than fusion procedures described in P-101.5.1 for describing certain types of heterocyclic rings fused to a fundamental parent structure, for instance 'epoxy' to denote a bridge rather than 'oxireno' to denote the ring fused as an attached component. The use of atomic bridges is preferred to fusion nomenclature to connect two nonadjacent atoms in a fundamental parent structure [epoxides and thioepoxides are exceptions as they can be named substitutively (see P-63.5)]. The prefixes used to denote bridges are nondetachable; they are cited in a name in front of the prefixes used to denote skeletal modifications, preceded by appropriate locants.

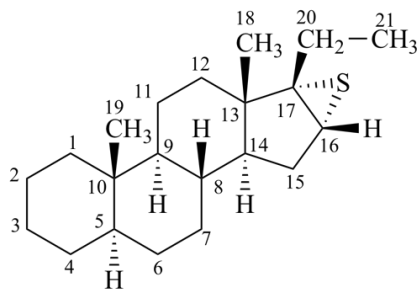
Examples:



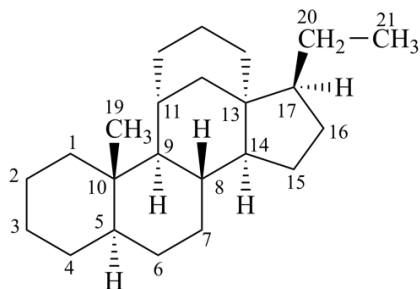
4,5 $\alpha$ -epoxymorphinan  
(5 $\beta$ H)-5,13-dihydrofuro[2',3',4',5':4,12,13,5]morphinan



3 $\alpha$ ,8-epidioxy-5 $\alpha$ ,8 $\alpha$ -androstane

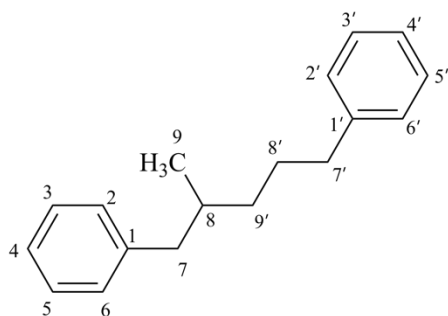


(16 $\beta$ H)-thiireno[2',3':16,17]-5 $\alpha$ -pregnane (fusion name)  
16 $\alpha$ ,17-epithio-5 $\alpha$ -pregnane

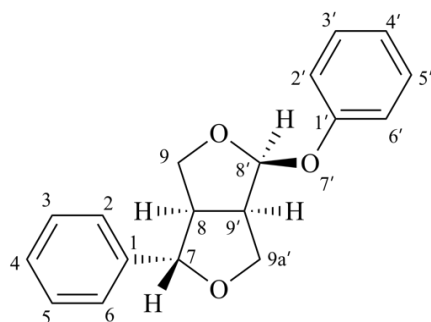


11 $\alpha$ ,18-ethano-5 $\alpha$ ,13 $\alpha$ -pregnane  
11 $\alpha$ ,13-propano-18-nor-5 $\alpha$ ,13 $\alpha$ -pregnane  
11 $\beta$ ,18-cyclo-12a,12b-dihomo-5 $\alpha$ -pregnane  
11 $\alpha$ ,18b-cyclo-18a,18b-dihomo-5 $\alpha$ ,13 $\alpha$ -pregnane

**Explanation:** The inversion of configuration at C-13 is not counted as an operation since it is not modifying the skeleton of the molecule.



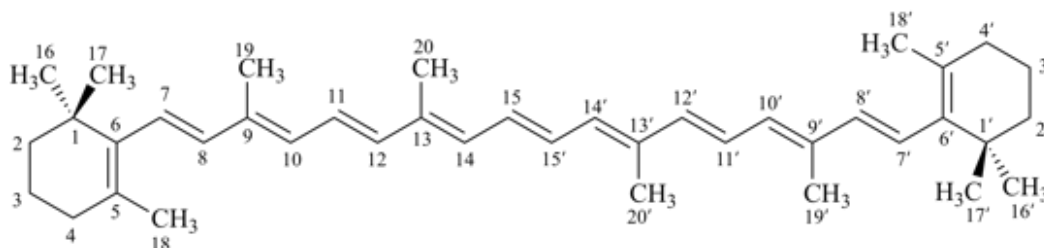
8,9'-neolignane  
(fundamental parent structure)  
1,1'-(2-methylpentane-1,5-diyl)dibenzene



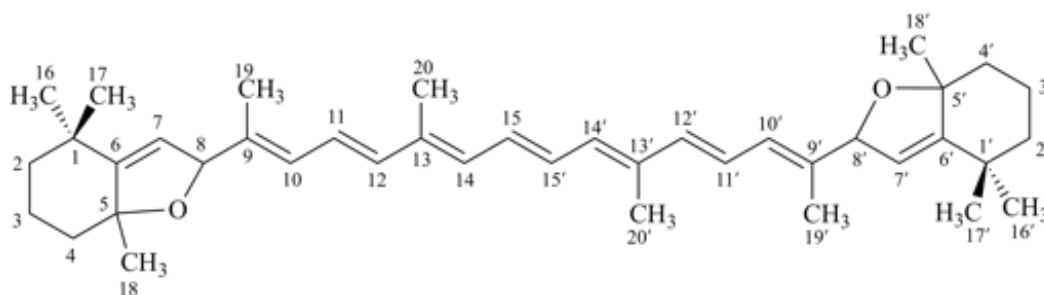
(7 $\alpha$ ,8 $\alpha$ ,8' $\beta$ ,9' $\alpha$ )-7,9a':8',9-diepoxy-7'-oxa-9a'-homo-8,9'-neolignane  
(1*S*,3*aR*,4*S*,6*aR*)-1-phenoxy-4-phenyltetrahydro-1*H*,3*H*-furo[3,4-*c*]furan

Contrary to the recommendations for systematic nomenclature of organic compounds, in carotenoid nomenclature (see ref. 40), the bridge named 'epoxy' is considered detachable and the hydro/dehydro prefixes nondetachable. The name of the following bridged  $\beta,\beta$ -carotene is written in conformity with the rules of carotenoid nomenclature (ref. 40, Carotenoid Rule 7.3) but in contradiction to Rule P-15.1.5.

Examples:



$\beta,\beta$ -carotene  
(fundamental parent structure)

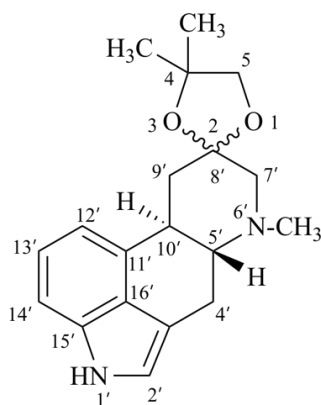


5,8:5',8'-diepoxy-5,8,5',8'-tetrahydro- $\beta,\beta$ -carotene  
(see P-14.3.1 for use of unprimed and primed locants)

#### P-101.5.3 Rings and ring systems incorporated by spiro nomenclature

Spiro compounds are named as prescribed in P-24.5 for monospiro compound having at least one polycyclic component.

Example:



(2 $\xi$ )-4,4,6'-trimethylspiro[1,3-dioxolane-2,8'-ergoline]

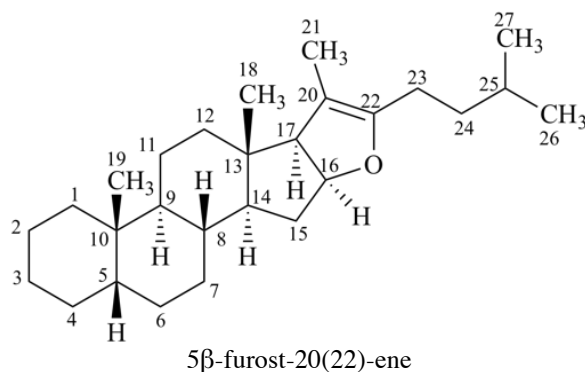
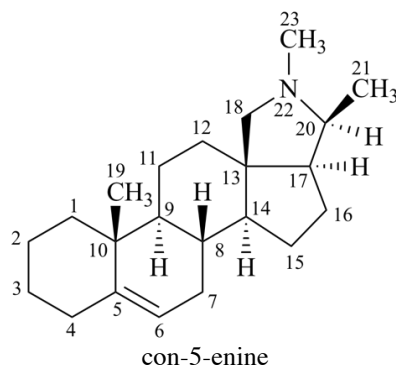
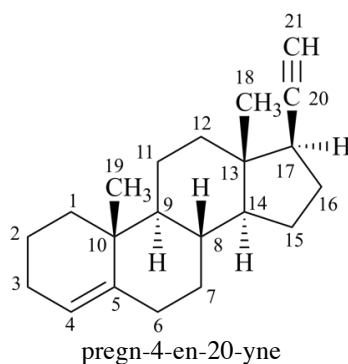
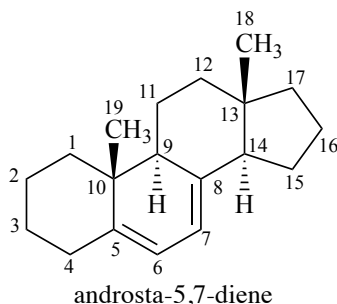
#### P-101.6 MODIFICATION OF THE DEGREE OF HYDROGENATION OF PARENT STRUCTURES

The general principles and rules for modifying the degree of hydrogenation of parent hydrides prescribed in Section P-31 are applied to parent structures. The endings 'ene' and 'yne' (see P-31.1) and the prefixes 'hydro' and 'dehydro' (see P-31.2) are used, depending on the subtractive or additive operation required. There is no limit to the introduction

of double bonds in parent hydrides; 'hydro/dehydro' prefixes can be used in any number, as required, provided that no mancude structure is generated.

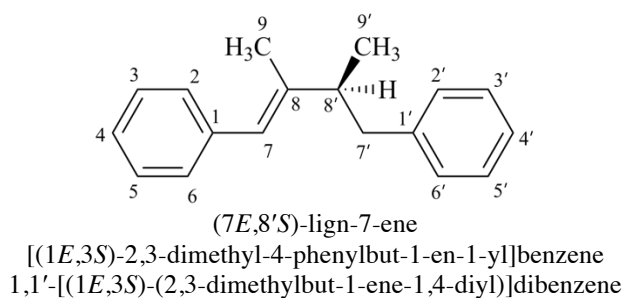
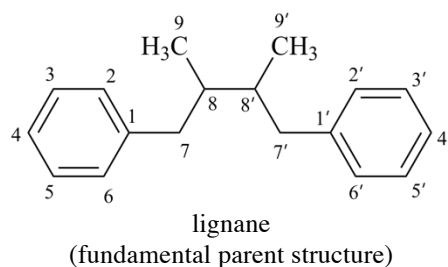
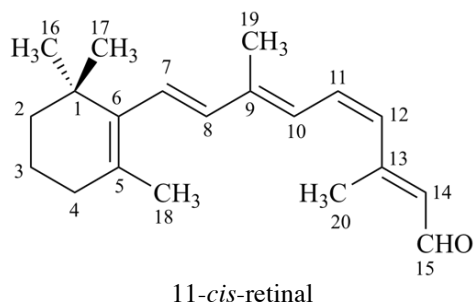
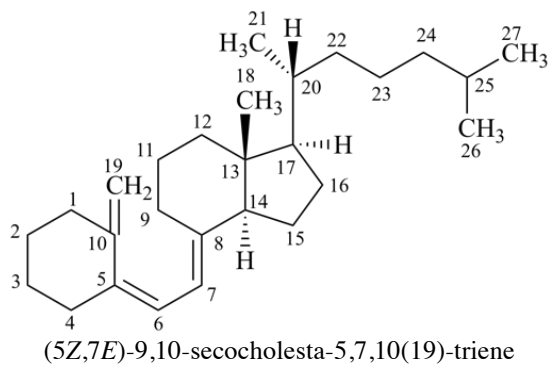
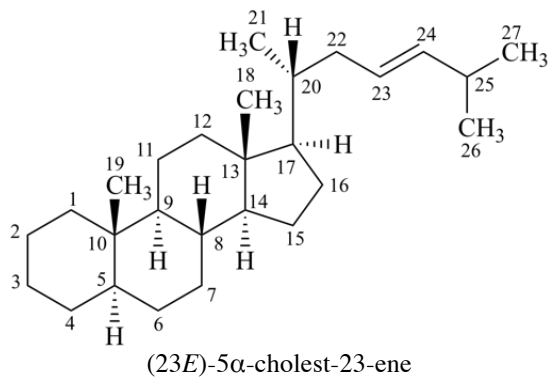
**P-101.6.1** Unsaturation in a compound whose parent structure is fully saturated or in the portion of a parent structure that is otherwise fully saturated and whose name ends in 'an', 'ane', or 'anine' is indicated by changing 'an' or 'ane' to 'ene' or 'yne' and 'anine' to 'enine' or 'ynine' and by adding numerical multiplying prefixes as prescribed in P-31.1.1.2. Locants are placed immediately before the part of the name to which they relate.

Examples:



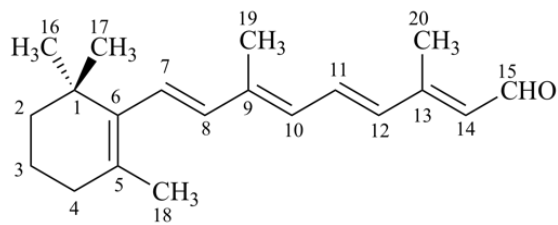
**P-101.6.2** The descriptors '*E*' and '*Z*', preceded by appropriate locants, are used to describe modified or additional stereochemical configurations for double bonds. The stereodescriptors '*cis*' and '*trans*' are used in carotenoid nomenclature (ref. 40) and retinoid nomenclature (ref. 49).

Examples:



**P-101.6.3** The prefix '*all*' is used in front of stereodescriptors to indicate that all configurations are identical. This prefix is used only in the nomenclature of natural products, for example, '*all-trans*' to denote the fact that in retinal all double bonds are '*trans*'.

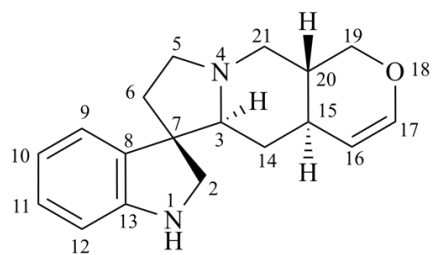
Example:



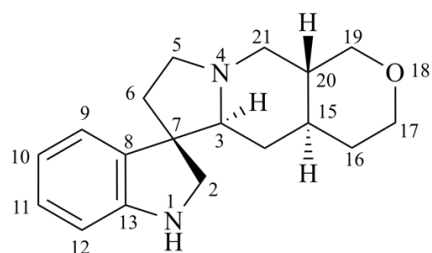
*all-trans-retinal*

**P-101.6.4** Saturation of double bonds in a parent structure whose name implies the presence of isolated double bonds and/or systems of conjugated double bonds is described by the prefix 'hydro', itself preceded by the locants of the saturated positions. The 'hydro' prefix is detachable and always cited immediately in front of the fundamental parent structure (see P-31.2).

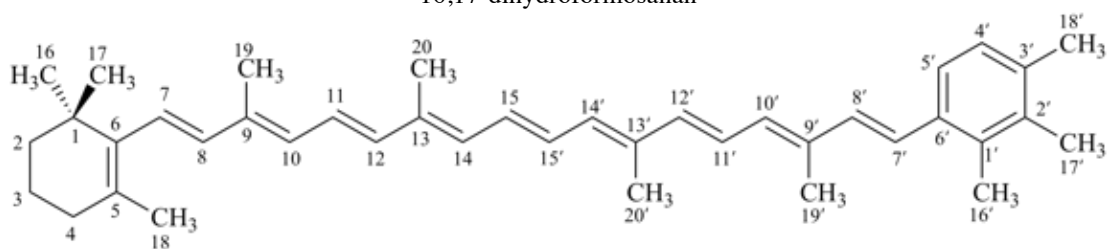
Examples:



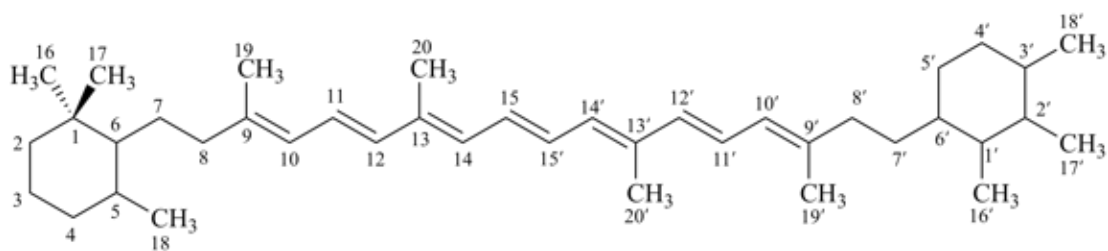
formosanan  
(fundamental parent structure)



16,17-dihydroformosanan



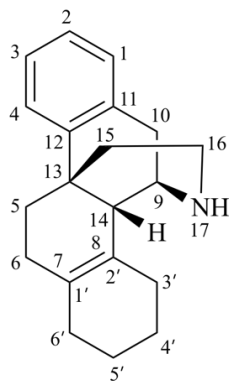
$\beta,\gamma$ -carotene



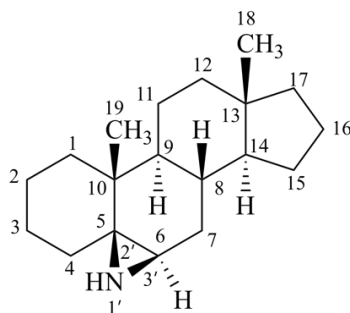
5,6,7,8,1',2',3',4',5',6',7',8'-dodecahydro- $\beta,\gamma$ -carotene  
(see P-14.3.1 for use of unprimed and primed locants)

**P-101.6.5** Saturated, or partially saturated, carbocyclic and heterocyclic ring components fused to a parent structure are named using 'hydro' prefixes. When there is a choice between primed and unprimed locants, the unprimed locants are used.

Examples:



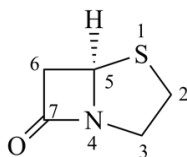
3',4',5',6'-tetrahydrobenzo[7,8]morphinan



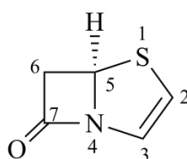
(6 $\alpha$ H)-1',6-dihydroazirino[2',3':5,6]-5 $\beta$ -androstane  
[for the symbol (6 $\alpha$ H), see P-101.5.1.2]

**P-101.6.6** The introduction of unsaturation additional to any unsaturation implied in a parent structure whose name does not end in 'an', 'ane', or 'anine', the conversion of an implied double bond into a triple bond, and the introduction of an additional double bond with rearrangement of an implied double bond are denoted by the prefix 'dehydro', itself prefixed by a numerical multiplying term equal to the number of hydrogen atoms removed and the appropriate locants. The 'dehydro' prefix is detachable and always cited at the front of the fundamental parent structure, after any detachable alphabetized prefixes, when present.

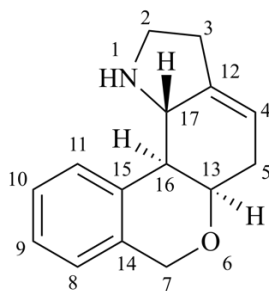
Examples:



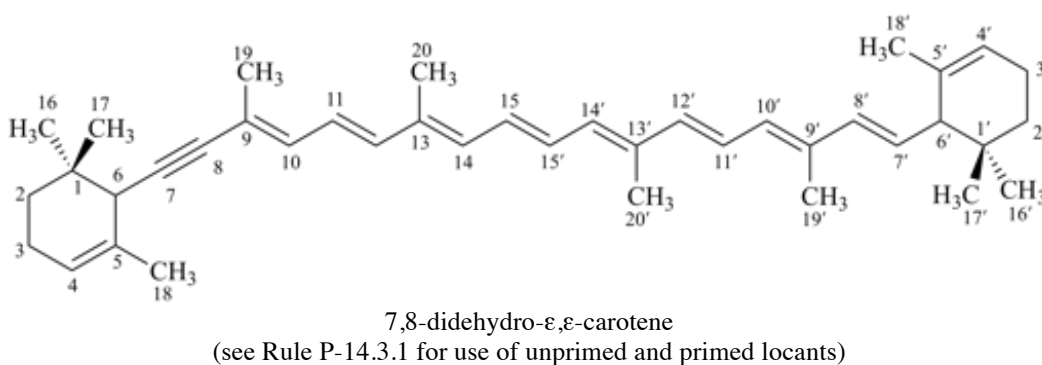
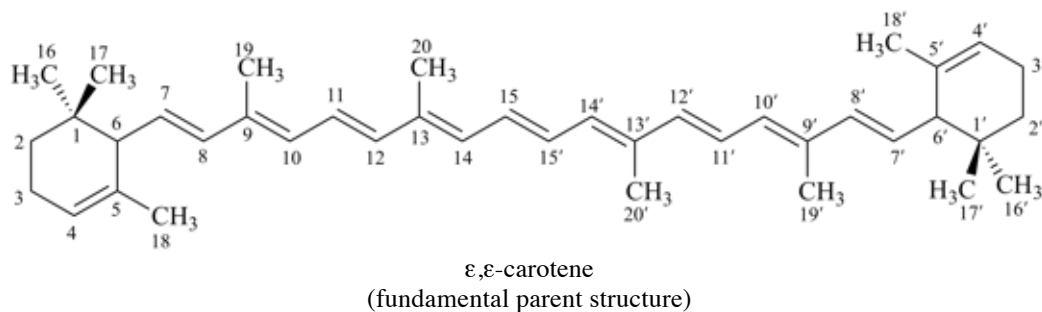
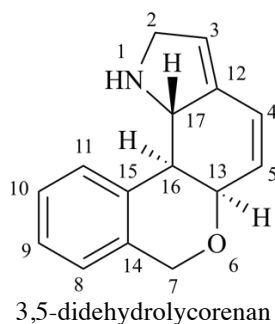
penam  
(fundamental parent structure)  
(note new numbering)



2,3-didehydropenam

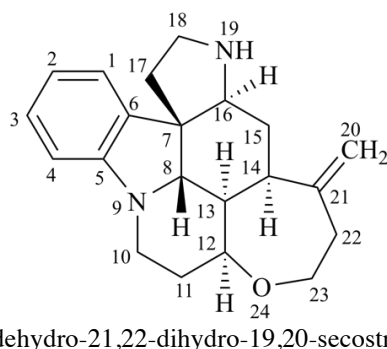
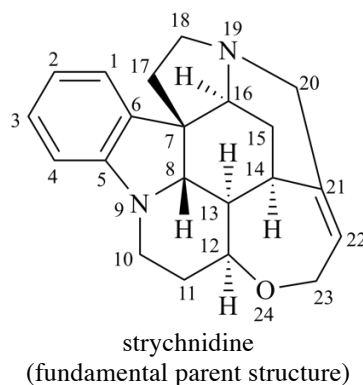


lycorenan  
(fundamental parent structure)



**P-101.6.7** Rearrangement of double bonds may be indicated by a combination of 'hydro' and 'dehydro' prefixes. The 'dehydro' prefix is cited before the 'hydro' prefix, in accordance with the alphanumerical order.

Example:



### P-101.7 DERIVATIVES OF PARENT STRUCTURES

Derivatives of parent structures are named according to principles, rules, and conventions described in Chapters P-1 through P-9.



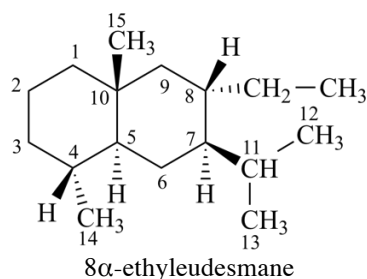
**P-101.7.1** The suffixes and prefixes of the nomenclature of organic compounds are used in the prescribed manner to name atoms and groups that are considered to substitute for hydrogen atoms of parent structures. The stereodescriptors  $\alpha$ ,  $\beta$ , and  $\xi$  are used to describe the configuration; they are cited in front of the prefix or suffix, preceded by the appropriate locant. Substitutive names so constructed are preferred to those that are formed by functional class nomenclature, except for some cyclic functional classes.

Substitution on rings and substitution on terminal segments are considered separately.

**P-101.7.1.1** Substitution by alkyl groups

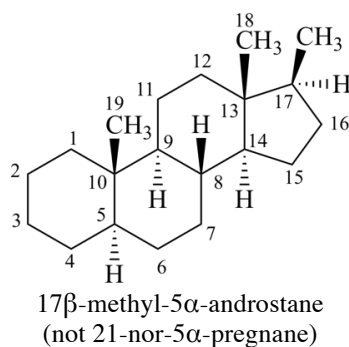
**P-101.7.1.1.1** Organyl groups such as aryl groups and alkyl groups are introduced by substitutive nomenclature.

Example:



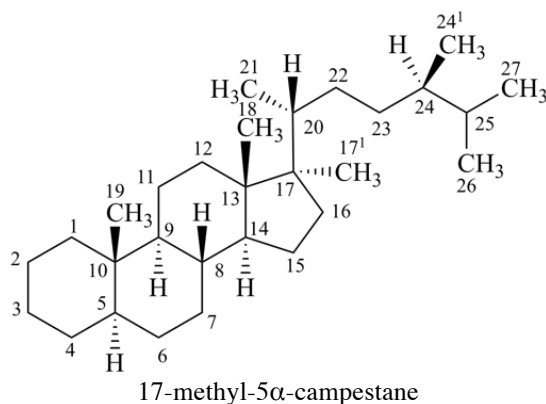
**P-101.7.1.1.2** The substitutive procedure is used to introduce a methyl group in androstane at position 17 $\beta$ ; the alternative method of subtracting a methylene group from pregnane by using the nondetachable prefix 'nor' is not recommended (see P-101.3.7.1).

Example:

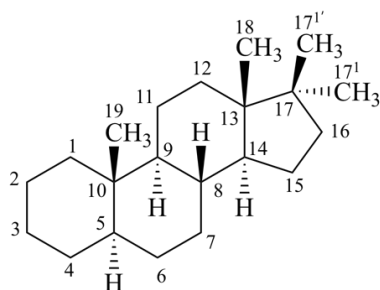


**P-101.7.1.1.3** Rule 3S-2.7 in ref. 16 describes the methodology to name steroids with a side chain as part of the parent carbocycle and an alkyl substituent at C-17. Rule 3S-2.7 also describes the methodology to name steroids with two alkyl substituents at C-17. This methodology is applicable to any fundamental parent structure described in Section P-101. Locants with superscript numbers are intended for the identification of the atoms, e.g. in  $^{13}\text{C}$ -nmr assignments, not as locants for further substitution.

Examples:



(the additional methyl group in position 17 is numbered 17<sup>1</sup>; other atoms are numbered as usual)

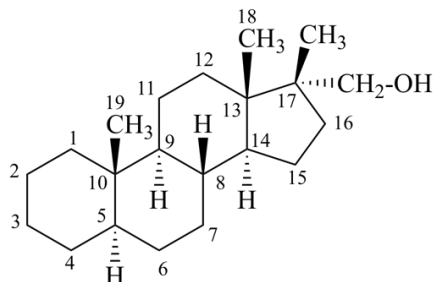


17,17-dimethyl-5 $\alpha$ -androstane

(both additional methyl groups are numbered 17<sup>1</sup>; the  $\beta$ -methyl group is primed)

**P-101.7.1.1.4** The principles, rules, and conventions of substitutive nomenclature are used when a characteristic group cited as a suffix is present on an alkyl substituent group added to a fundamental parent structure.

Example:

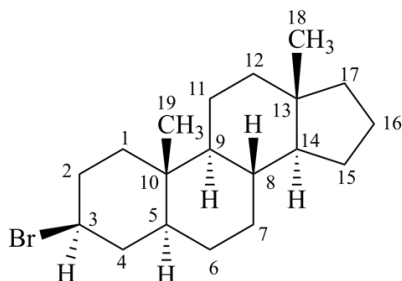


(17 $\beta$ -methyl-5 $\alpha$ -androstan-17 $\alpha$ -yl)methanol  
[not (21-nor-5 $\alpha$ -pregnan-17 $\alpha$ -yl)methanol]

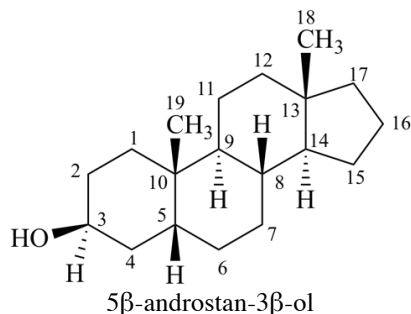
**P-101.7.1.2** Substitution on rings

Suffixes are used in accordance with the seniority order of suffixes, considering the cyclic nature of the parent hydride. Detachable prefixes are cited in alphanumerical order. The endings 'ene' and 'yne' are cited in the normal way; the 'hydro-dehydro' prefixes are detachable but cited last among detachable prefixes.

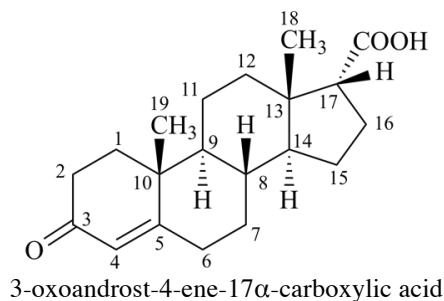
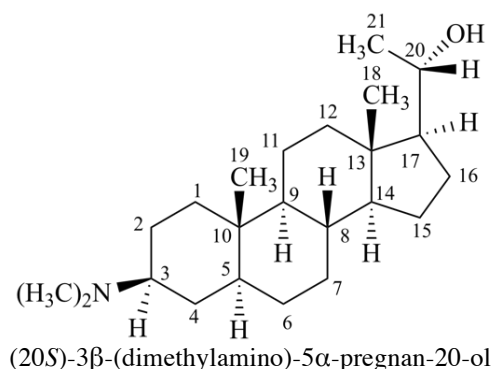
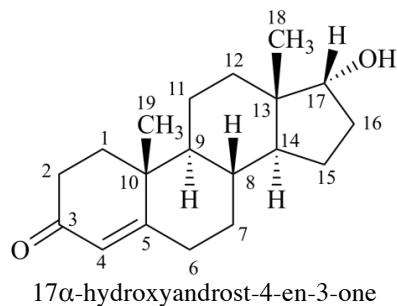
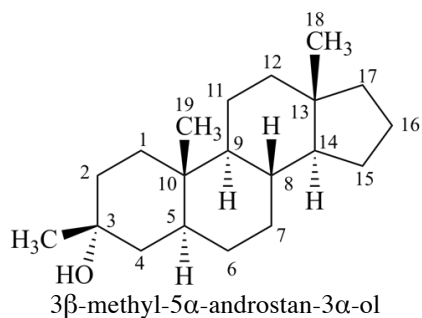
Examples:



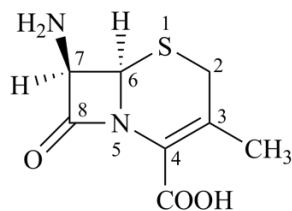
3 $\beta$ -bromo-5 $\alpha$ -androstane  
5 $\alpha$ -androstan-3 $\beta$ -yl bromide



5 $\beta$ -androstan-3 $\beta$ -ol



[not 21-nor-5α-pregnan-20-oic acid;  
the correct name involves the fewest number of operations (see P-101.3.7.1)]

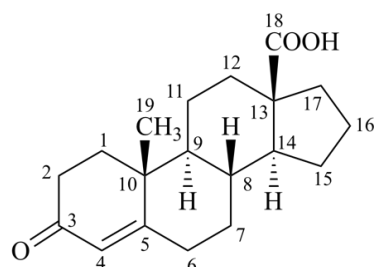


(6*R*,7*R*)-7-amino-3-methyl-8-oxo-5-thia-1-azabicyclo[4.2.0]oct-2-en-2-carboxylic acid  
7β-amino-3-methyl-3,4-didehydrocepham-4-carboxylic acid  
(note the new numbering for cepham; it is different than that reported in ref. 9)

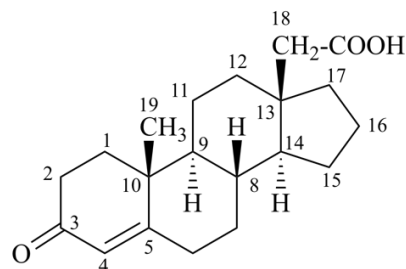
### P-101.7.1.3 Substitution on terminal segments

Substitution on terminal segments by prefixes and suffixes expressing characteristic groups is recommended, even when a carbon atom included in a characteristic group is present. Lengthening a terminal segment by two methylene groups is allowed and is denoted by the use of the prefix 'dihomo'. Further lengthening is possible, but alkyl groups must be used, as an exception to the rule related to seniority of the longest chain.

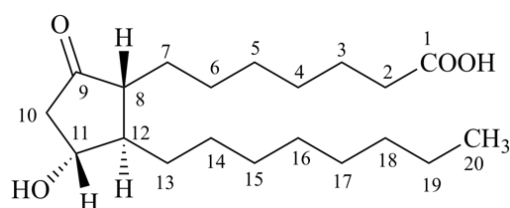
Examples:



3-oxoandrost-4-en-18-oic acid



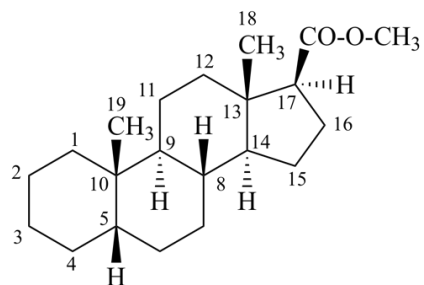
3-oxoandrost-4-ene-18-carboxylic acid



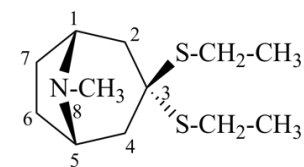
11 $\alpha$ -hydroxy-9-oxoprostan-1-oic acid

**P-101.7.2** Modifications to principal characteristic groups such as esters (see P-65.6.3.2), acetals (see P-66.6.5), etc. are named by the usual methods described in Chapter P-6. Cyclic modifications, such as lactones, cyclic acetals, etc. are named preferably as such rather than as fused or spiro ring systems, even if these names are functional class names (see also P-101.7.4).

Examples:



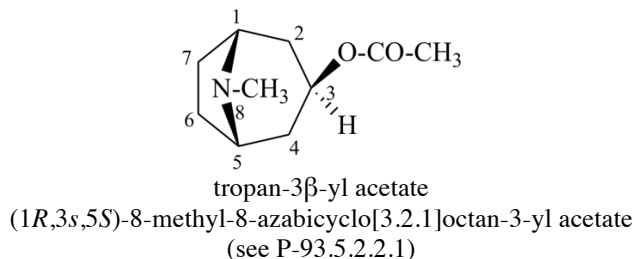
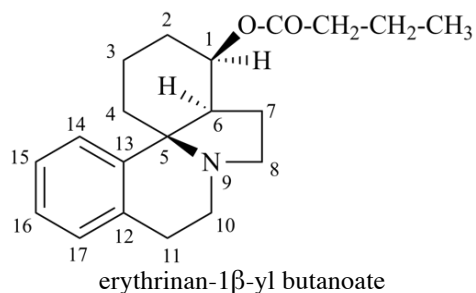
methyl 5 $\beta$ -androstane-17 $\beta$ -carboxylate



(1*R*,5*S*)-3,3-bis(ethylsulfanyl)-8-methyl-8-azabicyclo[3.2.1]octane  
3,3-bis(ethylsulfanyl)tropane  
tropan-3-one diethyl dithioether

**P-101.7.3** Names of substituent groups derived from parent structures are formed, by the general method described in P-29, by adding the suffixes 'yl', 'ylidene', or 'ylidyne', as appropriate, to the name of a parent, with elision of the final letter 'e', if present, before the letter 'y'.

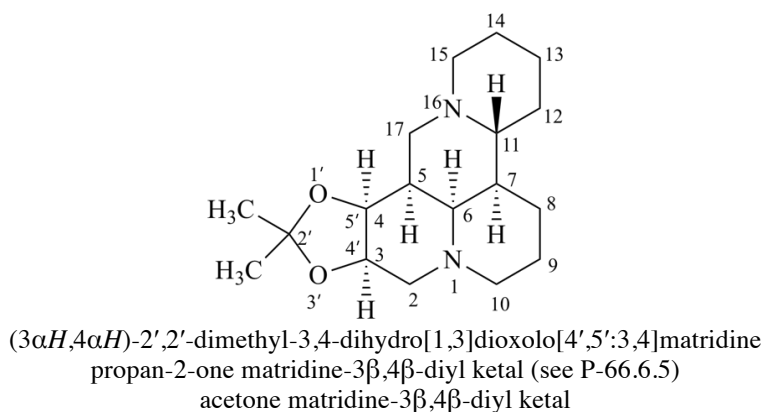
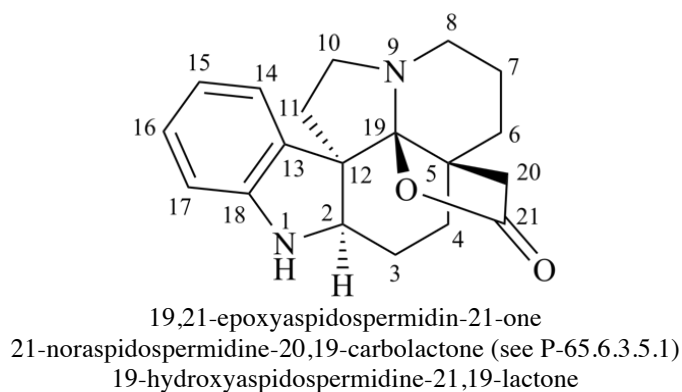
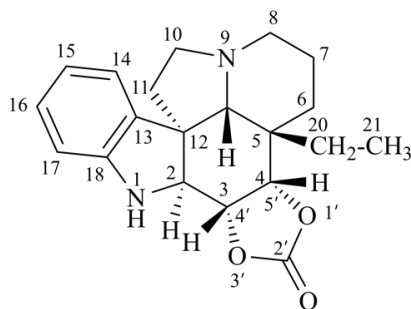
Examples:



#### P-101.7.4 Addition of rings denoting functional groups

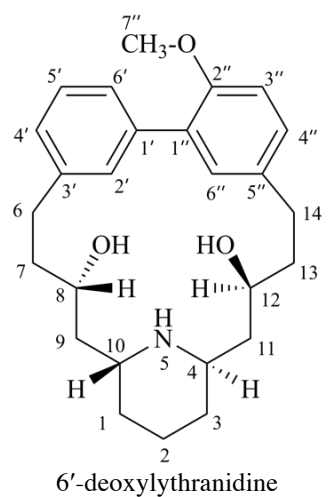
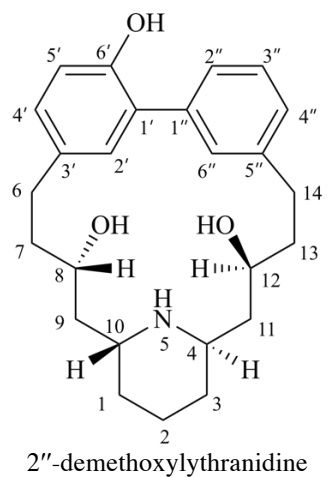
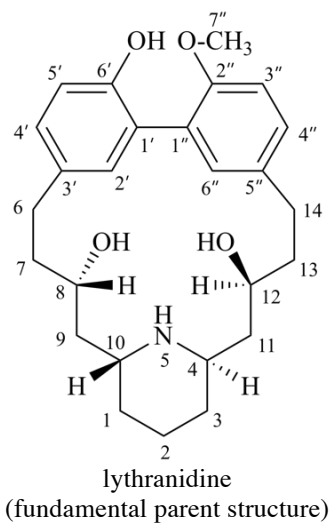
Rings denoting functional groups are preferably named by the usual methods described for constructing systematic names. Cyclic esters and lactones are named by the general method described for naming esters (see P-65.6.3.5). Names of acetals are formed by using the principles of functional class nomenclature (see P-66.6.5) rather than by fusion nomenclature described in P-101.5. When a choice is possible, a fusion name is preferred.

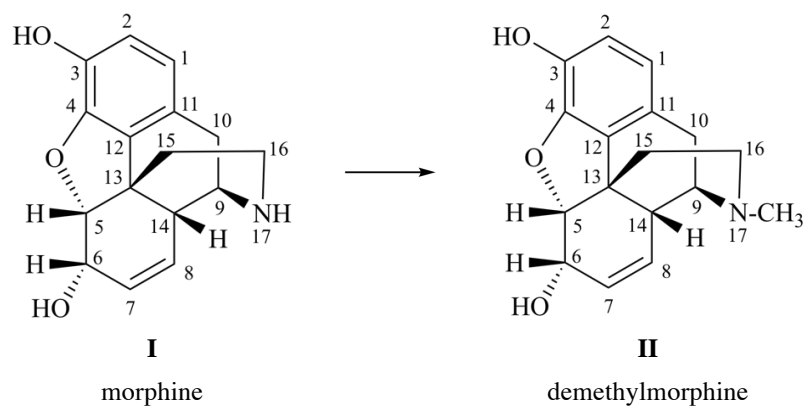
Examples:



**P-101.7.5** The prefix 'de' (not 'des'), followed by the name of a group or atom (other than hydrogen), denotes removal of that group or atom and addition of hydrogen atoms as necessary. The prefix 'de' is currently used in carbohydrate nomenclature (see P-102.5.3) to indicate the removal of an oxygen atom from –OH with reconnection of the hydrogen atom.

Examples:





**I** (5 $\beta$ H)-17-methyl-7,8-didehydrofuro[2',3',4',5':4,12,13,5]morphinan-3,6 $\alpha$ -diol  
**II** (5 $\beta$ H)-7,8-didehydrofuro[2',3',4',5':4,12,13,5]morphinan-3,6 $\alpha$ -diol

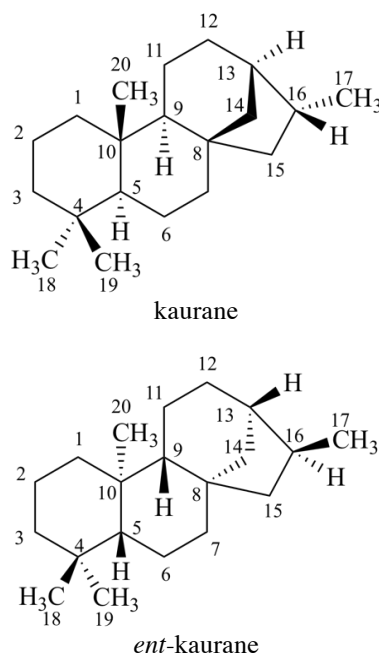
### P-101.8 FURTHER ASPECTS OF CONFIGURATIONAL SPECIFICATION

In addition to the specification of the absolute configuration of fundamental and modified parent structures using ' $\alpha$ ', ' $\beta$ ', ' $\xi$ ', ' $R$ ', and ' $S$ ' stereodescriptors, many other stereochemical features have to be described. The principles, rules and conventions described in Chapter P-9 are applied.

#### P-101.8.1 Inversion of configuration

Configurational inversion of all chirality centers is indicated by the italicized prefix '*ent*' (a contracted form for '*enantio*') placed at the front of the complete name of the compound. This prefix denotes inversion at all chirality centers (including those due to named substituents) whether these are cited separately or are implied in the name. For kaurane and *ent*-kaurane illustrated below, correct structures and names are given; the designations are reversed by Chemical Abstracts (see ref. 22).

Example:



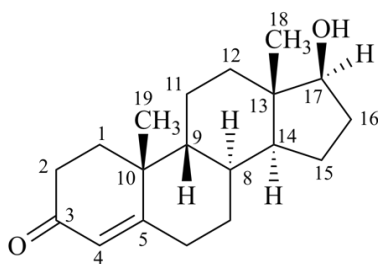
#### P-101.8.2 Racemates

Racemates are named by citing the italicized stereodescriptor '*rac*' (an abbreviation for *racemo*) in front of the whole name of the compound including the prefix '*epi*', if present. In the case of a racemic compound, the enantiomeric structure drawn should be the one that shows the lowest numbered chirality center in the  $\alpha$ -configuration. This may differ from the usual practice, which is to draw the enantiomeric structure having the same absolute configuration as the naturally occurring substance.

#### P-101.8.3 Relative configuration

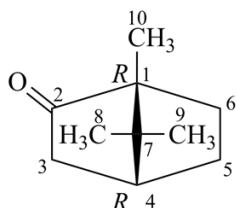
When the relative, but not the absolute configurational relationships among chirality centers are known, the symbol '*rel*' in association with  $R$  or  $S$  (preferred to ' $R^*$ ' and/or ' $S^*$ ') is used in accordance with Rule P-93.5.1.2. Alternatively, enantiomers of known relative, but unknown absolute, configuration may be distinguished by the compound stereodescriptor (+)-*rel*- or (-)-*rel*-, where the plus and minus sign refer to the direction of rotation of polarized

light at the sodium D line. Hence, the dextrorotatory form of the following structure would be named (+)-*rel*-17 $\beta$ -hydroxy-8 $\alpha$ ,9 $\beta$ -androst-4-en-3-one.



**P-101.8.4** The stereodescriptors '*R*' and '*S*' are used to describe the absolute configuration of stereogenic centers for a compound whose parent structure is achiral, for example bornane. They are also used, in place of ' $\alpha$ ', ' $\beta$ ', ' $\xi$ ', when a ring is opened creating two chiral portions, one of which may rotate, as shown for vitamin D.

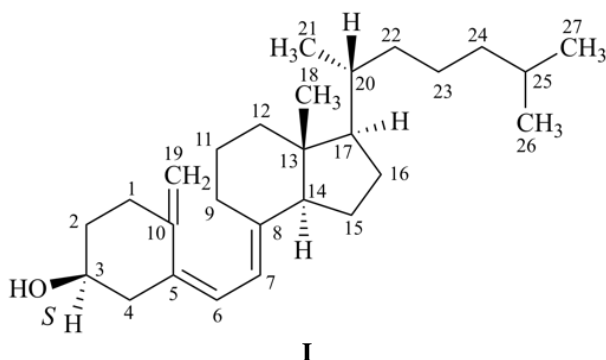
Example:



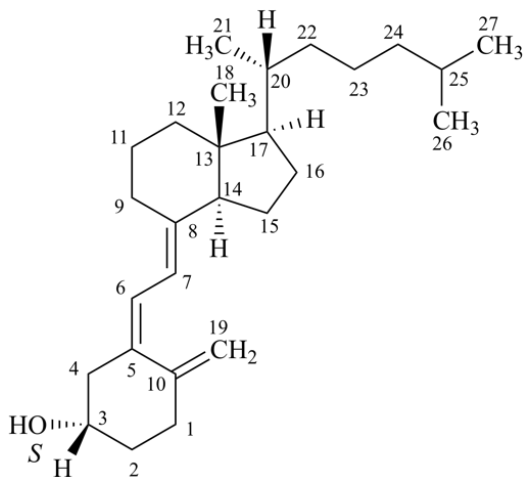
(1*R*,4*R*)-bornan-2-one

(+)-camphor

(1*R*,4*R*)-1,7,7-trimethylbicyclo[2.2.1]heptan-2-one



**I**



**II**

**I is equivalent to II**

(3*S*,5*Z*,7*E*)-9,10-secocholesta-5,7,10(19)-trien-3-ol

[structures **I** and **II** are two conformations of the same 3-hydroxy derivative]

## P-102 CARBOHYDRATE NOMENCLATURE

### P-102.0 INTRODUCTION

Nomenclature of carbohydrates is based on the concept of parent monosaccharides having retained names. These structures and names can be modified to indicate the nature of characteristic groups that are present, such as aldehydes, carboxylic acids, alcohols. They can also be combined to form di-, tri-, and oligosaccharides.



The nomenclature has been recently revised (ref. 27). This Section describes the basic concepts of this specific type of nomenclature, in particular the extensive system of symbols and stereodescriptors to indicate the configuration of the many diastereoisomers and enantiomers.

P-102.1 Definitions

P-102.2 Parent monosaccharides

P-102.3 Configurational symbolism

P-102.4 Choice of a parent structure

P-102.5 Monosaccharides: aldoses and ketoses; deoxy and amino sugars

P-102.6 Monosaccharides and derivatives as substituent groups

P-102.7 Disaccharides and oligosaccharides

## **P-102.1 DEFINITIONS**

P-102.1.1 Carbohydrates

P-102.1.2 Monosaccharides

P-102.1.3 Oligosaccharides

P-102.1.4 Polysaccharides

### **P-102.1.1 Carbohydrates**

The generic term 'carbohydrates' includes monosaccharides, oligosaccharides, and polysaccharides as well as substances derived from monosaccharides by reduction of the carbonyl group (alditols), by oxidation of one or more terminal groups to carboxylic acids, or by replacement of one or more hydroxy group(s) by a hydrogen atom, an amino group, a thiol group, or similar heteroatomic groups. It also includes derivatives of these compounds. The term 'sugar' is frequently applied to monosaccharides and lower oligosaccharides.

Cyclitols are generally not regarded as carbohydrates. For nomenclature of cyclitols, see P-104 and ref. 39.

### **P-102.1.2 Monosaccharides**

Parent monosaccharides are polyhydroxy aldehydes  $H-[CHOH]_n-CHO$  or polyhydroxy ketones  $H-[CHOH]_m-CO-[CHOH]_n-H$  with three or more carbon atoms.

The generic term 'monosaccharide' (as opposed to oligosaccharide or polysaccharide) denotes a single unit without glycosidic connections to other such units and includes aldoses, dialdoses, aldoketoses, ketoses, diketoses, as well as deoxy sugars and amino sugars, and their derivatives, provided that the parent compound has a (potential) carbonyl group.

Names for monosaccharides are either trivial or systematic. Many trivial names such as glucose, fructose, etc. are retained and used to describe the corresponding functional parents. This aspect of carbohydrate nomenclature is limited because it applies only to monosaccharides having four to six carbon atoms. A 'systematic carbohydrate nomenclature' has been developed that is applicable to compounds with four or more carbon atoms, and is used extensively by carbohydrate chemists for compounds with more than six carbon atoms, and for unsaturated or branched sugars. In these recommendations, these names are called 'systematic carbohydrate names' to differentiate them from names formed systematically by applying principles, rules and conventions of substitutive nomenclature discussed in Chapters P-1 to P-9 of these Recommendations that are called 'substitutive names' or 'systematic substitutive names' (see P-102.5.2.3 for a discussion on and the illustration of these two types of nomenclature).

#### **P-102.1.2.1 Aldoses and ketoses**

Monosaccharides with an aldehydic carbonyl or potential aldehydic carbonyl group are called aldoses; those with a ketonic carbonyl or potential carbonyl group, ketoses.

Addition of a numerical prefix (e.g. 'pent', 'hex') indicates the number of carbon atoms present (e.g. 'aldopentose', 'ketoheptose').

The term 'potential aldehydic group' refers to the hemiacetal group arising from ring closure; the term 'potential ketonic group' refers to the hemiketal structure.

Cyclic hemiacetals or hemiketals of sugars with a five-membered ring (oxolane or tetrahydrofuran) ring are called 'furanoses', those with a six-membered ring (oxane or tetrahydropyran) ring 'pyranoses'.

Dialdoses are monosaccharides containing two (potential) aldehydic groups.

Diketoses are monosaccharides containing two (potential) ketonic groups.

Ketoaldoses are monosaccharides containing one (potential) aldehydic group and one (potential) ketonic group; this term is preferred to 'aldoketoses' and 'aldosuloses'.

### P-102.1.2.2 Deoxysugars

Monosaccharides in which an alcoholic hydroxy group has been replaced by a hydrogen atom are called 'deoxy sugars'.

### P-102.1.2.3 Amino sugars

Monosaccharides in which an alcoholic hydroxy group has been replaced by an amino group are called 'amino sugars'. When the hemiacetal group is replaced by an amino group, the compounds are called 'glycosylamines'.

### P-102.1.2.4 Glycosides

Glycosides are mixed acetals formally arising by elimination of water between the hemiacetal or hemiketal hydroxy group of a sugar and a hydroxy group of a second compound. The bond between the two components is called a 'glycosidic bond'.

### P-102.1.3 Oligosaccharides

Oligosaccharides are compounds in which monosaccharide units are joined by glycosidic linkages. According to the number of units, they are called disaccharides, trisaccharides, etc. The maximum number of units is not defined.

### P-102.1.4 Polysaccharides

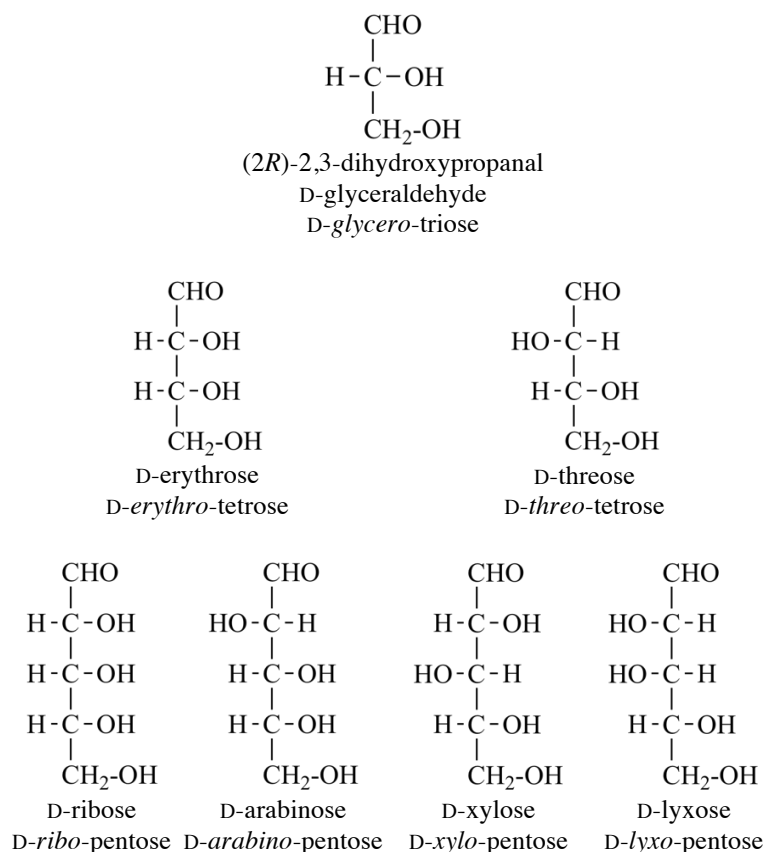
'Polysaccharide' (glycan) is the name given to a macromolecule consisting of a large number of monosaccharide (glycose) residues joined to each other by glycosidic linkages. The term 'poly(glycose)' is not a synonym for polysaccharide (glycan), because it includes monosaccharide residues joined to each other by nonglycosidic linkages.

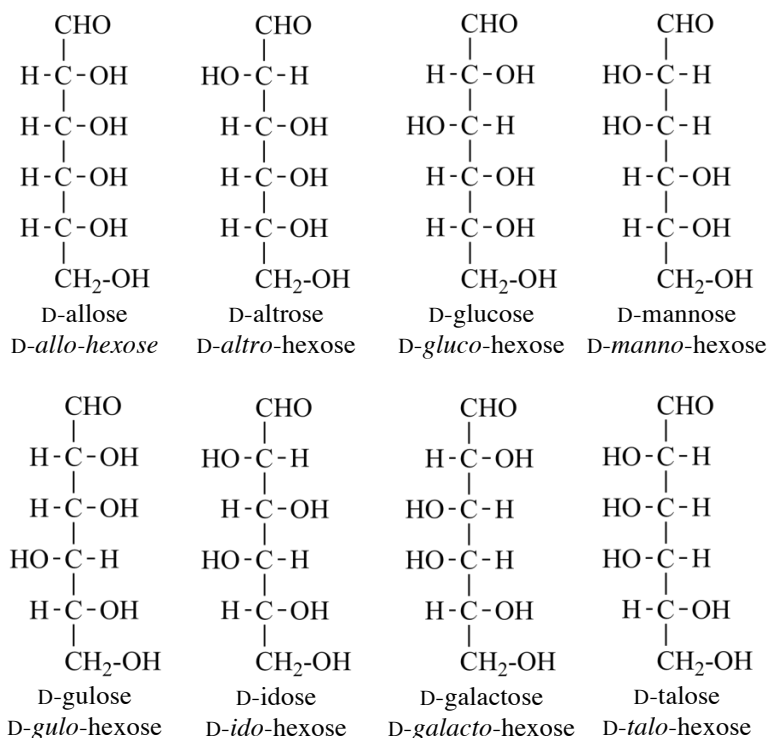
## P-102.2 PARENT MONOSACCHARIDES

**P-102.2.1** The bases for carbohydrate names are the structures of the parent monosaccharides in their acyclic form. Tables 10.2 and 10.3 give retained names for parent aldoses and ketoses with up to six carbon atoms. These retained names are customarily used when the acyclic aldose or ketose has a carbon chain consisting of four, five, or six carbon atoms. Names of monosaccharides whose carbon skeleton is composed of more than six carbon atoms are systematic carbohydrate names.

In Table 10.2 structures and retained names of the aldoses (in the aldehydic, acyclic form) with three through six carbon atoms are described. Only the D-forms are shown; the L-forms are the mirror images.

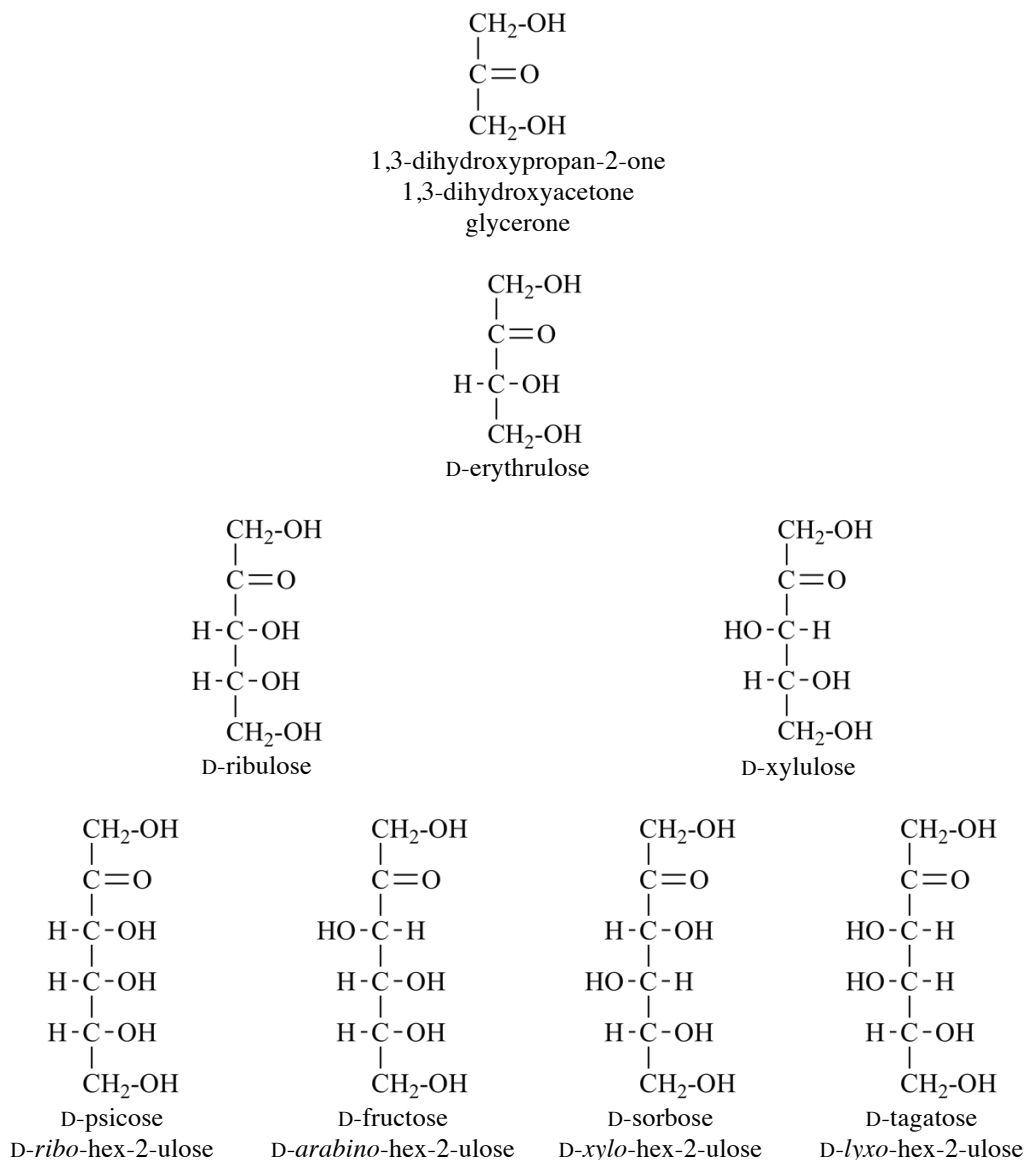
**Table 10.2 Retained and systematic carbohydrate names and structures (in the aldehydic acyclic form) of the aldoses with three through six carbon atoms**





In Table 10.3 structures and retained names of the 2-ketoses (in the ketonic, acyclic form) with three through six carbon atoms are described. Only the D-forms are shown; the L-forms are the mirror images.

**Table 10.3 Structures, and carbohydrate names, of the 2-ketoses with three through six carbon atoms**

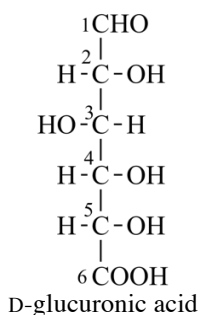
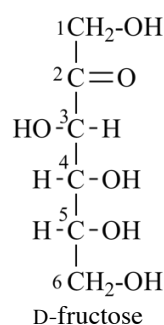
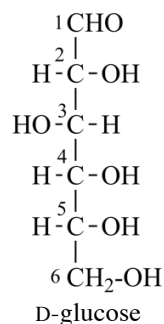


### P-102.2.2 Numbering parent structures

The carbon atoms of a monosaccharide are numbered consecutively in such a way that:

- (1) a (potential) aldehyde group receives the locant 1 (even if a more senior characteristic group is present);
- (2) the most senior of other characteristic groups expressed in the suffix receives the lowest possible locant, i.e. carboxylic acid (derivatives) > (potential) ketonic carbonyl groups.

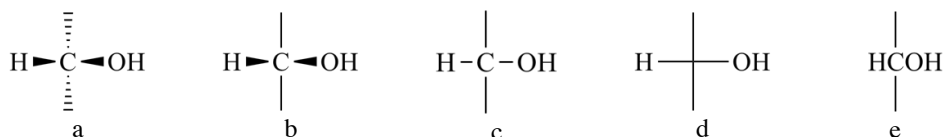
Examples:



### P-102.3 CONFIGURATIONAL SYMBOLISM

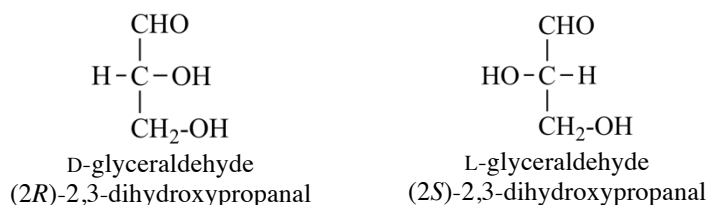
#### P-102.3.1 The Fischer projection of the acyclic form

In this representation of a monosaccharide, the carbon chain is written vertically with the lowest numbered carbon at the top, as indicated in P-102.2.2. To define the configuration, each carbon atom is considered in turn and placed in the plane of the paper. Neighboring carbon atoms are below, and the H atoms and OH groups are above the plane of the paper. Various representations 'b', 'c', 'd', 'e' and 'f' of a carbon atom in a monosaccharide in the Fischer projection are as follows (structure 'a' is a three-dimensional representation; the real Fischer projection is 'd'). The representation 'c' is commonly used in these recommendations.



#### P-102.3.2 The stereodescriptors 'D' and 'L'

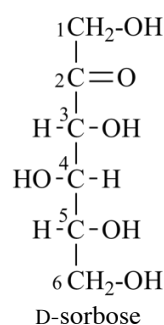
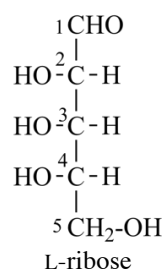
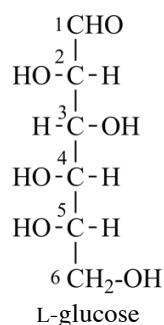
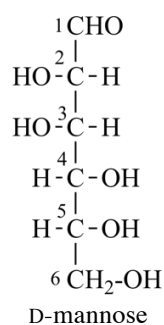
The simplest aldose is glyceraldehyde. It contains one center of chirality and occurs therefore in two enantiomeric forms, called D-glyceraldehyde and L-glyceraldehyde; these are represented by the Fischer projection formulas given below. It is known that these projections correspond to the absolute configurations. The configurational stereodescriptors 'D' and 'L' must be written in small capital letters and linked by a hyphen to the name of the sugar. The configuration is often described by the preferred CIP stereodescriptors *R* and *S*.



### P-102.3.3 The configurational atom

A monosaccharide is assigned to the 'D' or 'L' series according to the configuration of the highest numbered chirality center. This asymmetrically substituted carbon atom is called the 'configurational atom'. Thus if the hydroxy group projects to the right in the Fischer projection, the sugar belongs to the 'D' series, and receives the 'D' stereodescriptor.

Examples:



### P-102.3.4 Cyclic forms of monosaccharides

Most monosaccharides exist as cyclic hemiacetals or hemiketals. Two aspects of the internal cyclisation must be examined: first, the size of the ring, and secondly, the configuration of the newly created chirality center.

#### P-102.3.4.1 Ring size

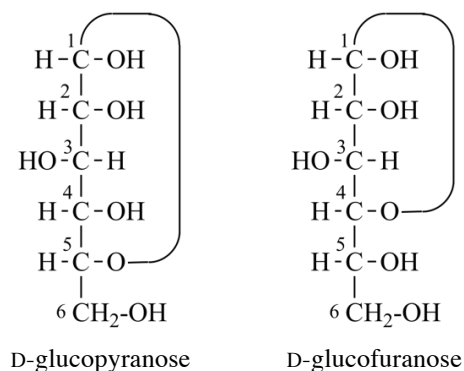
Out of the various possible heterocyclic ring sizes resulting from hemiacetal or hemiketal formation, those with five and six members, including an oxygen atom, prevail and are discussed in this Section. Their names are based on those of the parent heterocycles furan and pyran, respectively. Names are formed by including the terms 'furan' and 'pyran' before

the ending 'ose' in the name of a sugar. For example, D-mannose is changed to D-mannopyranose to indicate the cyclic form having a six-membered ring; furthermore, the generic term 'pyranose' includes all the sugars having a six-membered ring structure. Similarly, the sugars having a five-membered ring structure are 'furanoses'; oxiroses, oxetoses and septanoses have a three-, four- or seven-membered cyclic structure, respectively.

Different representations of cyclic forms are to be considered.

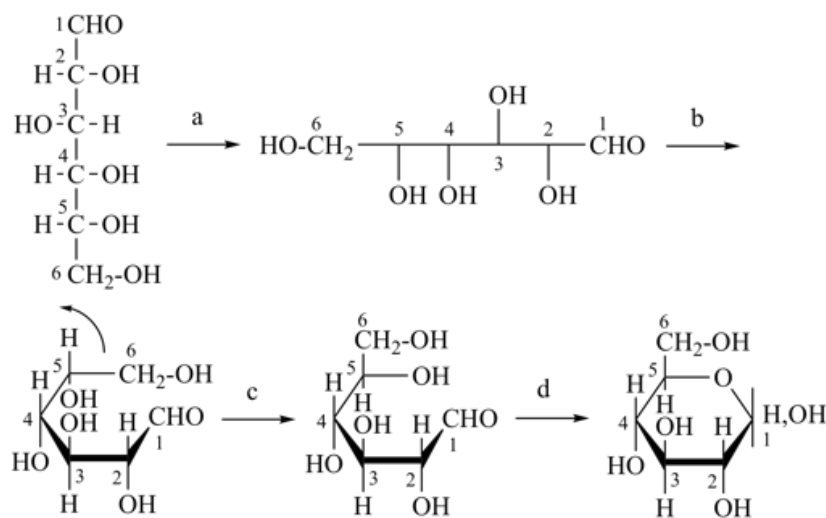
**P-102.3.4.1.1** Hemiacetal or hemiketal formation is indicated in the Fischer projection of the cyclic form by a long bond joining the original aldehydic or ketonic group to the oxygen atom included in the ring.

Examples:



#### P-102.3.4.1.2 The Haworth representation

The Haworth representation is a perspective drawing. The ring is orientated almost perpendicular to the plane of the paper, but viewed from slightly above so that the edge closer to the viewer is drawn below the most distant edge, with the oxygen behind and 'C-1' at the right hand end. The cyclisation process is envisaged as proceeding stepwise, as exemplified for D-glucopyranose in Fig. 10.1.



**Fig. 10.1** Reorientation of a Fischer projection to a Haworth projection

Two reorientations are necessary from the standard Fischer projection to prepare the acetalization or ketalization procedure: the first reorientation, step (a), consists in placing the nonterminal hydroxy groups vertically; the second reorientation, step (c), is the rotation at carbon C-5 to place the oxygen atom in the plane of the ring. The structure is defined completely by expressing the configuration at carbon '1'.

#### P-102.3.4.2 Anomeric forms; the stereodescriptors 'α' and 'β'

**P-102.3.4.2.1** In the cyclic form, the configuration of the newly created chirality center 'C-1' must be expressed. This center is called the 'anomeric center'. The two stereoisomers are called 'anomers'; they are designated by the stereodescriptors 'α' and 'β' according to the configurational relationship between the anomeric center and the so called 'reference center'.

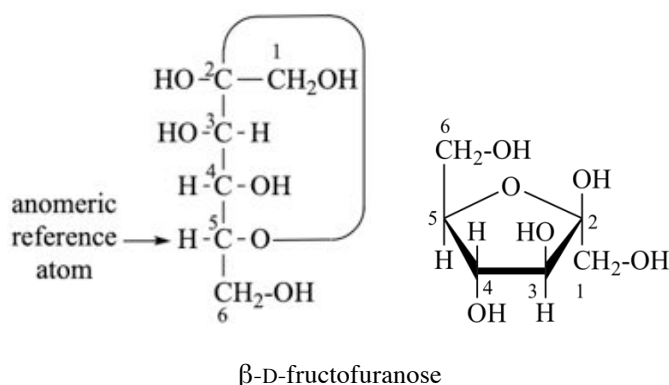
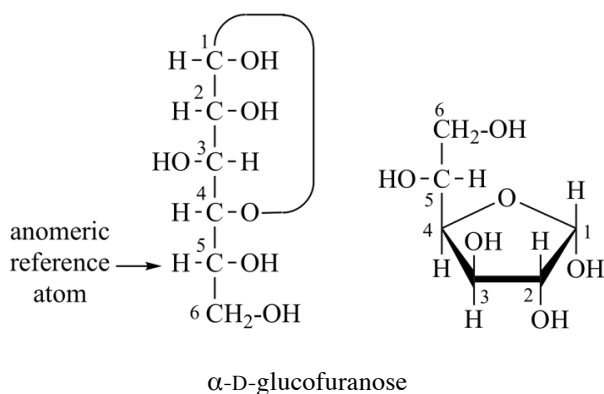
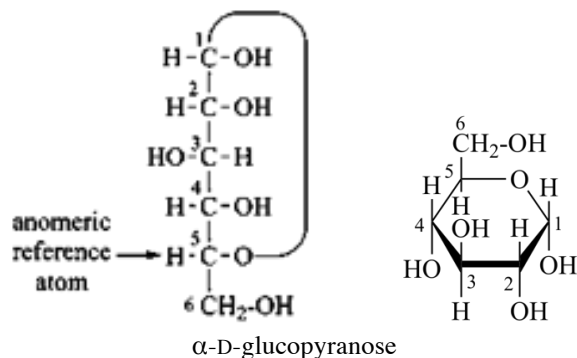
#### P-102.3.4.2.2 Configurations 'α' and 'β' for monosaccharides

The anomeric reference center in a monosaccharide having a retained name is the configurational atom as defined in P-102.3.3. In the Fischer projection, the α-anomer has the exocyclic oxygen atom at the anomeric center formally 'cis' to the oxygen atom attached to the anomeric reference atom; in the β-anomer, the relationship is 'trans'. The reference

plane for determining the configurations 'cis' and 'trans' is perpendicular to the Fischer projection, including all carbon atoms of the monosaccharide.

The anomeric stereodescriptor 'α' or 'β', followed by a hyphen, is placed immediately before the configurational stereodescriptor 'D' or 'L' of the carbohydrate name.

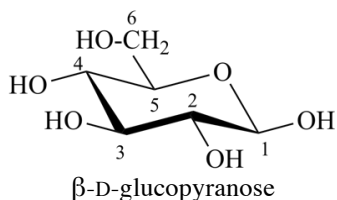
Examples:



### P-102.3.5 Conformation of monosaccharides

Pyranoses assume conformations that are not planar. For example β-D-glucopyranose assumes a chair conformation with characteristic substituent groups in equatorial conformations (hydrogen atoms attached to the ring are not shown):

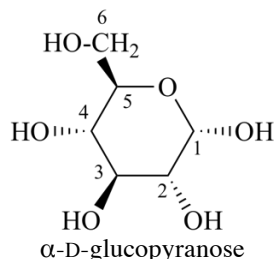
Example:



### P-102.3.6 The Mills depiction

In this depiction, the main hemiacetal ring is drawn in the plane of the paper. Hashed wedges denote substituents below this plane, and solid wedges those above.

Example:

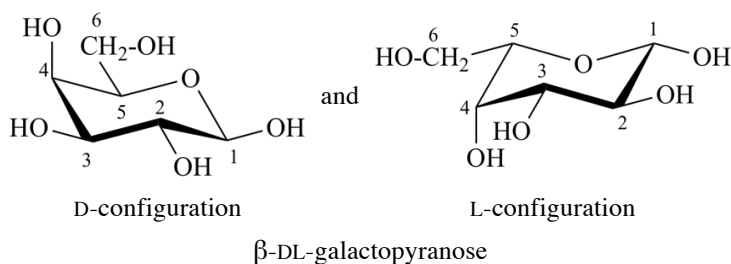
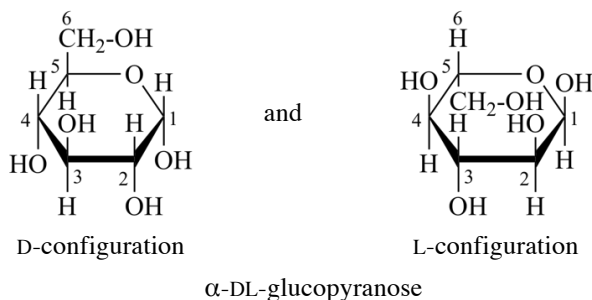


### P-102.3.7 Stereodescriptors for denoting racemates and uncertain configurations

#### P-102.3.7.1 Stereodescriptors for denoting racemates

Racemates are indicated by the stereodescriptor 'DL'.

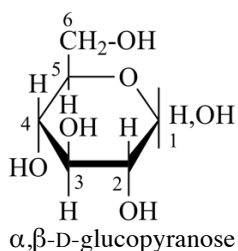
Examples:



#### P-102.3.7.2 Mixtures of anomers

When a mixture of anomers has to be described, the stereodescriptors ' $\alpha$ ' and ' $\beta$ ' are placed at the front of the name, separated by a comma; in Haworth representations, the symbols H and OH replace the formal bonds at the anomeric carbon atom.

Example:



### P-102.4 CHOICE OF PARENT STRUCTURE

In cases where more than one monosaccharide structure is embedded in a large molecule, a parent structure is chosen on the basis of the following criteria, applied in the order given until a decision is reached:

(a) the parent that includes the functional group most senior in the order of classes (see P-41). If there is a choice, it is made on the basis of the greatest number of occurrences of the most senior functional group.

Thus, ketoaldaric acid/aldaric acid > ketouronic acid/uronic and ketoaldonic acid/aldonic acid > dialdose > ketoaldose/aldose > diketose > ketose;

(b) the parent with the greatest number of carbon atoms in the chain, for example, heptose rather than hexose;

(c) the parent with the name that comes first in an alphabetical listing based on the following:



(i) the trivial name or the configurational prefix(es) of the systematic name, for example, glucose rather than gulose; a gluco rather than a gulo derivative;

Example: D-glucitol; not L-gulitol (see P-102.5.6.5.1);

(ii) the configurational symbol D rather than L;

Example: 5-*O*-methyl-D-galactitol; not 2-*O*-methyl-L-galactitol (see P-102.5.6.5.2);

(iii) the anomeric stereodescriptor  $\alpha$  rather than  $\beta$ ;

Example:  $\alpha$ -D-fructofuranose  $\beta$ -D-fructofuranose 1,2':1',2-dianhydride; not  $\beta$ -D-fructofuranose  $\alpha$ -D-fructofuranose 1,2':1',2-dianhydride (see P-102.5.6.7.2);

(d) the parent with the most substituent groups cited as prefixes (bridging substitution for example, 2,3-*O*-methylene is regarded as multiple substitution for this purpose); the prefixes 'deoxy' and 'anhydro' are detachable and alphabetized, thus regarded as substituent groups;

(e) the parent with the lowest locants for substituent prefixes;

Example: 2,3,5-tri-*O*-methyl-D-mannitol; not 2,4,5-tri-*O*-methyl-D-mannitol [see P-102.5.6.5.3 (a)]

(f) the parent with the lowest locant for the first cited substituent.

Example: 2-*O*-acetyl-5-*O*-methyl-D-mannitol; not 5-*O*-acetyl-2-*O*-methyl-D-mannitol [see P-102.5.6.5.3 (b)].

## **P-102.5 MONOSACCHARIDES: ALDOSES AND KETOSES; DEOXY AND AMINO SUGARS**

P-102.5.1 Aldoses

P-102.5.2 Ketoses

P-102.5.3 Deoxy sugars

P-102.5.4 Amino sugars

P-102.5.5 Thio sugars

P-102.5.6 Substituted monosaccharides

### **P-102.5.1 Aldoses**

Names of aldoses are retained or substitutively formed. Retained and semisystematic carbohydrate names for aldoses with three through six carbon atoms are listed in Table 10.2.

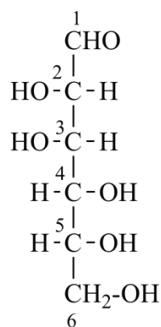
Names of aldoses having more than six carbon atoms are formed in two ways: by the procedures of systematic carbohydrate nomenclature, and by those of systematic substitutive nomenclature.

#### **P-102.5.1.1 Systematic carbohydrate names**

Systematic carbohydrate names of aldoses are formed from a stem name and a configurational prefix or prefixes. Stem names for the aldoses with three through ten carbon atoms are triose, tetrose, pentose, hexose, heptose, octose, nonose, and decose. The chain is numbered so that the carbonyl group receives the locant '1'.

**P-102.5.1.1.1** The configuration of  $>\text{CH-OH}$  groups of the sugar is designated by the configurational prefix(es) listed in Table 10.2, such as '*glycero*', '*gluco*', '*manno*', etc. Each name is qualified by a 'D' or 'L' stereodescriptor, as defined in P-102.3.2.

Example:



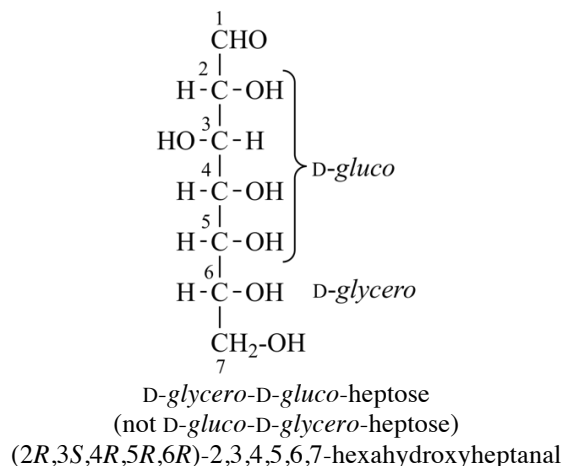
D-*manno*-hexose (systematic carbohydrate name)

D-mannose (retained name)

**P-102.5.1.1.2** Aldoses composed of more than four chirality centers are named by adding two or more configurational prefixes (listed in Table 10.2) to the stem name. Prefixes are assigned in order to the chirality centers in groups of four,

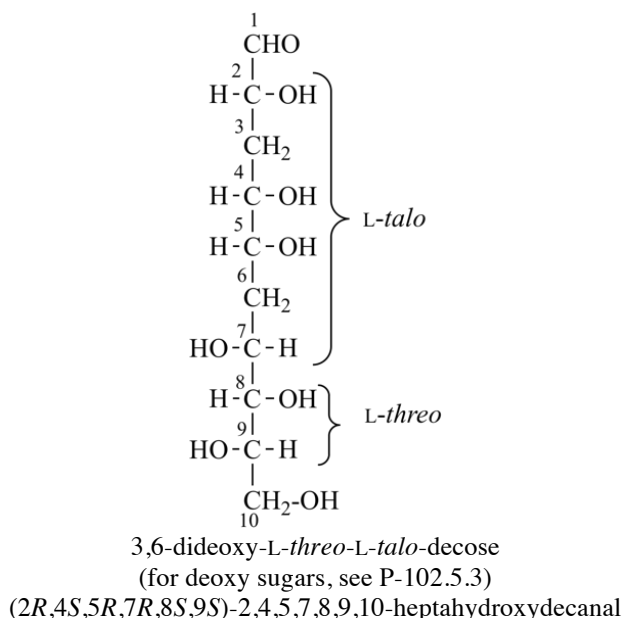
beginning with the group located next to the aldehydic group. The prefix relating to the group of carbon atoms farthest from the aldehydic group (which may contain fewer than four chirality centers) is cited first.

Example:



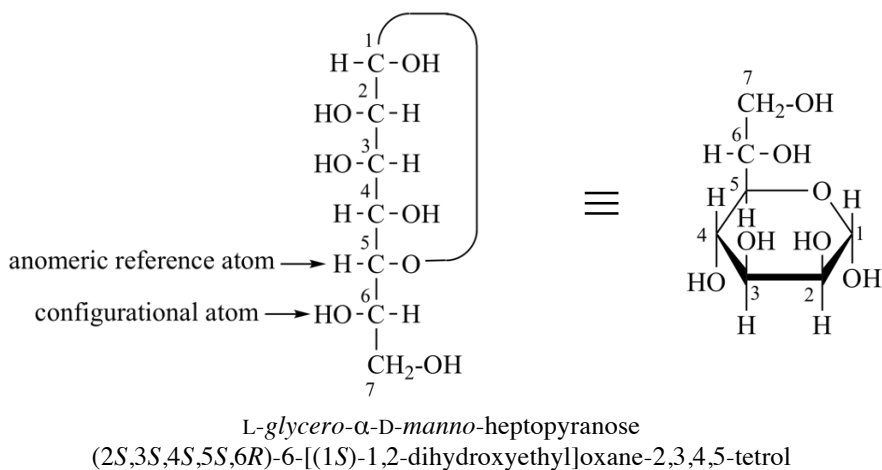
**P-102.5.1.1.3** When sequences of chirality centers are separated by nonchiral centers, the nonchiral centers are ignored, and the remaining set of chirality centers is assigned the appropriate configurational prefix (for four centers or less) or prefixes (for more than four centers).

Example:



#### P-102.5.1.1.4 Cyclic forms

For monosaccharides having more than six carbon atoms, the anomeric reference center is the highest numbered atom of the group of chirality centers next to the anomeric center that is involved in the heterocyclic ring and specified by a single configurational prefix. In the  $\alpha$ -anomer, the exocyclic oxygen atom at the anomeric center is formally 'cis', in the Fischer projection, to the oxygen atom attached to the anomeric reference atom; in the  $\beta$ -anomer these oxygen atoms are formally 'trans'.



## P-102.5.2 Ketoses

### P-102.5.2.1 Classification

Ketoses are classified as 2-ketoses, 3-ketoses, etc. according to the lowest locant for the position of the (potential) carbonyl group.

### P-102.5.2.2 Retained names

Retained names and structures are shown in Table 10.3; the configuration is specified by a 'D' or 'L' stereodescriptor, as defined in P-102.3.2.

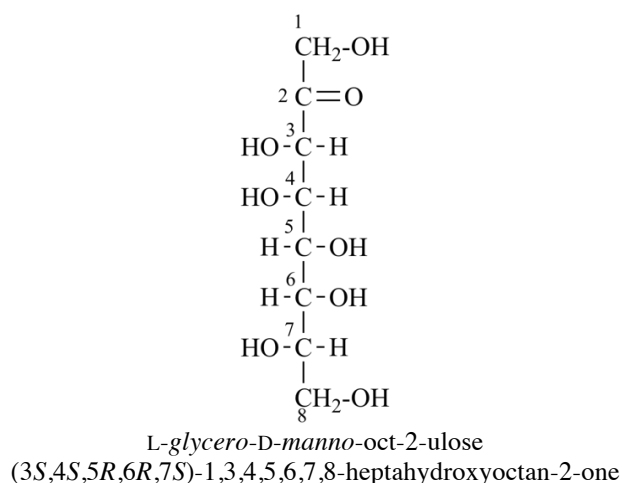
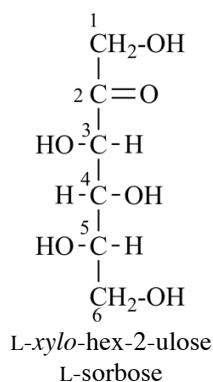
### P-102.5.2.3 Systematic carbohydrate names

The systematic carbohydrate names of ketoses having four through six carbon atoms are formed from the stem name and the appropriate configurational prefix listed in Table 10.3. The stem names are formed from the corresponding aldoses stem names by replacing the ending 'ose' with 'ulose', preceded by the locant of the carbonyl group, e.g. 'pent-2-ulose' and 'hex-3-ulose'. The chain is numbered so that the carbonyl group receives the lowest possible locant. When the carbonyl group is in the middle of a chain with an odd number of carbon atoms, a choice between alternative names is made according to P-102.4.

For 2-ketoses, configurational prefixes are given in the same way as for aldoses.

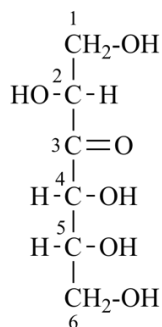
Substitutive names are given after the 'systematic carbohydrate names' to illustrate the two types of systematic names that are recommended in carbohydrate nomenclature.

Examples:

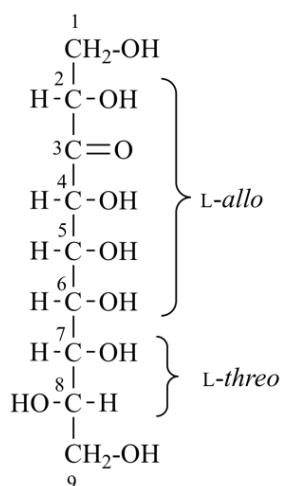


For ketoses with the carbonyl group at C-3, or at a higher-numbered carbon atom, the carbonyl group is ignored and the set of chirality centers is given the appropriate prefix or prefixes according to Table 10.3.

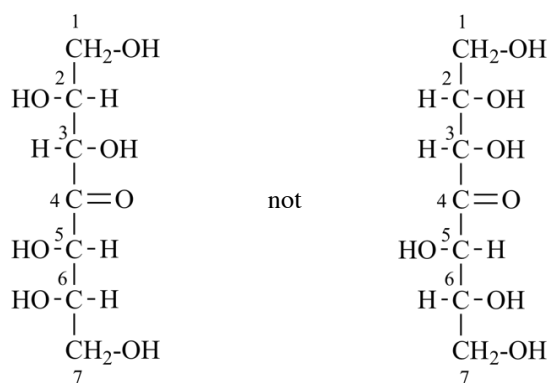
Examples:



D-*arabino*-hex-3-ulose  
(2*R*,4*R*,5*R*)-1,2,4,5,6-pentahydroxyhexan-3-one



L-*threo*-D-*allo*-non-3-ulose  
(2*S*,4*R*,5*R*,6*R*,7*R*,8*S*)-1,2,4,5,6,7,8,9-octahydroxynonan-3-one



L-*gluco*-hept-4-ulose

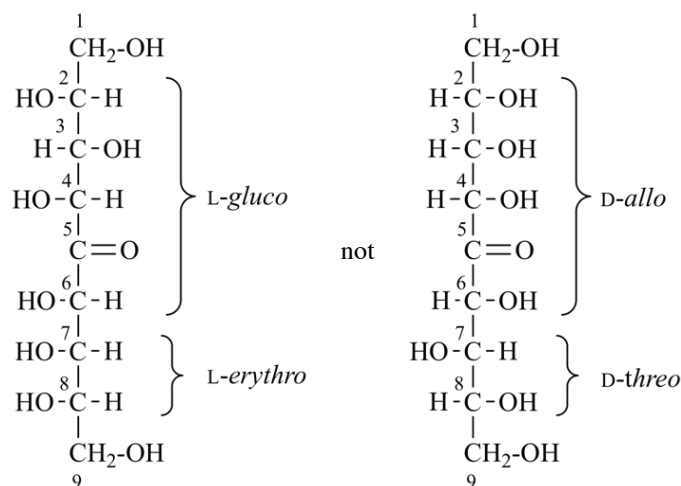
[not D-*gulo*-hept-4-ulose;

*gluco* is earlier in alphanumerical order, see P-102.4 (c)]

(2*R*,3*S*,5*S*,6*S*)-1,2,3,5,6,7-hexahydroxyheptan-4-one

[not (2*S*,3*S*,5*S*,6*R*)-1,2,3,5,6,7-hexahydroxyheptan-4-one;

when there is a choice, the *R* configuration is assigned the lowest locant, see P-14.4 (j)]



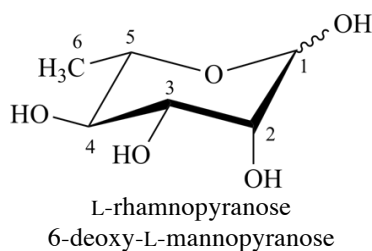
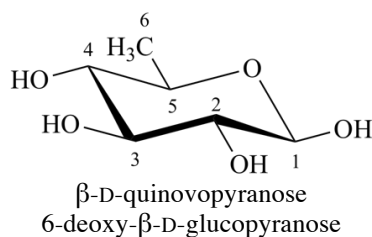
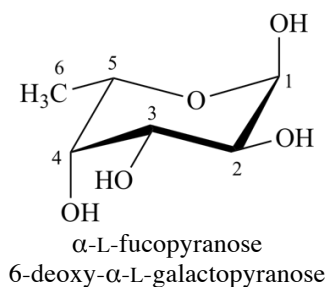
*L-erythro-L-gluco-non-5-ulose*  
 [not *D-threo-D-allo-non-5-ulose*;  
*erythro-gluco* is earlier in alphanumerical order]  
 (2*R*,3*S*,4*R*,6*S*,7*S*,8*S*)-1,2,3,4,6,7,8,9-octahydroxynonan-5-one

### P-102.5.3 Deoxy sugars

**P-102.5.3.1** The prefix ‘deoxy’ describes the removal of an ‘oxy’ group, –O–, with rejoining of the hydrogen atom. In these recommendations, the prefix ‘deoxy’ is classified as detachable; i.e., it is alphabetized among the substituents arising from substitutive nomenclature. This is a change from the previous status (see R-0.1.8.4, ref. 2) that classified the prefix ‘deoxy’ among nondetachable prefixes (see also the prefix ‘anhydro’ which is now classified as detachable and alphabetized among all detachable prefixes).

#### P-102.5.3.2 Trivial names.

The following names are retained: fucose, quinovose and rhamnose. The corresponding structures are shown in the pyranose form.



#### P-102.5.3.3 Carbohydrate names derived from retained names

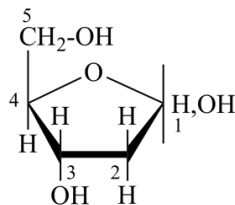
The prefix ‘deoxy’ is used in combination with a retained name when the deoxygenation does not involve the configuration at any chirality center, for example, 6-deoxy-D-allose. However the 6-deoxy derivatives of glucose, mannose, and galactose have their own retained trivial names (see P-102.5.3.2). When the prefix ‘deoxy’ modifies a chirality center, a carbohydrate name is preferred: names formed by substitutive nomenclature with CIP stereodescriptors are appropriate (see P-102.5.3.4 for examples).

The combination of 'amino' and 'deoxy' at the same position (and also the prefixes always cited as prefixes in substitutive nomenclature described in P-59.1.9 and 'deoxy' at the same position) is allowed.

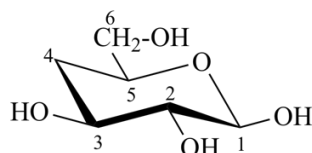
#### P-102.5.3.4 Systematic carbohydrate names

The systematic carbohydrate name consists of the prefix 'deoxy', preceded by the appropriate locant and followed by the stem name with such configurational prefixes as necessary to describe the chirality centers present in the deoxy compound. Configurational prefixes are cited in order commencing at the end furthest from C-1. The use of the prefix 'deoxy' with retained names of aldoses and ketoses is not recommended.

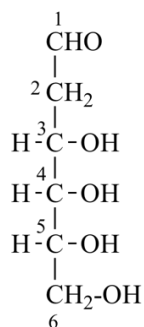
Examples:



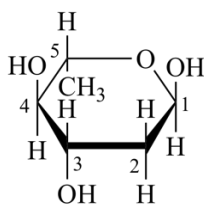
2-deoxy-D-*erythro*-pentofuranose  
(often referred to as 2-deoxy-D-ribofuranose  
or 2-deoxy-D-ribose)  
(2*ξ*,4*S*,5*R*)-5-(hydroxymethyl)oxolane-2,4-diol



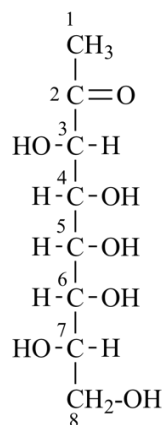
4-deoxy-β-D-*xylo*-hexopyranose  
(not 4-deoxy-β-D-galactopyranose)  
(2*R*,3*R*,4*S*,6*S*)-6-(hydroxymethyl)oxane-2,3,4-triol



2-deoxy-D-*ribo*-hexose (not 2-deoxy-D-allose)  
(3*S*,4*S*,5*R*)-3,4,5,6-tetrahydroxyhexanal



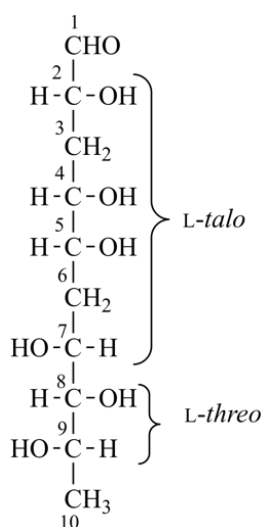
2,6-dideoxy-α-L-*arabino*-hexopyranose  
(2*R*,4*S*,5*R*,6*S*)-6-methyloxane-2,4,5-triol



1-deoxy-L-glycero-D-altro-oct-2-ulose  
(3*S*,4*R*,5*R*,6*R*,7*S*)-3,4,5,6,7,8-hexahydroxyoctan-2-one

When the  $-\text{CH}_2-$  group divides the chirality centers into two sets, it is ignored for the purpose of assigning the configurational prefix; the prefix(es) assigned should cover the entire sequence of chirality centers (see aldoses) (see P-102.5.1.1.3).

Example:



3,6,10-trideoxy-L-threo-L-talo-decose  
(2*R*,4*S*,5*R*,7*R*,8*R*,9*S*)-2,4,5,7,8,9-hexahydroxydecanal

#### P-102.5.4 Amino sugars

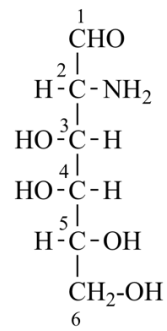
The replacement of a hydroxy group that is not an anomeric hydroxy group of a monosaccharide or a monosaccharide derivative by an amino group is envisaged as substitution of the appropriate hydrogen atom of the corresponding deoxy monosaccharide by an amino group. The configuration at the carbon atom carrying the amino group is expressed as that of an aldose, considering that the amino group has replaced a hydroxy group.

To the contrary, the replacement of a hydroxy group by a sulfanyl group is considered to be a functional replacement indicated by the prefix 'thio'.

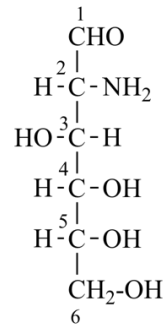
##### P-102.5.4.1 Amino sugars

###### P-102.5.4.1.1 Trivial names

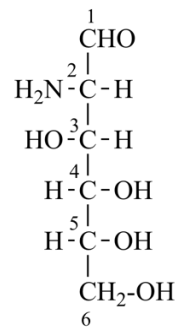
The following glycosamine names are retained.



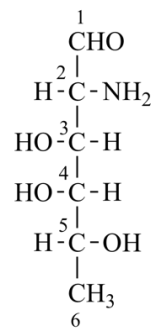
D-galactosamine  
2-amino-2-deoxy-D-galactose



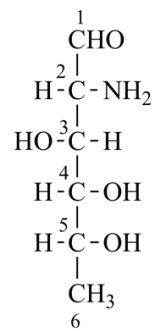
D-glucosamine  
2-amino-2-deoxy-D-glucose



D-mannosamine  
2-amino-2-deoxy-D-mannose

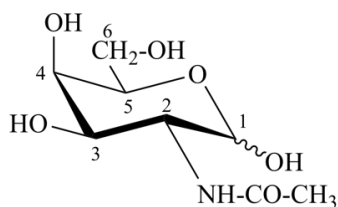


D-fucosamine  
2-amino-2,6-dideoxy-D-galactose



D-quinovosamine  
2-amino-2,6-dideoxy-D-glucose



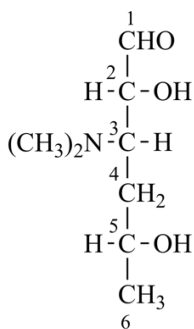


*N*-acetyl-D-galactosamine  
2-acetamido-2-deoxy-D-galactopyranose

#### P-102.5.4.1.2 Systematic carbohydrate names

Systematic carbohydrate names are formed, in two steps: in a first step a deoxy sugar is created by deoxygenation at the carbon atom where the amino group is to be introduced by substitution in a second step. Names of substituted amines are formed by using the name of the substituted amino group as a prefix.

Example:



3,4,6-trideoxy-3-(dimethylamino)-D-xylo-hexose  
(2*R*,3*S*,5*R*)-3-(dimethylamino)-2,5-dihydroxyhexanal

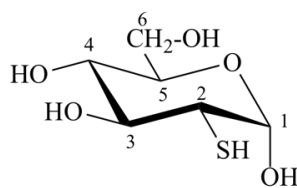
#### P-102.5.5 Thio sugars and other chalcogen analogues

The replacement of a hydroxy oxygen atom of an aldose or ketose, or of the oxygen atom of the carbonyl group of an acyclic aldose or ketose, by sulfur, selenium or tellurium is indicated by placing the prefix 'thio', 'seleno' or 'telluro', respectively, preceded by the appropriate locant, at the front of the systematic or trivial name of the aldose or ketose. In carbohydrate nomenclature, the prefixes 'thio', 'seleno' and 'telluro' are considered as detachable, alphabetized prefixes.

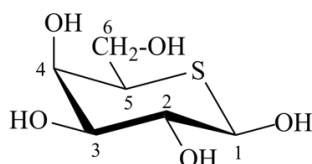
Replacement of the ring oxygen atom of the cyclic form of an aldose or ketose by sulfur, selenium, or tellurium is indicated in the same way, the number of the non-anomeric adjacent carbon atom of the ring being used as locant. In such a case, skeletal replacement expressed by an 'a' replacement prefix is not recommended.

Sulfoxides (and selenoxides or telluroxides) and sulfones (and selenones or tellurones) are named by functional class nomenclature (see P-63.6 for functional class names of sulfoxides and sulfones).

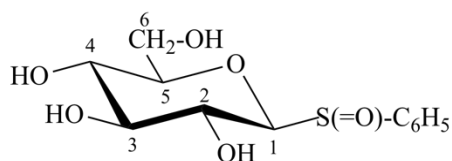
Examples:



2-thio- $\alpha$ -D-glucopyranose



5-thio- $\beta$ -D-galactopyranose



$\beta$ -D-glucopyranosyl phenyl sulfoxide  
(for glycosyl groups, see P-102.6.1.1)  
(2*S*,3*R*,4*S*,5*S*,6*R*)-2-(benzenesulfinyl)-6-(hydroxymethyl)oxane-3,4,5-triol

## P-102.5.6 Monosaccharide derivatives

P-102.5.6.1 *O*-Substitution

P-102.5.6.2 Glycosides

P-102.5.6.3 *C*-Substitution

P-102.5.6.4 *N*-Substitution

P-102.5.6.5 Alditols

P-102.5.6.6 Monosaccharide carboxylic acids

P-102.5.6.7 Anhydrides

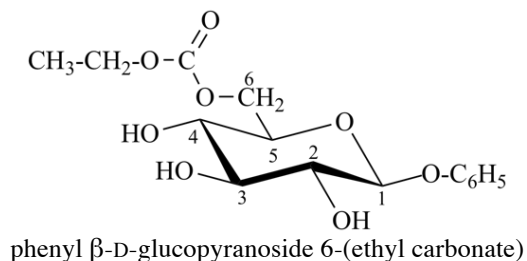
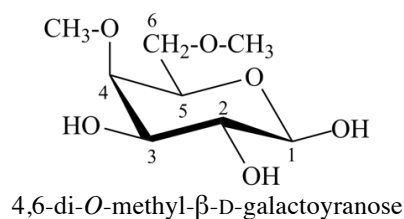
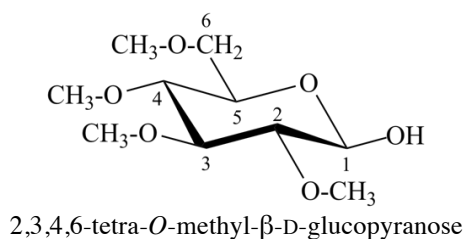
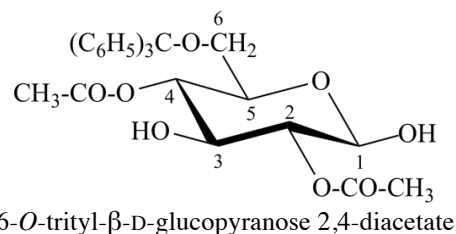
### P-102.5.6.1 *O*-Substitution

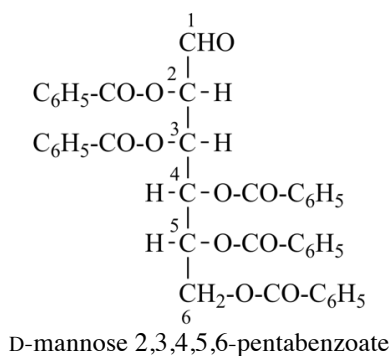
In order to maintain the integrity of structures and take advantage of retained names to imply the absolute configuration, *O*-substitution is allowed in carbohydrate nomenclature. Substituents replacing the hydrogen atom of an alcoholic hydroxy group of a monosaccharide or monosaccharide derivative are denoted as *O*-substituents. The substitution of an anomeric hydroxy group is discussed in P-102.5.6.3.2. The *O*-locant is not repeated for multiple substitution by the same atom or group. Number locants are used as necessary to specify the positions of substituents; they are not required for compounds fully substituted by identical atoms or groups.

#### P-102.5.6.1.1 *O*-Acetyl and *O*-alkyl functionalization.

For *O*-acyl derivatives, names with the acid component cited as a separate word ending in 'ate' after the monosaccharide name are preferred to names using *O*-acyl group prefixes. However when the ose ending is changed (e.g. to denote a glycosyl or an acid function having seniority over an ester) *O*-acyl prefixes are required. *O*-Alkyl derivatives are always expressed by prefixes.

Examples:





### P-102.5.6.1.2 Phosphoric acid esters

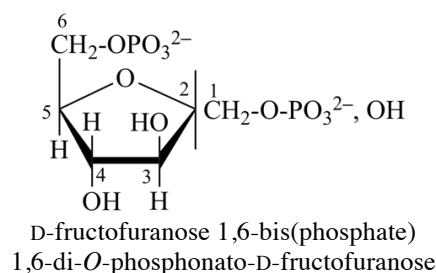
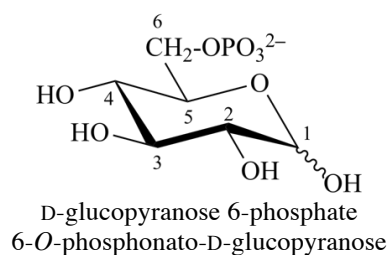
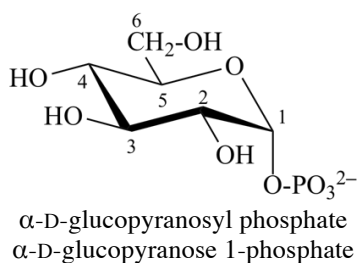
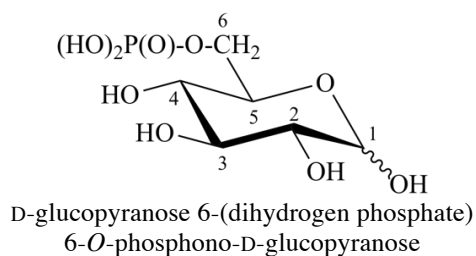
Esters of sugars with phosphoric acid are generally termed 'phosphates'. In biochemical usage, the term 'phosphate' indicates the phosphate residue regardless of the state of ionization or the counter ions present. However, systematically the names must differentiate between a true phosphate,  $-\text{O-PO}(\text{O}^-)_2$ , and an acid phosphate, i.e.,  $-\text{O-PO}(\text{OH})_2$ , called a (dihydrogen phosphate). The prefixes 'phosphono', for  $-\text{PO}(\text{OH})_2$ , and 'phosphonato', for  $\text{PO}(\text{O}^-)_2$ , are also used, to denote *O*-phosphonic acid derivatives.

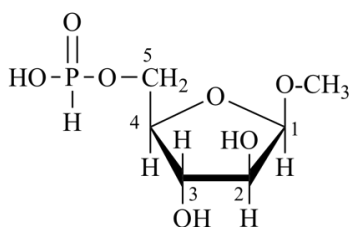
The term 'phospho' is used in place of 'phosphono' and 'phosphonato' in biochemical contexts.

When the sugar is esterified by two or more phosphate groups, the numerical terms 'bis', 'tris' are used, as 'bis(phosphate)', 'tris(phosphate)'.

Phosphonates are treated in the same way as phosphates.

Examples:



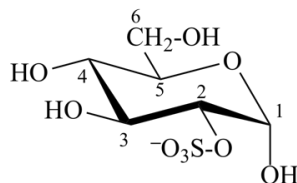


methyl  $\beta$ -D-arabinofuranoside 5-(hydrogen phosphonate)  
methyl 5-deoxy- $\beta$ -D-arabinofuranosid-5-yl hydrogen phosphonate

### P-102.5.6.1.3 Esters with sulfuric acid

Esters of sugars with sulfuric acid are named by adding the term 'sulfate' after the name of the sugar, with the appropriate locant. The prefixes 'sulfo' for  $-\text{SO}_3\text{H}$ , and 'sulfonato' for  $\text{SO}_3^-$ , can be used to denote *O*-derivatives.

Example:



$\alpha$ -D-glucopyranose 2-sulfate  
2-*O*-sulfonato- $\alpha$ -D-glucopyranose

### P-102.5.6.2 Glycosides

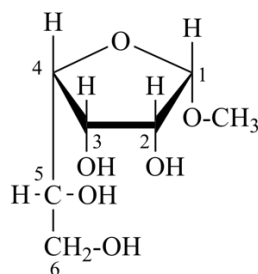
#### P-102.5.6.2.1 Definitions

Glycoside is a less frequently used term for a monosaccharide. Glycosides are mixed acetals (ketals) derived from cyclic forms of monosaccharides, thus, having an *O*-substituted anomeric  $-\text{OH}$  group, such as  $-\text{OR}$ . See ref. 27 for a full discussion on the use of the term glycoside.

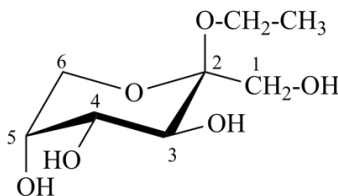
#### P-102.5.6.2.2 Names

Glycosides are named by using functional class nomenclature. The name of the class 'glycoside' is adapted to the name of each cyclic monosaccharide, by changing the letter 'e' at the end of the name to 'ide', for example glucopyranose becomes glucopyranoside, fructofuranose becomes fructofuranoside. The class name is preceded, as a separate word, by the name of the substituent group that is part of the acetal or ketal function.

Examples:



methyl  $\alpha$ -D-gulofuranoside



ethyl  $\beta$ -D-fructopyranoside

### P-102.5.6.3 C-Substitution

P-102.5.6.3.1 Substitution at a nonterminal carbon atom

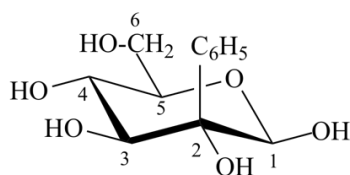
P-102.5.6.3.2 Substitution replacing a nonterminal hydroxy group

P-102.5.6.3.3 Substitution at a terminal carbon atom

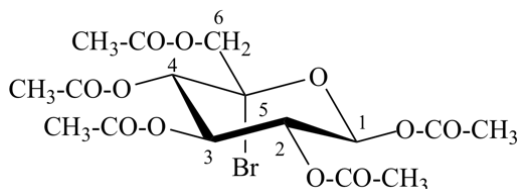
#### P-102.5.6.3.1 Substitution at a nonterminal carbon atom

The compound is named as a *C*-substituted monosaccharide. The group having priority in accordance with the CIP priority system is regarded as equivalent to –OH for assignment of configuration. Any ambiguity (e.g. at a carbon atom where ring formation occurs) is avoided by using the *R,S* system to specify the configuration at the modified chirality center.

Examples:



2-*C*-phenyl- $\beta$ -D-glucopyranose  
(2*R*,3*R*,4*S*,5*S*,6*R*)-6-(hydroxymethyl)-3-phenyloxane-2,3,4,5-tetrol

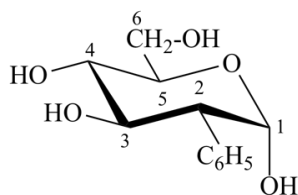


5-*C*-bromo- $\beta$ -D-glucopyranose pentaacetate  
(2*R*,3*R*,4*R*,5*S*,6*S*)-6-[(acetyloxy)methyl]-6-bromooxane-2,3,4,5-tetrayl tetraacetate

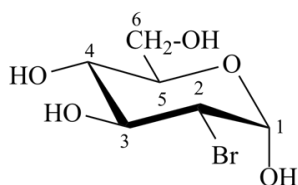
#### P-102.5.6.3.2 Substitution replacing a nonterminal hydroxy group

The compound is named as a substituted derivative of a deoxy sugar. The group replacing the –OH group determines the configuration. Any potential ambiguity must be dealt with by the use of the *R,S* system. The *R,S* system must be used to assign the preferred configuration of a chirality center twice substituted; this method is preferable to that establishing the configuration by making the substituent with high CIP priority equivalent to the –OH group.

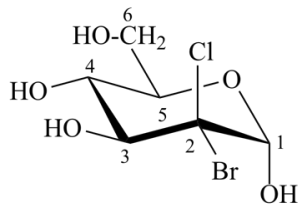
Examples:



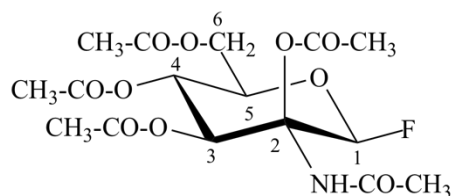
2-deoxy-2-phenyl- $\alpha$ -D-glucopyranose  
2-deoxy-2-*C*-phenyl- $\alpha$ -D-glucopyranose  
(2*R*)-2-deoxy-2-phenyl- $\alpha$ -D-*arabino*-hexopyranose  
(2*S*,3*R*,4*R*,5*S*,6*R*)-6-(hydroxymethyl)-3-phenyloxane-2,4,5-triol



2-bromo-2-deoxy- $\alpha$ -D-glucopyranose



(2*R*)-2-bromo-2-chloro-2-deoxy- $\alpha$ -D-*arabino*-hexopyranose  
2-bromo-2-chloro-2-deoxy- $\alpha$ -D-glucopyranose  
(2*S*,3*R*,4*S*,5*S*,6*R*)-3-bromo-3-chloro-6-(hydroxymethyl)oxane-2,4,5-triol

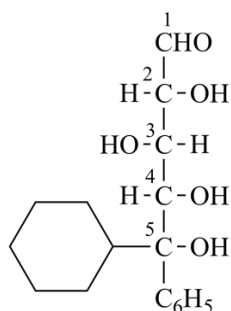


2-C-acetamido-2,3,4,6-tetra-*O*-acetyl- $\beta$ -D-mannopyranosyl fluoride  
 (2*S*,3*S*,4*S*,5*R*,6*R*)-3-acetamido-6-[(acetyloxy)methyl]-2-fluorooxane-3,4,5-triyl triacetate

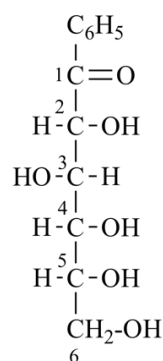
### P-102.5.6.3.3 Substitution at a terminal carbon atom

Substitution at a terminal carbon atom of a carbohydrate chain creates a new chirality center; the configuration is indicated by the '*R/S*' system. Preferred names are formed substitutively.

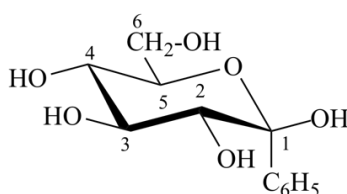
Examples:



(5*R*)-5-*C*-cyclohexyl-5-*C*-phenyl-D-xylose  
 (2*R*,3*S*,4*S*,5*R*)-5-cyclohexyl-2,3,4,5-tetrahydroxy-5-phenylpentanal



1-phenyl-D-glucose  
 (2*R*,3*S*,4*R*,5*R*)-2,3,4,5,6-pentahydroxy-1-phenylhexan-1-one



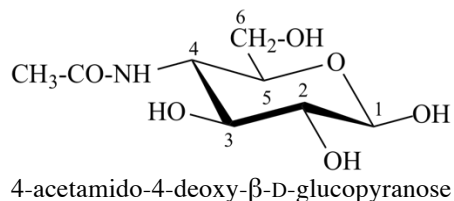
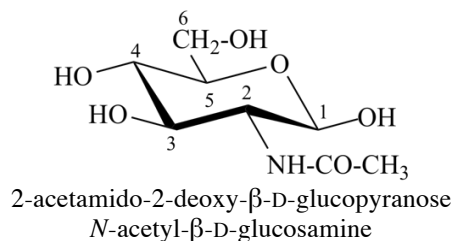
1-*C*-phenyl- $\beta$ -D-glucopyranose  
 (2*R*,3*R*,4*S*,5*S*,6*R*)-6-(hydroxymethyl)-2-phenyloxane-2,3,4,5-tetrol

### P-102.5.6.4 *N*-Substitution

Substitution at the  $\text{-NH}_2$  group of an amino sugar is dealt with in two different ways:

- (1) The whole substituted amino group is designated as a prefix as in 2-acetamido-2-deoxy-D-glucose or 2-(butylamino)-2-deoxy-D-glucose.
- (2) If the amino sugar has a retained trivial name, the substitution is indicated by a prefix preceded by the capital italicized letter *N*.

Examples:



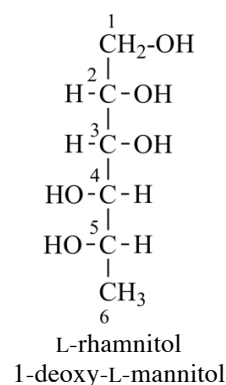
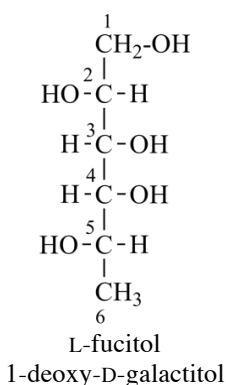
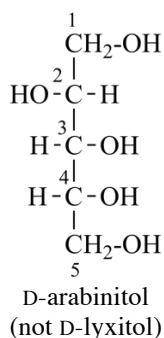
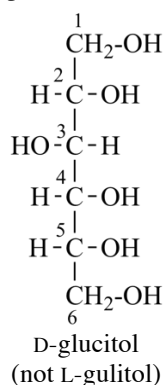
### P-102.5.6.5 Alditols

Alditols are named by changing the ending 'ose' in the name of the corresponding aldose into 'itol'.

#### P-102.5.6.5.1 Choice of a parent structure

When the same alditol can be derived from either of two different aldoses, or from an aldose or a ketose, the recommended structure is derived from Rule P-102.4, with the exception of the retained names fucitol and rhamnitol.

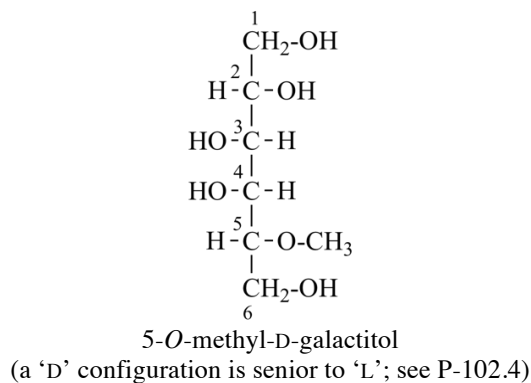
Examples:

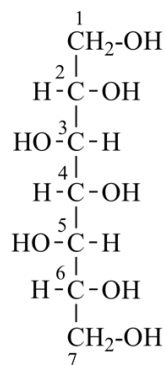


#### P-102.5.6.5.2 meso-Forms

The prefix 'meso' may be included in the preferred names of erythritol, ribitol and galactitol. The stereodescriptor 'D' or 'L' must be given when a derivative of a 'meso' form has become asymmetric by substitution. It is also necessary to use the stereodescriptor 'D' or 'L' in the case where there are more than four contiguous chirality centers.

Examples:





*meso*-D-glycero-L-ido-heptitol

(a 'D' configuration is senior to 'L': see P-102.4)

(2*R*,3*S*,4*r*,5*R*,6*S*)-heptane-1,2,3,4,5,6,7-heptol

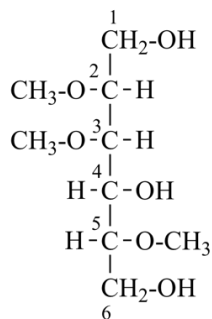
(note that locant '1' must be shifted to the other end of the alditol)

### P-102.5.6.5.3 Choice of parent structure for substituted alditols

The parent structure must have:

(a) the lowest locants for substituent prefixes in accordance with criterion (e) in Rule P-102.4;

Example:

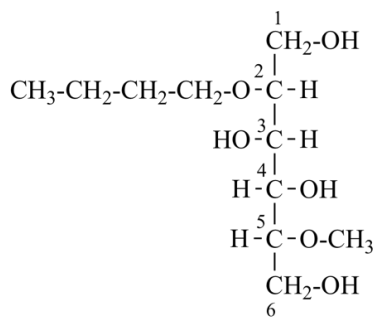


2,3,5-tri-*O*-methyl-D-mannitol

(not 2,4,5-tri-*O*-methyl-D-mannitol)

(b) the lowest locant for the first cited substituent in alphanumerical order, in accordance with criterion (f) in Rule P-102.4.

Example:



2-*O*-butyl-5-*O*-methyl-D-mannitol

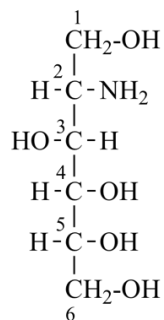
(not 5-*O*-butyl-2-*O*-methyl-D-mannitol)

### P-102.5.6.5.4 Aminoalditols

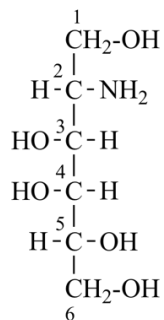
Alditols derived from galactosamine and glucosamine are aminoalditols. They have retained names, galactosaminitol and glucosaminitol, respectively.



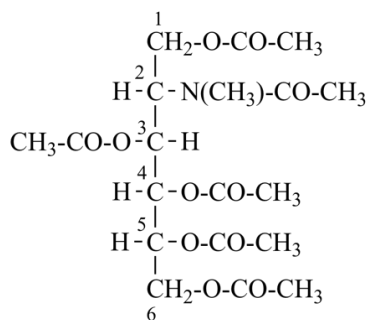
Examples:



D-glucosaminitol  
2-amino-2-deoxy-D-glucitol



D-galactosaminitol  
2-amino-2-deoxy-D-galactitol



2-deoxy-2-(*N*-methylacetamido)-D-glucitol 1,3,4,5,6-pentaacetate

#### P-102.5.6.6 Monosaccharide carboxylic acids

##### P-102.5.6.6.1 Classifications

P-102.5.6.6.2 Aldonic acids. Monocarboxylic acids formally derived from aldoses by oxidation of the aldehydic group to a carboxylic acid are called aldonic acids.

P-102.5.6.6.3 Ketoaldonic acids. Oxo carboxylic acids formally derived from aldonic acids by oxidation of a secondary –CHOH group to a carbonyl group are called ketoaldonic acids.

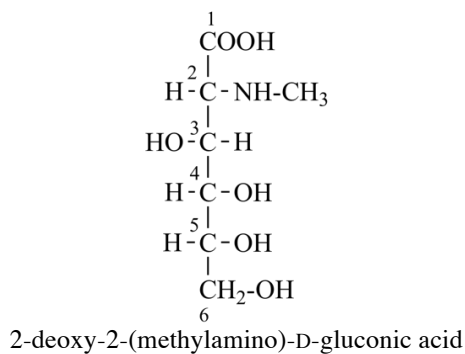
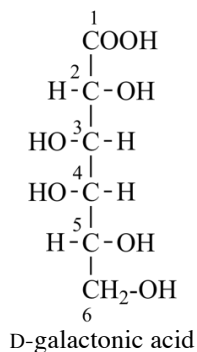
P-102.5.6.6.4 Uronic acids. Carboxylic acids formally derived from aldoses by oxidation of the terminal –CH<sub>2</sub>OH group to a carboxy group are called uronic acids.

P-102.5.6.6.5 Aldaric acids. Carboxylic acids formed by the oxidation of both terminal groups (–CHO and –CH<sub>2</sub>OH) of aldoses to carboxy groups are called aldaric acids.

##### P-102.5.6.6.2 Aldonic acids

Aldonic acids are divided into aldotronic acids, aldotetronic acids, etc. according to the number of carbon atoms in the chain. The names of individual compounds are formed by changing the ending ‘ose’ of the retained or systematic name of the aldose to ‘onic acid’. The locant 1 is assigned to the carboxy group.

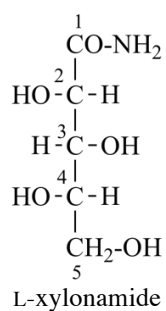
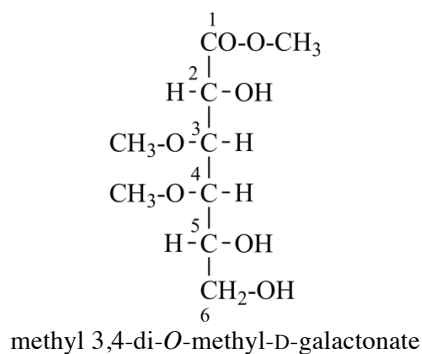
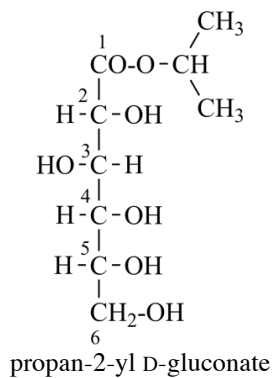
Examples:

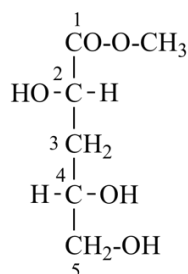


**P-102.5.6.6.2.1** Derivatives of aldonic acids

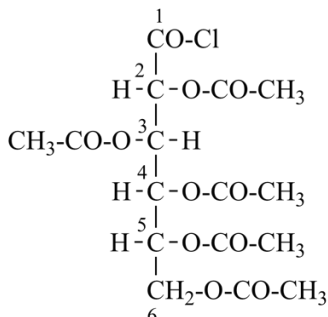
Aldonic acids are treated as carboxylic acids having a retained name. They can form salts, esters, anhydrides, acyl groups and acid halides and pseudohalides, amides, hydrazides, nitriles and chalcogen analogues as described in Sections P-65 and P-66 for systematic nomenclature.

Examples:





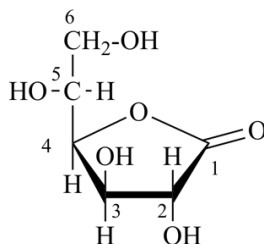
methyl 3-deoxy-D-*threo*-pentonate



2,3,4,5,6-penta-*O*-acetyl-D-gluconoyl chloride

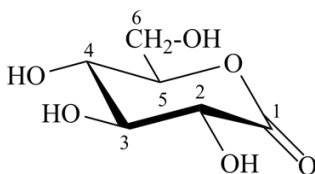
**P-102.5.6.6.2.2** Lactones and lactams are named by adapting Rules P-65.6.3.5.1 and P-66.1.5, respectively. Two locants are used before the lactone or lactam term: the first one is the locant 1 denoting the carboxy group position; the second locant denotes the position of attachment on the carbon chain. To name lactams, the amino group,  $-\text{NH}_2$ , must be generated and cited. The use of Greek letters to indicate the size of a lactone or lactam ring is not recommended. Names can also be formed substitutively on the basis of heterocyclic rings in accordance with the rules described in Chapters P-1 through P-9.

Examples:



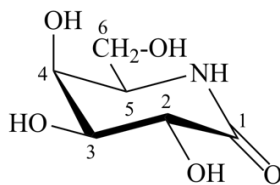
D-glucono-1,4-lactone

(3*R*,4*R*,5*R*)-5-[(1*R*)-1,2-dihydroxyethyl]-3,4-dihydroxyoxolan-2-one



D-glucono-1,5-lactone

(3*R*,4*S*,5*S*,6*R*)-3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-one



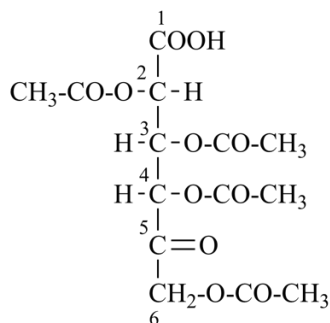
5-amino-5-deoxy-D-galactono-1,5-lactam

(3*R*,4*S*,5*S*,6*R*)-3,4,5-trihydroxy-6-(hydroxymethyl)piperidin-2-one

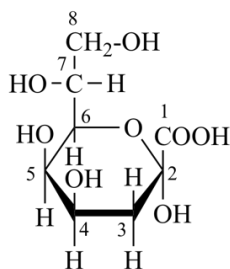
### P-102.5.6.6.3 Ketoaldonic acids

**P-102.5.6.6.3.1** Names of individual ketoaldonic acids are formed by changing the ending 'ulose' in the name of the corresponding ketose to 'ulosonic acid', preceded by the locant of the ketonic group. The numbering starts at the carboxy group.

Examples:



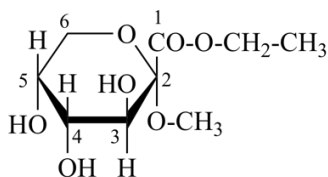
2,3,4,6-tetra-*O*-acetyl-*D*-arabino-hex-5-ulosonic acid  
 (2*S*,3*R*,4*S*)-2,3,4,6-tetrakis(acetyloxy)-5-oxohexanoic acid



3-deoxy- $\alpha$ -*D*-manno-oct-2-ulopyranosonic acid  
 (2*R*,4*R*,5*R*,6*R*)-6-[(1*R*)-1,2-dihydroxyethyl]-2,4,5-trihydroxyoxane-2-carboxylic acid

**P-102.5.6.6.3.2** Glycosides of ketoaldonic acids are named by changing the component 'pyranose' into 'pyranoside' in the name, to give '-ulopyranosidonic acid'. Names of derivatives of ketoaldonic acids are formed as described in P-102.5.6.6.2 for aldonic acids. When a glycoside is esterified, parentheses are used to isolate the glycosidic portion of the name.

Example:

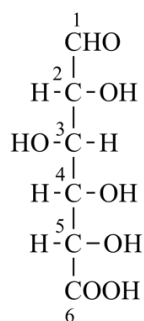


ethyl (methyl  $\alpha$ -*D*-fructopyranosid)onate  
 ethyl (methyl  $\alpha$ -*D*-arabino-hex-2-ulopyranosid)onate  
 ethyl (2*R*,3*S*,4*R*,5*R*)-3,4,5-trihydroxy-2-methoxyoxane-2-carboxylate

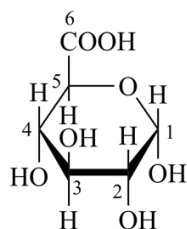
#### P-102.5.6.6.4 Uronic acids

**P-102.5.6.6.4.1** Names of individual uronic acids are formed by changing the ending 'ose' in the retained or systematic name of the corresponding aldose to 'uronic acid'. The numbering of the aldose is kept intact; the locant '1' is still assigned to the (potential) aldehydic group.

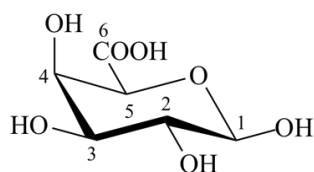
Examples:



*D*-glucuronic acid



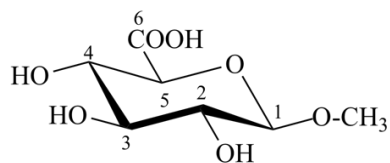
$\alpha$ -D-glucopyranuronic acid



$\beta$ -D-galactopyranuronic acid

**P-102.5.6.6.4.2** Glycosides of uronic acids are named by changing the 'pyran' component in the name of the acid to 'pyranoside', with elision of the final letter 'e', to give 'pyranosiduronic acid'.

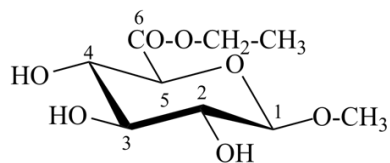
Example:



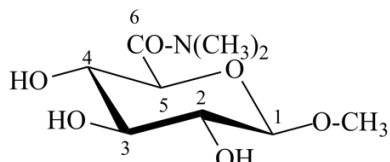
methyl  $\beta$ -D-glucopyranosiduronic acid

**P-102.5.6.6.4.3** Derivatives of uronic acid are named as indicated in P-102 and P-65 and P-66.

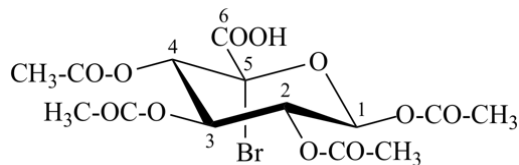
Examples:



ethyl (methyl  $\beta$ -D-glucopyranosid)uronate



*N,N*-dimethyl(methyl  $\beta$ -D-glucopyranosid)uronamide

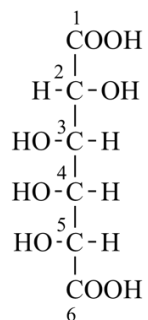


(5*R*)-1,2,3,4-tetra-*O*-acetyl-5-*C*-bromo- $\alpha$ -D-xylo-hexopyranuronic acid  
(2*R*,3*S*,4*R*,5*R*,6*R*)-3,4,5,6-tetrakis(acetyloxy)-2-bromooxane-2-carboxylic acid

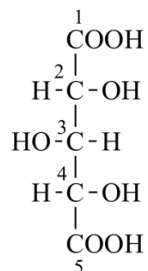
**P-102.5.6.6.5** Aldaric acids

**P-102.5.6.6.5.1** Names of aldaric acids are formed by changing the 'ose' ending in retained or systematic names of parent aldoses to 'aric acid'. Choice of a parent structure is made in accordance with P-102.4 and P-102.5.6.5.1. The stereodescriptor '*meso*' may be added for sake of clarity to the names of the appropriate aldaric acids.

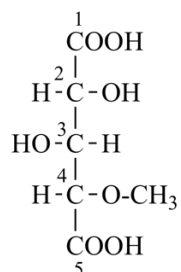
Examples:



L-altraric acid  
(not L-talaric acid)



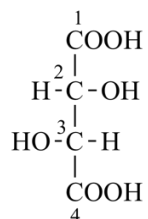
*meso*-xylaric acid



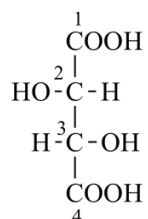
4-*O*-methyl-D-xylaric acid  
(not 2-*O*-methyl-L-xylaric acid)

**P-102.5.6.6.5.2** Tartaric acid is the retained name to describe the aldaric acids corresponding to the parent aldoses, erythrose and threose. '*R*' and '*S*' are preferred stereodescriptors for denoting the configuration of tartaric acid. Salts and esters are referred to as tartrates.

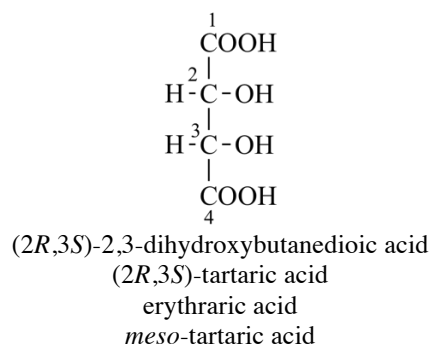
Examples:



(2*R*,3*R*)-2,3-dihydroxybutanedioic acid  
(2*R*,3*R*)-tartaric acid  
L-threonic acid  
(+)-tartaric acid

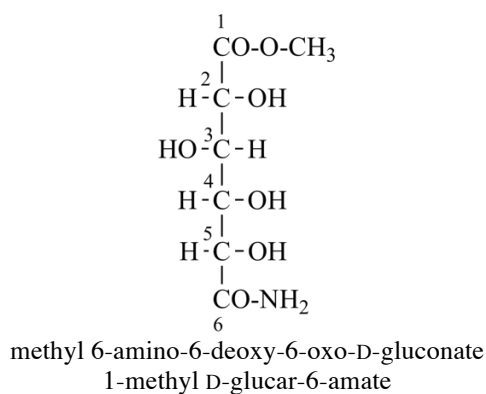
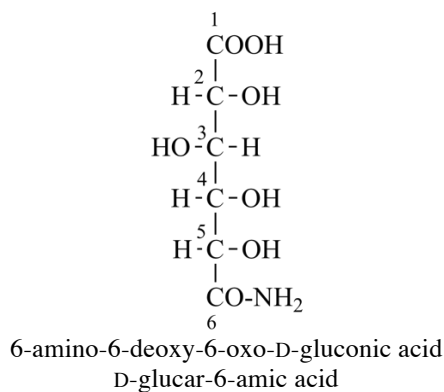
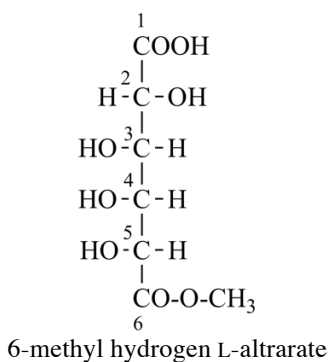
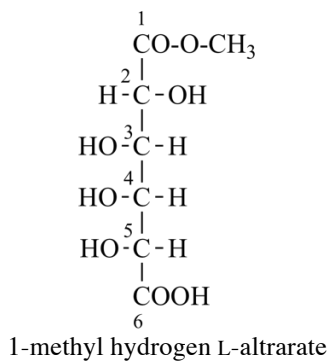


(2*S*,3*S*)-2,3-dihydroxybutanedioic acid  
(2*S*,3*S*)-tartaric acid  
D-threonic acid  
(-)-tartaric acid



**P-102.5.6.6.5.3** Derivatives of aldaric acids formed by modifying the carboxy group (into esters, amides, hydrazides, nitriles, amic acids, etc.) are named by the methods described in P-102.5.6.6.2.1, P-65 and P-66.

Examples:



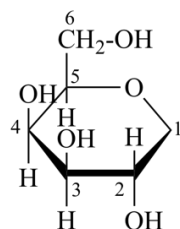
### P-102.5.6.7 Anhydrides

Anhydrides are intramolecular or intermolecular derivatives of monosaccharides.

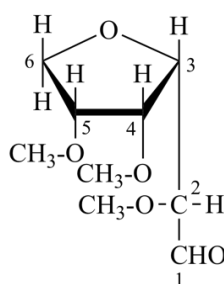
#### P-102.5.6.7.1 Intramolecular anhydrides

An intramolecular ether (commonly called an intramolecular anhydride), formally arising by elimination of water from two hydroxy groups of a single molecule of a monosaccharide (aldose, ketose) or monosaccharide derivative, is named by adding the detachable prefix 'anhydro', preceded by a pair of locants identifying the two hydroxy groups, to the name of the monosaccharide.

Examples:



1,5-anhydro-D-galactitol  
(2*R*,3*R*,4*R*,5*S*)-2-(hydroxymethyl)oxane-3,4,5-triol

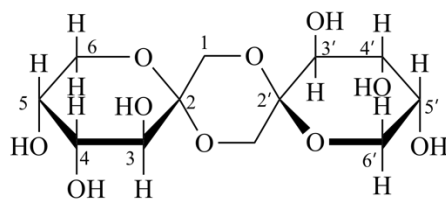


3,6-anhydro-2,4,5-tri-*O*-methyl-D-glucose  
(2*R*)-2-[(2*S*,3*R*,4*R*)-3,4-dimethoxyoxolan-2-yl]-2-methoxyacetaldehyde

#### P-102.5.6.7.2 Intermolecular anhydrides

The cyclic product of condensation of two monosaccharide molecules with elimination of two molecules of water (commonly called an intermolecular anhydride) is named by placing the term 'dianhydride' after the names of the two parent monosaccharides. When the two parents are different, the senior parent, according to the selection criteria for selecting the parent structure (see P-102.4), is cited first. The position of each anhydride link is indicated by a pair of locants showing the position of the two hydroxy groups involved, the locants relating to one monosaccharide (in a mixed anhydride, the second monosaccharide named) are primed. The pair of locants immediately precedes the term 'dianhydride'.

Example:



$\alpha$ -D-fructopyranose  $\beta$ -D-fructopyranose 1,2':1',2-dianhydride  
[ $\alpha$ -D-fructopyranose is cited first; according to P-102.4 (c),  $\alpha$  precedes  $\beta$ ]  
(3*R*,4*R*,5*S*,6*R*,9*S*,12*R*,13*R*,14*S*)-1,7,10,15-tetraoxadispiro[5.2.5<sup>9</sup>.2<sup>6</sup>]hexadecane-3,4,5,12,13,14-hexol

### P-102.6 MONOSACCHARIDES AND DERIVATIVES AS SUBSTITUENT GROUPS

#### P-102.6.1 Glycosyl groups

#### P-102.6.2 Substituent groups other than glycosyl groups

#### P-102.6.1 Glycosyl groups

##### P-102.6.1.1 Glycosyl groups

##### P-102.6.1.2 *O*-Glycosyl compounds

##### P-102.6.1.3 *N*-Glycosyl compounds (glycosylamines)

##### P-102.6.1.4 *C*-Glycosyl compounds

##### P-102.6.1.5 Glycosyl halides, pseudohalides and esters

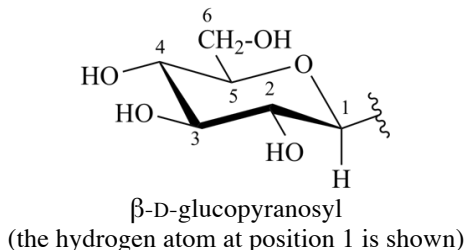


### P-102.6.1.1 Glycosyl groups

**P-102.6.1.1.1** Substituent groups formed by removal of the anomeric hydroxy group from a cyclic monosaccharide are named by replacing the final letter 'e' of the monosaccharide name by 'yl'. The term 'glycosyl residue' is used in the nomenclature of carbohydrates. Terms of this nature are widely used in naming glycosides, when they are not the parent structures, and oligosaccharides.

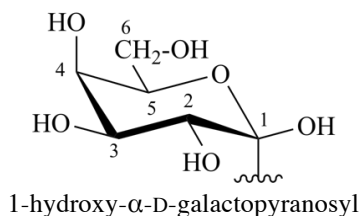
No locant is added to the name of the substituent to indicate the position of the free valence. A sinuous line denotes the free valence, as recommended for cyclic substituent groups in systematic nomenclature.

Examples:



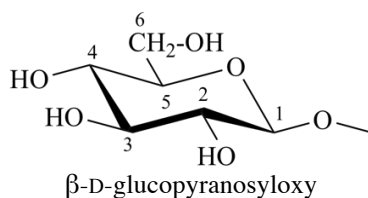
**P-102.6.1.1.2** When the free valence is formed at carbon '1' by subtraction of a hydrogen atom, the substituent group is named as a glycosyl group but the presence of the hydroxy group is denoted by substitution at carbon '1'. In this case, the stereodescriptor ' $\alpha$ ' or ' $\beta$ ' refers to the free valence, not to the -OH group.

Example:

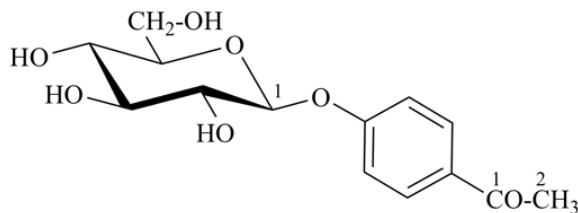


### P-102.6.1.2 O-Glycosyl compounds

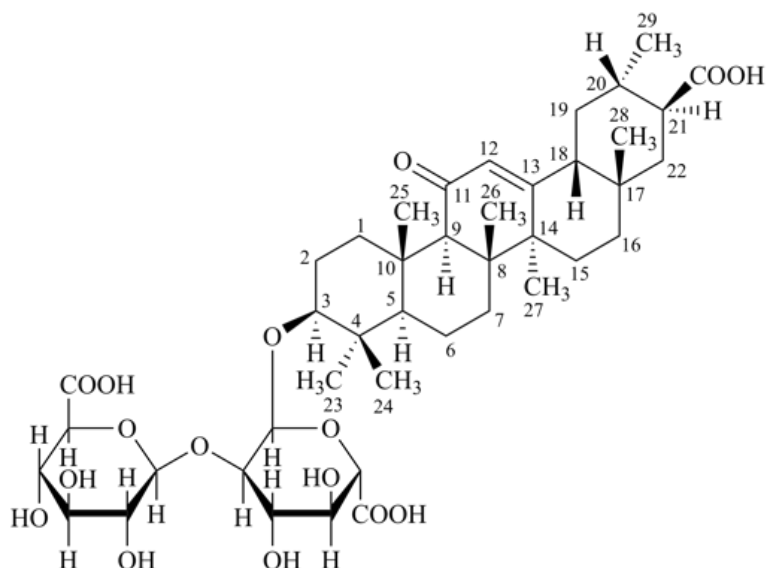
The substituent group formed by removal of a hydrogen atom from the anomeric -OH group is considered as a compound substituent group formed by the 'glycosyl' group and an 'oxy' group. In the examples, names are formed by using the seniority of class to determine the principal characteristic group to be assigned to the monosaccharide or to the aglycone component.



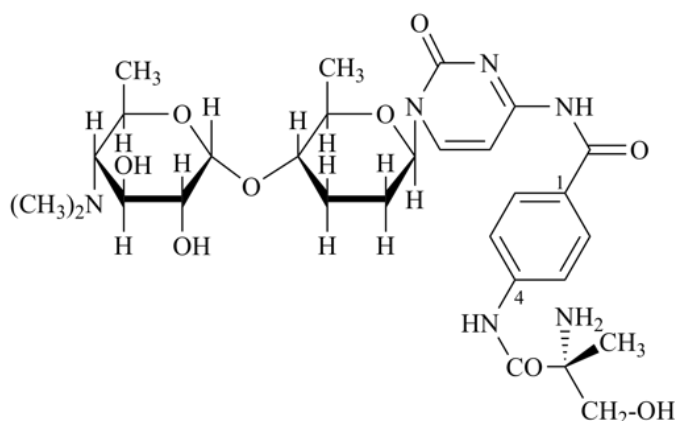
Examples:



1-[4-( $\beta$ -D-glucopyranosyloxy)phenyl]ethan-1-one  
[not 4'-( $\beta$ -D-glucopyranosyloxy)acetophenone;  
acetophenone cannot be substituted (see P-64.2.1.2)]  
(not 4-acetylphenyl  $\beta$ -D-glucopyranoside;  
a ketone is senior to a hydroxy compound)



21 $\beta$ -carboxy-11-oxo-30-norolean-12-en-3 $\beta$ -yl (2-O- $\beta$ -D-glucopyranosyluronic acid)- $\alpha$ -D-glucopyranosiduronic acid



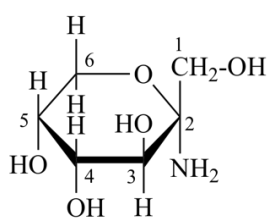
4-[(2*R*)-2-amino-3-hydroxy-2-methylpropanamido]-*N*-{1-[(2*R*,5*S*,6*R*)-5-{[4,6-dideoxy-4-(dimethylamino)- $\alpha$ -D-glucopyranosyl]oxy}-6-methyloxan-2-yl]-2-oxo-1,2-dihydropyrimidin-4-yl}benzamide

**Explanation:** the principal function is an amide; the cyclic amide, benzamide, is senior to the acyclic amide, propanamide.

#### P-102.6.1.3 *N*-Glycosyl compounds (glycosylamines)

*N*-Glycosyl derivatives are named as glycosylamines.

Example:

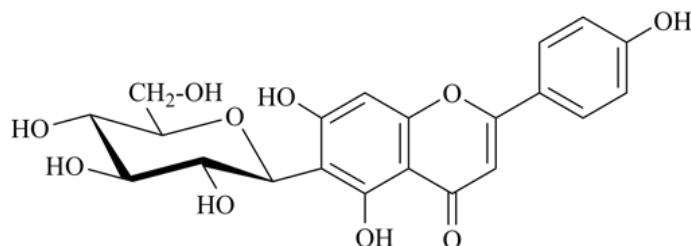


$\alpha$ -D-fructopyranosylamine

#### P-102.6.1.4 *C*-Glycosyl compounds

Compounds arising formally from the elimination of water from the glycosidic hydroxy group and a hydrogen atom bound to a carbon atom (thus creating a C-C bond) are named using the appropriate glycosyl group.

Example:

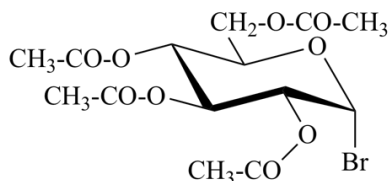


6-( $\beta$ -D-glucopyranosyl)-5,7-dihydroxy-2-(4-hydroxyphenyl)-4*H*-chromen-4-one  
6-( $\beta$ -D-glucopyranosyl)-5,7-dihydroxy-2-(4-hydroxyphenyl)-4*H*-1-benzopyran-4-one  
6-( $\beta$ -D-glucopyranosyl)-4',5,7-dihydroxyflavone

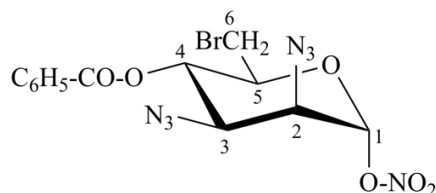
### P-102.6.1.5 Glycosyl halides, pseudohalides and esters

Glycosyl halides and pseudohalides are named by using functional class nomenclature, by adding, as a separate word, the class name 'chloride', 'isocyanate', etc. to the name of the appropriate glycosyl group. Esters of oxoacids in position 1 are treated as described for esters at other positions (see P-102.5.6.1).

Examples:



2,3,4,6-tetra-*O*-acetyl- $\alpha$ -D-glucopyranosyl bromide

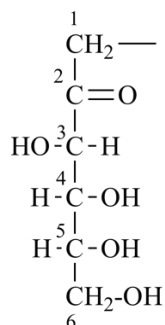


2,3-diazido-6-bromo-2,3,6-trideoxy- $\alpha$ -D-mannopyranose 4-benzoate 1-nitrate

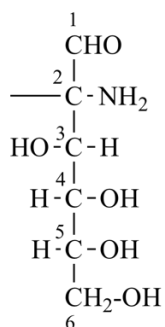
### P-102.6.2 Substituent groups other than glycosyl groups

A hydrogen atom may be removed from any position of a monosaccharide other than C-1. This formation of a free valence is denoted by the suffix 'yl', but a locant is necessary to indicate the position of the free valence and to distinguish such a name from that of glycosyl substituents for which the locant '1' is omitted. These prefixes can be formed by replacing the final letter 'e' of the systematic or trivial name of a monosaccharide by *n*-C-yl, *n*-O-yl. The symbol 'C' is omitted when the free valence is derived from a position at which hydrogen atoms only are attached.

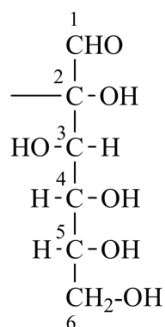
Examples:



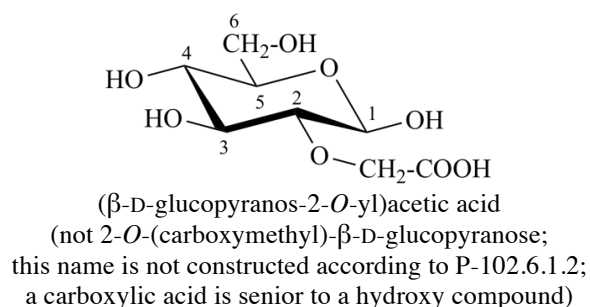
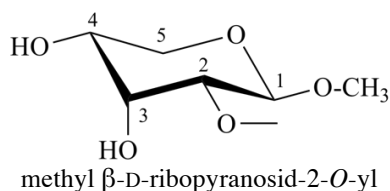
1-deoxy-D-fructos-1-yl



2-amino-2-deoxy-D-glucos-2-C-yl



D-glucos-2-C-yl



## P-102.7 DISACCHARIDES AND OLIGOSACCHARIDES

Names of disaccharides and oligosaccharides are formed by the principles, rules, and conventions described above for monosaccharides.

P-102.7.1 Disaccharides

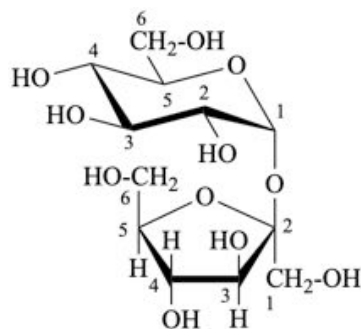
P-102.7.2 Oligosaccharides

### P-102.7.1 Disaccharides

#### P-102.7.1.1 Disaccharides without a free hemiacetal group

Disaccharides which can be regarded as formed by elimination of one molecule of water from two glycosidic (anomeric) hydroxy groups, are named as glycosyl glycosides. The parent (cited as the 'glycoside') is chosen in accordance with criteria described in P-102.4. Both anomeric descriptors must be cited in the name.

Example:



$\beta$ -D-fructofuranosyl  $\alpha$ -D-glucopyranoside  
(not  $\alpha$ -D-glucopyranosyl  $\beta$ -D-fructofuranoside  
fructo precedes gluco in the alphabetical order)  
sucrose (trivial name)

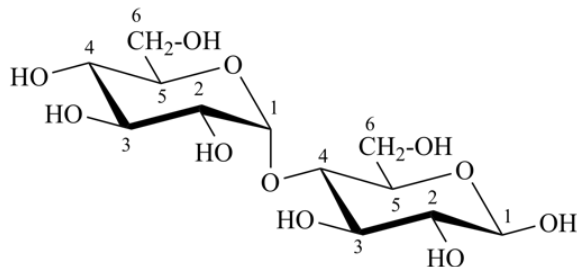
#### P-102.7.1.2 Disaccharides with a free hemiacetal group

Disaccharides which can be regarded as formed by elimination of one molecule of water from one glycosidic (anomeric) hydroxy group and one alcoholic hydroxy group, are named as glycosylglycoses. Locants and anomeric descriptors must be cited in the full name.

There are two established methods for citing locants:

- (1) in parentheses between the components with an arrow going from the locant of the glycosyl component to that of the glycoside component;
- (2) at the front of the glycosyl component.

Example:



- (1)  $\alpha$ -D-glucopyranosyl-(1 $\rightarrow$ 4)- $\beta$ -D-glucopyranose  
 (2) 4-O- $\alpha$ -D-glucopyranosyl- $\beta$ -D-glucopyranose  
 $\beta$ -maltose (trivial name;  
 not  $\beta$ -D-maltose)

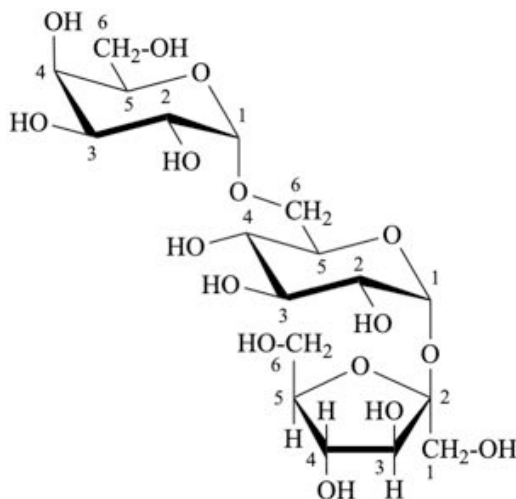
### P-102.7.2 Oligosaccharides

Oligosaccharides are multicomponent saccharides generally with more than two monosaccharide units. According to the number of units, they are called trisaccharides, tetrasaccharides, etc. The number of units involved before they become polysaccharides is not defined.

#### P-102.7.2.1 Oligosaccharides without a free hemiacetal group

A trisaccharide, for example, is named as a glycosylglycosyl glycoside or glycosyl glycosylglycoside as required. A choice between the two residues linked through their anomeric positions for citation as the 'glycoside' portion can be made on the basis of P-102.4. Alternatively, a sequential (end-to-end) naming approach may be used, regardless of P-102.4. The name is formed by the preferred method for naming disaccharides.

Example:

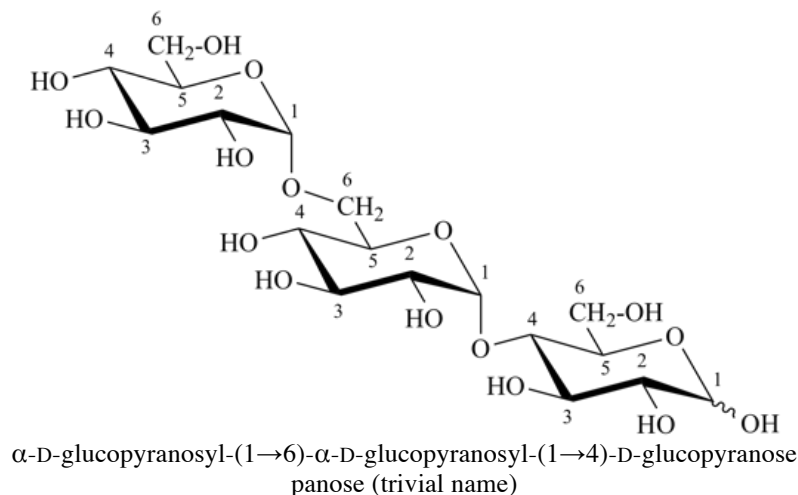


- $\beta$ -D-fructofuranosyl  $\alpha$ -D-galactopyranosyl-(1 $\rightarrow$ 6)- $\alpha$ -D-glucopyranoside  
 (glucose, not fructose, is selected as the 'glycoside')  
 $\alpha$ -D-galactopyranosyl-(1 $\rightarrow$ 6)- $\alpha$ -D-glucopyranosyl  $\beta$ -D-fructofuranoside (sequential method)  
 raffinose (trivial name)

#### P-102.7.2.2 Oligosaccharides with a free hemiacetal group

An oligosaccharide of this type is named as a glycosyl[glycosyl]<sub>n</sub>glucose, the 'glucose' portion being the parent. The conventional depiction has the 'glucose' portion on the right. Names are formed as described in P-102.7.2.1.

Example:



## P-103 AMINO ACIDS AND PEPTIDES

P-103.0 Introduction

P-103.1 Names, numbering, and configuration specification of amino acids

P-103.2 Derivatives of amino acids

P-103.3 Nomenclature of peptides

### P-103.0 INTRODUCTION

This Section describes the nomenclature of amino acids that constitute the building blocks of peptides and proteins. They are functional parents having retained names listed in Table 10.4. Less common amino acids also have retained names (see Table 10.5). The nomenclature of amino acids is composed of two types of names: names based on retained names for functional parents, with a limited capacity of functionalization and substitution, and systematic substitutive names for all other compounds.

The nomenclature of these amino acids and peptides is described in the document entitled 'Nomenclature and Symbolism for Amino Acids and Peptides' (ref. 18). A document covering the nomenclature of cyclic peptides is in preparation. In this Section, the nomenclature of these amino acids and peptides is restricted to their derivatives outside the field of peptides and proteins.

### P-103.1 NAMES, NUMBERING, AND CONFIGURATION SPECIFICATION OF AMINO ACIDS

P-103.1.1 Retained and systematic names

P-103.1.2 Numbering of  $\alpha$ -amino carboxylic acids

P-103.1.3 Configuration of  $\alpha$ -amino carboxylic acids

#### P-103.1.1 Retained and systematic names

P-103.1.1.1 Retained names of the 'common' amino acids

P-103.1.1.2 Retained names of 'less common' amino acids

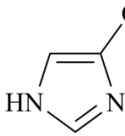
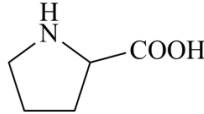
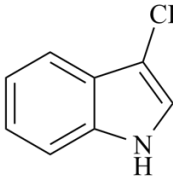
P-103.1.1.3 Systematic substitutive names

#### P-103.1.1.1 Retained names of the 'common' amino acids

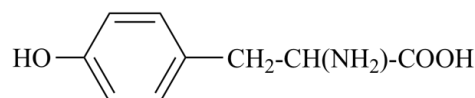
The retained names of the  $\alpha$ -amino acids that are commonly found in proteins and are represented in the genetic code, together with their systematic names, symbols (3-letter and/or 1-letter), and formulas, are given in Table 10.4. Some less common amino acids are discussed in P-103.1.1.2 and listed in Table 10.5.

**Table 10.4 Retained names of 'common'  $\alpha$ -amino acids**

Retained name	Symbols		Formula
	3-letter	1-letter	
alanine 2-aminopropanoic acid	Ala	A	$\text{CH}_3\text{-CH}(\text{NH}_2)\text{-COOH}$
arginine 2-amino-5-(carbamimidoylamino)pentanoic acid	Arg	R	$\text{H}_2\text{N-C(=NH)-NH-[CH}_2\text{]}_3\text{-CH}(\text{NH}_2)\text{-COOH}$

asparagine 2,4-diamino-4-oxobutanoic acid	Asn	N	$\text{H}_2\text{N}-\text{CO}-\text{CH}_2-\text{CH}(\text{NH}_2)-\text{COOH}$
aspartic acid aminobutanedioic acid	Asp	D	$\text{HOOC}-\text{CH}_2-\text{CH}(\text{NH}_2)-\text{COOH}$
cysteine 2-amino-3-sulfanylpropanoic acid	Cys	C	$\text{HS}-\text{CH}_2-\text{CH}(\text{NH}_2)-\text{COOH}$
glutamine 2,5-diamino-5-oxopentanoic acid	Gln	Q	$\text{H}_2\text{N}-\text{CO}-[\text{CH}_2]_2-\text{CH}(\text{NH}_2)-\text{COOH}$
glutamic acid 2-aminopentanedioic acid	Glu	E	$\text{HOOC}-[\text{CH}_2]_2-\text{CH}(\text{NH}_2)-\text{COOH}$
glycine aminoacetic acid	Gly	G	$\text{H}_2\text{N}-\text{CH}_2-\text{COOH}$
histidine 2-amino-3-(1 <i>H</i> -imidazol-4-yl)propanoic acid	His	H	$\text{CH}_2-\text{CH}(\text{NH}_2)-\text{COOH}$ 
isoleucine <i>rel</i> -(2 <i>R</i> ,3 <i>R</i> )-2-amino-3-methylpentanoic acid (see P-103.1.3.2.1 for configuration specification)	Ile	I	$\text{CH}_3-\text{CH}_2-\text{CH}(\text{CH}_3)-\text{CH}(\text{NH}_2)-\text{COOH}$
leucine 2-amino-4-methylpentanoic acid	Leu	L	$(\text{CH}_3)_2\text{CH}-\text{CH}_2-\text{CH}(\text{NH}_2)-\text{COOH}$
lysine 2,6-diaminohexanoic acid	Lys	K	$\text{H}_2\text{N}-[\text{CH}_2]_4-\text{CH}(\text{NH}_2)-\text{COOH}$
methionine 2-amino-4-(methylsulfanyl)butanoic acid	Met	M	$\text{CH}_3-\text{S}-[\text{CH}_2]_2-\text{CH}(\text{NH}_2)-\text{COOH}$
phenylalanine 2-amino-3-phenylpropanoic acid	Phe	F	$\text{C}_6\text{H}_5-\text{CH}_2-\text{CH}(\text{NH}_2)-\text{COOH}$
proline pyrrolidine-2-carboxylic acid	Pro	P	
serine 2-amino-3-hydroxypropanoic acid	Ser	S	$\text{HO}-\text{CH}_2-\text{CH}(\text{NH}_2)-\text{COOH}$
threonine <i>rel</i> -(2 <i>R</i> ,3 <i>S</i> )-2-amino-3-hydroxybutanoic acid (see P-103.1.3.2.1 for configuration specification)	Thr	T	$\text{CH}_3-\text{CH}(\text{OH})-\text{CH}(\text{NH}_2)-\text{COOH}$
tryptophan 2-amino-3-(1 <i>H</i> -indol-3-yl)propanoic acid	Trp	W	$\text{CH}_2-\text{CH}(\text{NH}_2)-\text{COOH}$ 

tyrosine	Tyr	Y
2-amino-3-(4-hydroxyphenyl)propanoic acid		
valine	Val	V
2-amino-3-methylbutanoic acid		
unspecified amino acid	Xaa	X



### P-103.1.1.2 Retained names of 'less common' amino acids

Several other less common trivial names and their symbols are described in Table 10.5. The publication 'Nomenclature and symbolism for amino acids and peptides' (ref. 18) must be consulted for the complete description of the naming of 'less common' amino acids.

**Table 10.5 Amino acids with trivial names (other than those listed in Table 10.4)**

Retained name Systematic name	Symbol	Formula
$\beta$ -alanine 3-aminopropanoic acid	$\beta$ Ala	$\text{H}_2\text{N}-\text{CH}_2-\text{CH}_2-\text{COOH}$
alloisoleucine <i>rel</i> -(2 <i>R</i> ,3 <i>S</i> )-2-amino-3-methylpentanoic acid (see P-103.1.3.2.1 for configuration specification)	alle	$\text{CH}_3-\text{CH}_2-\text{CH}(\text{CH}_3)-\text{CH}(\text{NH}_2)-\text{COOH}$
allothreonine <i>rel</i> -(2 <i>R</i> ,3 <i>R</i> )-2-amino-3-hydroxybutanoic acid (see P-103.1.3.2.1 for configuration specification)	aThr	$\text{CH}_3-\text{CH}(\text{OH})-\text{CH}(\text{NH}_2)-\text{COOH}$
allysine 2-amino-6-oxohexanoic acid	—	$\text{HCO}-[\text{CH}_2]_3-\text{CH}(\text{NH}_2)-\text{COOH}$
citrulline <i>N</i> <sup>5</sup> -carbamoylornithine	Cit	$\text{NH}_2-\text{CO}-\text{NH}-[\text{CH}_2]_3-\text{CH}(\text{NH}_2)-\text{COOH}$
cystathionine <i>S</i> -(2-amino-2-carboxyethyl)homocysteine	Ala   Hcy	$\text{HOOC}-\text{CH}(\text{NH}_2)-\text{CH}_2-\text{CH}_2-\text{S}-\text{CH}_2-\text{CH}(\text{NH}_2)-\text{COOH}$
cysteic acid 3-sulfoalanine 2-amino-3-sulfopropanoic acid	Cya	$\text{HO}_3\text{S}-\text{CH}_2-\text{CH}(\text{NH}_2)-\text{COOH}$
cystine 3,3'-disulfanediyldialanine	Cys   Cys	$\text{S}-\text{CH}_2-\text{CH}(\text{NH}_2)-\text{COOH}$   $\text{S}-\text{CH}_2-\text{CH}(\text{NH}_2)-\text{COOH}$
dopa 3-hydroxytyrosine	—	
homocysteine 2-amino-4-sulfanylbutanoic acid	Hcy	$\text{HS}-\text{CH}_2-\text{CH}_2-\text{CH}(\text{NH}_2)-\text{COOH}$

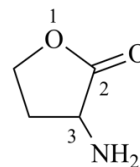


homoserine Hse HO-CH<sub>2</sub>-CH<sub>2</sub>-CH(NH<sub>2</sub>)-COOH

2-amino-4-hydroxybutanoic acid

homoserine lactone Hsl

3-aminooxolan-2-one



lanthionine

Ala

CH<sub>2</sub>-CH(NH<sub>2</sub>)-COOH

3,3'-sulfanediyldialanine

Cys

S-CH<sub>2</sub>-CH(NH<sub>2</sub>)-COOH

ornithine

Orn

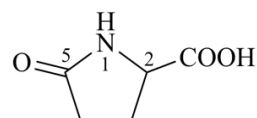
H<sub>2</sub>N-[CH<sub>2</sub>]<sub>3</sub>-CH(NH<sub>2</sub>)-COOH

2,5-diaminopentanoic acid

5-oxoproline

Glp

5-oxopyrrolidine-2-carboxylic acid



sarcosine

Sar

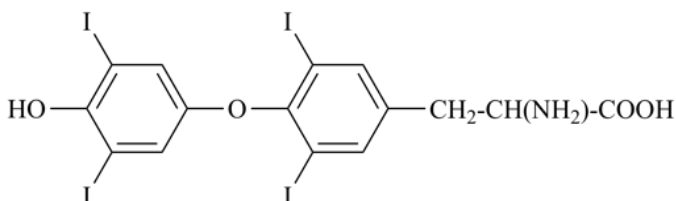
CH<sub>3</sub>-NH-CH<sub>2</sub>-COOH

*N*-methylglycine

thyroxine

Thx

*O*-(4-hydroxy-3,5-diiodophenyl)-3,5-diiodotyrosine

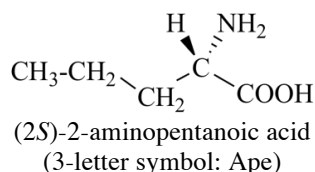


### P-103.1.1.3 Systematic substitutive names

When not denoted by a retained name, amino acids receive systematic substitutive names constructed by applying the principles, rules and conventions of substitutive nomenclature.

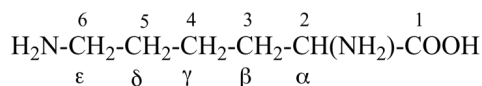
Systematic substitutive names are given to homologues of glycine and alanine, for example 2-aminobutanoic acid, 2-aminopentanoic acid (formerly 'norvaline') and 2-aminohexanoic acid (formerly 'norleucine'). The corresponding three-letter symbols are Abu, Ape, and Ahx. The stereodescriptors D and L are used to denote the configuration on C-2. These acids and their symbols are illustrated in 3AA-15.2.3 in ref. 18. The names 'norvaline' and 'norleucine' are not recommended (see 3AA-15.2.3, ref. 18).

Example:



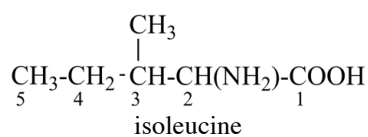
### P-103.1.2 Numbering of $\alpha$ -amino carboxylic acids

In acyclic amino acids, the carbon atom of the carboxy group next to the carbon atom carrying the amino group is numbered '1'. Alternatively, Greek letters may be used, with C-2 designated  $\alpha$ .

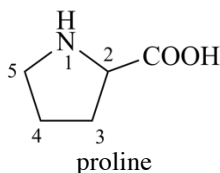


A heteroatom in a characteristic group has the same number as the carbon atom to which it is attached, e.g. *N*-2 is on C-2. When such numerals are used as locants they are written as superscripts, e.g. '*N*<sup>2</sup>'-acetyllysine (see P-103.2.1).

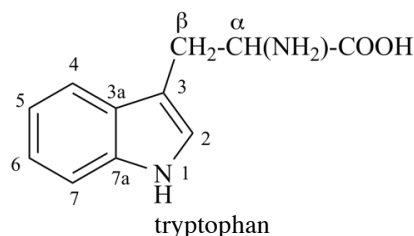
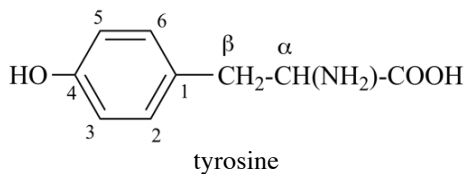
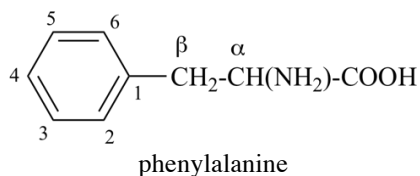
The carbon atoms of the methyl groups of valine are numbered '4' and '4''; likewise, those of leucine are '5' and '5''. Isoleucine is numbered as follows:



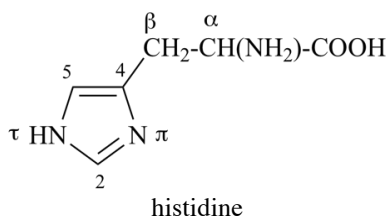
The atoms in proline are numbered as in pyrrolidine, the nitrogen atom being numbered '1', and the carbon atom bonded to the carboxy group is numbered '2'.



The carbon atoms in the aromatic rings of phenylalanine, tyrosine and tryptophan are numbered as in systematic nomenclature. The carbon atoms of the chains are designated ' $\alpha$ ' and ' $\beta$ ' as shown below:



A special fixed numbering, composed of numerals and Greek letters, is assigned to histidine; the Greek letters  $\pi$  and  $\tau$  are used to designate the nitrogen atoms of the ring near to and far from the side chain, respectively.

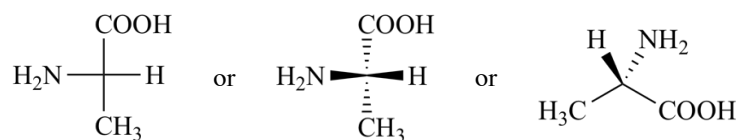


### P-103.1.3 Configuration of $\alpha$ -amino carboxylic acids

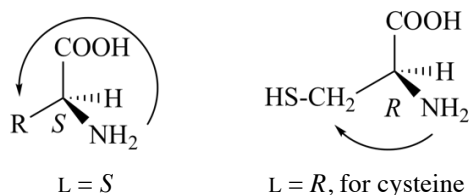
#### P-103.1.3.1 The stereodescriptors 'D' and 'L'

The absolute configuration at the  $\alpha$ -carbon atom of the  $\alpha$ -amino carboxylic acids is designated by the stereodescriptor 'D' or 'L' to indicate a formal relationship to 'D- or L-glyceraldehyde'. The stereodescriptor ' $\xi$ ' (Greek letter xi) indicates unknown configuration.

The structure of amino acids may be drawn to show configuration in several ways. A Fischer projection (see P-102.3.1) or a structural diagram including plain and wedged bonds (solid or hashed) may be used (see P-91.1), as drawn below for L-alanine:



The 'L' configuration corresponds to the 'S' configuration of the CIP system, except that cysteine has the 'R' configuration (and also cysteine, see P-103.1.1.2).

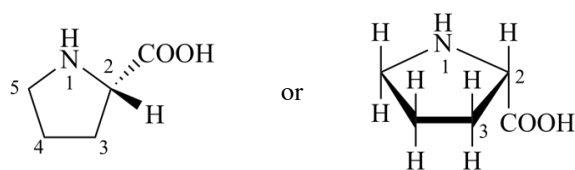


A mixture of equimolar amounts of 'D' and 'L' compounds is termed a 'racemate' and is designated by the stereodescriptor 'DL', for example 'DL-leucine'. The stereodescriptor 'DL' is preferred to 'rac', i.e. *rac*-leucine.

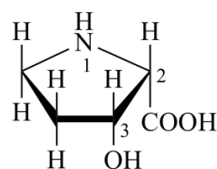
### P-103.1.3.2 Configuration of chirality centers other than the $\alpha$ -carbon atom

#### P-103.1.3.2.1 Use of CIP stereodescriptors

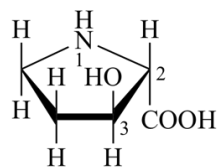
Stereodescriptors 'R' and 'S' are used for designating the configuration at centers other than the ' $\alpha$ -C' atom, while preserving stereodescriptors 'D' and 'L' for the ' $\alpha$ -C' atom to maintain homogeneity of stereodescriptors in peptides (see P-103.3.4). The use of stereodescriptors is exemplified here by the hydroxy prolines; for these compounds it is also acceptable to use the stereodescriptors *cis* and *trans* in the specialized nomenclature of amino acids and in general nomenclature.



L-proline  
(2*S*)-proline



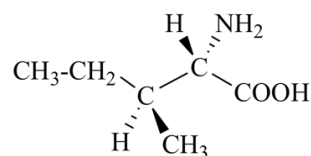
(3*R*)-3-hydroxy-L-proline  
(2*S*,3*R*)-3-hydroxyproline  
*cis*-3-hydroxy-L-proline



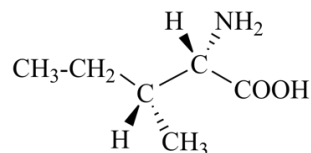
(3*S*)-3-hydroxy-L-proline  
(2*S*,3*S*)-3-hydroxyproline  
*trans*-3-hydroxy-L-proline

#### P-103.1.3.2.2 Use of the prefix 'allo'

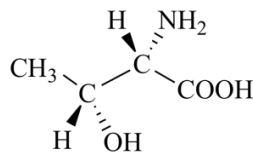
The prefix 'allo' is used to modify the retained names 'isoleucine' and 'threonine' when the configuration at 'C-3' has been inverted. The symbols are modified to 'alle' and 'aThr', respectively.



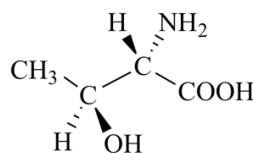
L-isoleucine (symbols 'Ile', 'I')  
(2*S*,3*S*)-2-amino-3-methylpentanoic acid



L-alloisoleucine (symbol 'alle')  
(2*S*,3*R*)-2-amino-3-methylpentanoic acid



L-threonine (symbols 'Thr', 'T')  
(2*S*,3*R*)-2-amino-3-hydroxybutanoic acid



L-allothreonine (symbol 'aThr')  
(2*S*,3*S*)-2-amino-3-hydroxybutanoic acid

## P-103.2 DERIVATIVES OF AMINO ACIDS

Retained names are used to generate names of salts, esters and acyl groups, and those of derivatives formed by substitution on carbon and nitrogen atoms or functionalization on oxygen and sulfur atoms.

The carboxy group,  $-\text{COOH}$ , can be transformed into various characteristic groups such as a hydroxymethyl group,  $-\text{CH}_2\text{-OH}$ , or an aldehyde group,  $-\text{CHO}$ . Some names derived from retained amino acid names are recommended to be used for naming amides, alcohols, aldehydes, and even ketones, in the context of peptide and protein nomenclature.

P-103.2.1 System for denoting locants

P-103.2.2 Names of substituent groups

P-103.2.3 Derivatives formed by substitution

P-103.2.4 Ionization of characteristic groups

P-103.2.5 Acyl groups

P-103.2.6 Esters

P-103.2.7 Amides, anilides, hydrazides, and other nitrogenous analogues

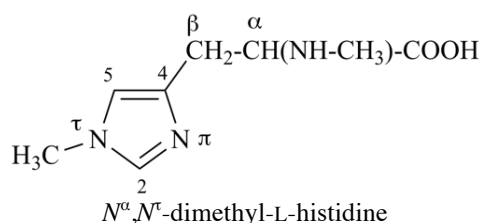
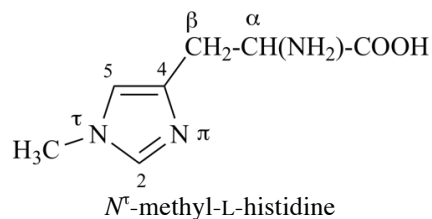
P-103.2.8 Alcohols, aldehydes, ketones and nitriles

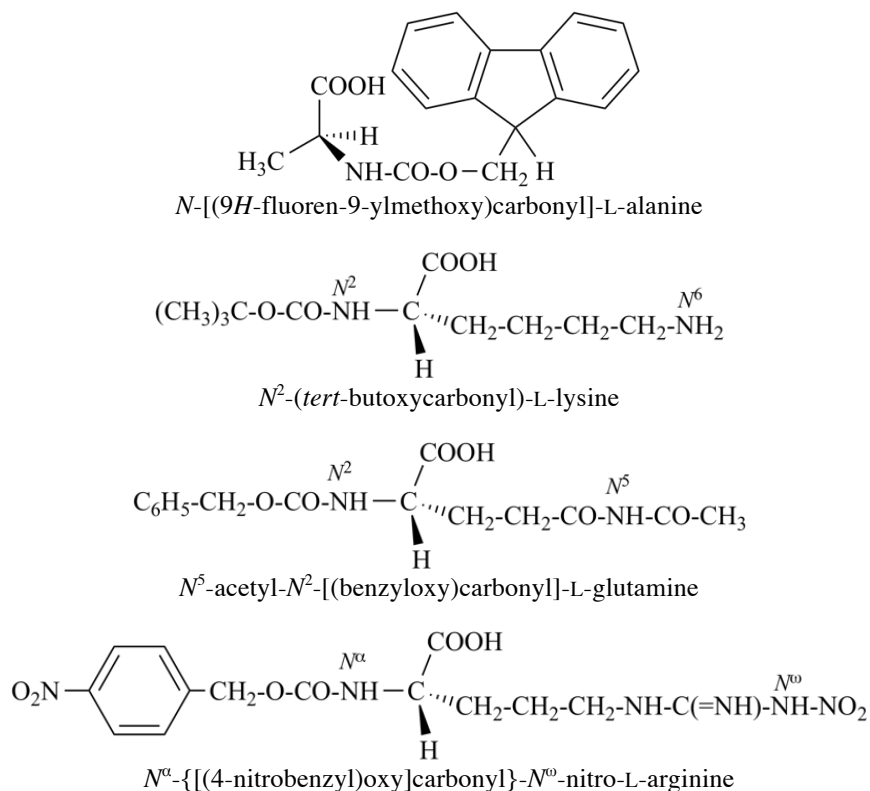
### P-103.2.1 System for denoting locants

It is recommended to use '*N*', '*O*' and '*S*' locants with the numerical locant of the attached carbon atom as a superscript to describe substitution on a nitrogen, oxygen, or sulfur atom when more than one is present. Locants ' $N^2$ ' and ' $N^6$ ' are recommended for lysine, ' $N^\alpha$ ', ' $N^\delta$ ' and ' $N^\omega$ ' for arginine, ' $N^2$ ' and ' $N^5$ ' for glutamine, ' $N^2$ ' and ' $N^4$ ' for asparagine, ' $N^\alpha$ ', ' $N^\pi$ ' and ' $N^\tau$ ' for histidine, according to the numbering of the corresponding  $\alpha$ -amino acids described in P-103.1.2. When only one nitrogen is present, the locant '*N*' is recommended; the numerical locant being omitted, even when other locants are present in the name.

When two identical substituent groups are present one letter locant is used between the multiplicative prefix and the name of the substituent group.

Examples:





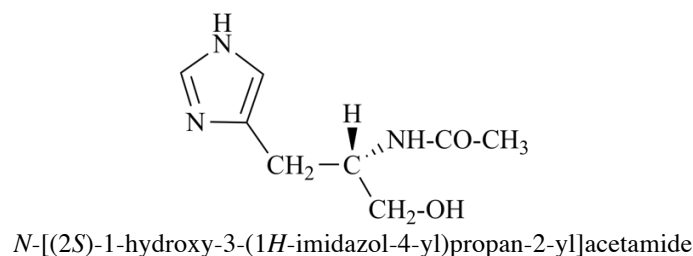
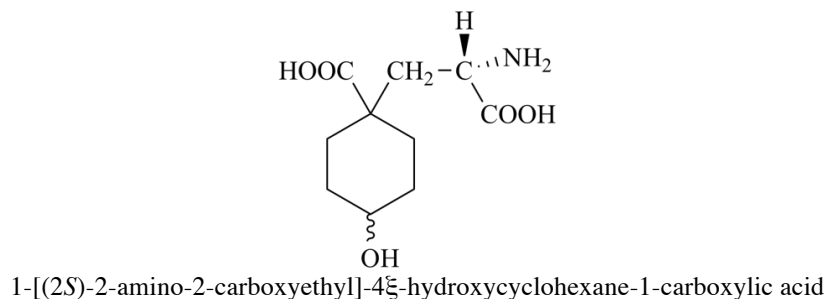
### P-103.2.2 Names of substituent groups

When  $\alpha$ -amino carboxylic acids must be cited as substituent groups in the presence of characteristic groups having seniority for citation as suffix, prefixes are formed according to the following principles.

#### P-103.2.2.1 Substituent groups with the free valence on a carbon atom

Substituent groups with the free valence on a carbon atom of an  $\alpha$ -amino carboxylic acid are formed according to the rules, principles and conventions of substitutive nomenclature as given in the previous chapters of these recommendations.

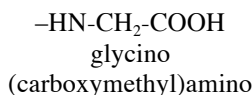
Examples:



#### P-103.2.2.2 Substituent groups with the free valence on a nitrogen atom

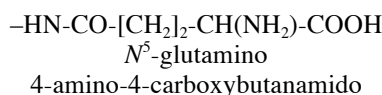
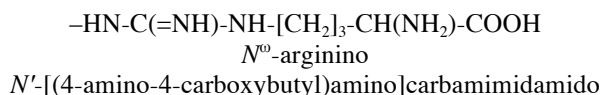
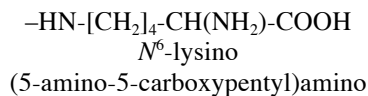
Substituent groups with the free valence on a nitrogen atom of an  $\alpha$ -amino carboxylic acid formed by the removal of a hydrogen atom from the amino group of an amino acid may be named by changing the ending 'e' of the name of the  $\alpha$ -amino acid into 'o', by adding the letter 'o' to tryptophan and by constructing the names asparto and glutamo, from aspartic acid and glutamic acid, respectively.

Example:



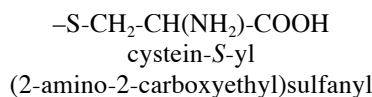
When there is more than one nitrogen atom in the amino acid, the use of a locant of the form '*N*' is recommended.

Examples:



**P-103.2.2.3** Substituent groups with the free valence on an oxygen or sulfur atom formed by subtraction of a hydrogen atom from an oxygen or sulfur atom may also be named by changing the final letter 'e', when appropriate, of the name of the  $\alpha$ -amino acid to '*x*-yl', *x* being the locant of the atom from which the hydrogen atom has been subtracted, for example cystein-*S*-yl, threonin-*O*-yl.

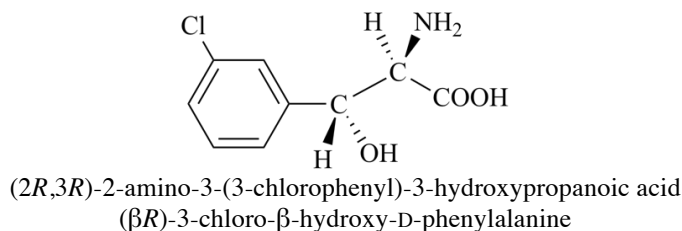
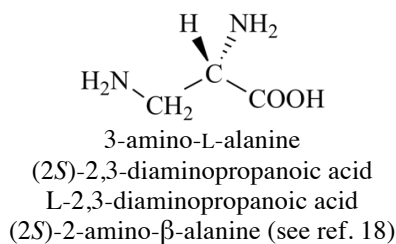
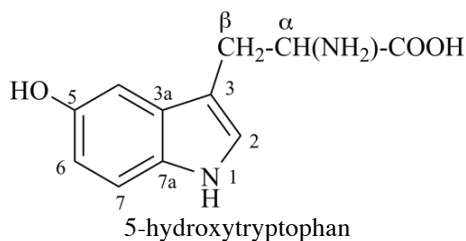
Example:

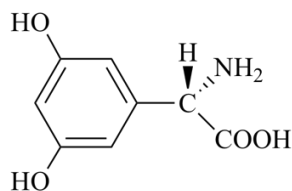


### P-103.2.3 Derivatives formed by substitution

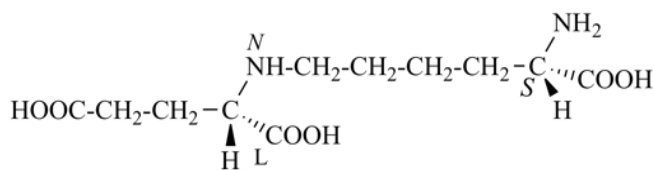
Retained names are used to indicate carbon, nitrogen, oxygen and sulfur substitution. Substitution on carbon atoms follows the principles, rules and conventions of substitutive nomenclature. Numerical locants and locants '*N*', '*O*', and '*S*' indicate the location of the substitution on nitrogen, oxygen, or sulfur atoms. For lysine, locants '*N*<sup>2</sup>' and '*N*<sup>6</sup>' are used to denote the two amino groups located at positions '2' and '6', respectively.

Examples:

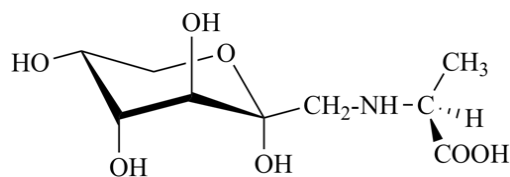




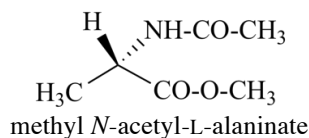
(2*S*)-2-amino-2-(3,5-dihydroxyphenyl)acetic acid  
2-(3,5-dihydroxyphenyl)-L-glycine



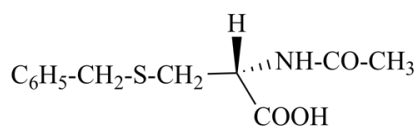
*N*-[(5*S*)-5-amino-5-carboxypentyl]-L-glutamic acid



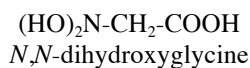
*N*-(1-deoxy- $\alpha$ -D-fructopyranos-1-yl)-L-alanine



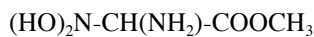
methyl *N*-acetyl-L-alaninate



*S*-benzyl-L-cysteine



*N,N*-dihydroxyglycine



*N*-(1-amino-2-methoxy-2-oxoethyl)azonous acid

[an acid, azonous acid for (HO)<sub>2</sub>NH, is senior to an ester; see P-67.1.1.1]

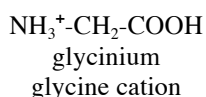
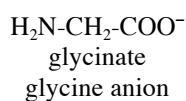
#### P-103.2.4 Ionization of characteristic groups

**P-103.2.4.1** The predominant form of a monoamino monocarboxylic acid in a neutral solution (pH 7) is R-CH(NH<sub>3</sub><sup>+</sup>)-COO<sup>-</sup> rather than R-CH(NH<sub>2</sub>)-COOH. It is nevertheless convenient to draw the conventional form as in Tables 10.4 and 10.5 and to name the amino acid alanine as 2-aminopropanoic acid rather than as 2-azaniumylpropanoate, 2-ammoniumylpropanoate or 2-ammoniopropanoate as in Chapter P-7 (P-74.1.3).

This is particularly so for representing the isoelectric forms of amino acids that contain other ionizing groups, such as a solution of lysine, which would contain appreciable amounts of both NH<sub>3</sub><sup>+</sup>-[CH<sub>2</sub>]<sub>4</sub>-CH(NH<sub>2</sub>)-COO<sup>-</sup> and NH<sub>2</sub>-[CH<sub>2</sub>]<sub>4</sub>-CH(NH<sub>3</sub><sup>+</sup>)-COO<sup>-</sup>.

**P-103.2.4.2** When it is desirable to mention or stress the ionic nature of an amino acid, cations or anions derived from a monoamino monocarboxylic acid may be indicated as follows (in indicating an anion the ending 'ate' replaces 'ic acid' or the final 'e' of the trivial name, or is added to the name tryptophan):

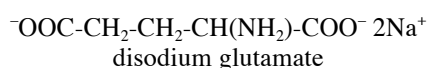
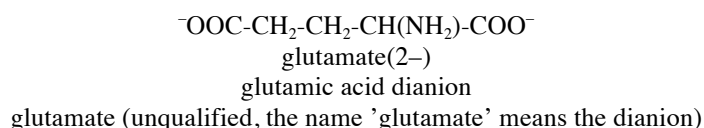
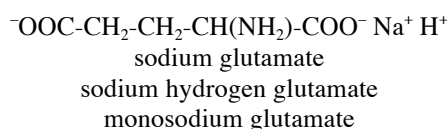
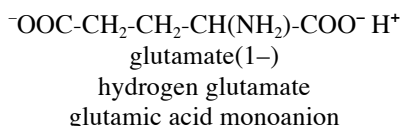
Examples:



**P-103.2.4.3** Further forms are required for amino acids that contain two amino groups or two carboxylic groups.

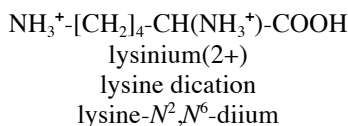
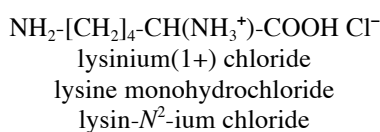
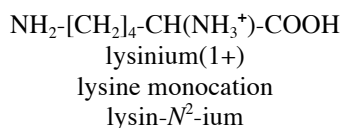
**P-103.2.4.3.1** The singly charged anions of aspartic and glutamic acids (strictly each has one positive charge and two negative charges, but this nomenclature refers to net charge) and may be distinguished from the doubly charged anions by placing the charge after the name or by stating the number of neutralizing cations.

Examples:



**P-103.2.4.3.2** The singly charged cations of arginine, histidine, and lysine (strictly each has two positive charges and one negative charge, but this nomenclature refers to net charge) may be distinguished from the doubly charged cations by placing the charge after the name, by the phrase 'monocation', or by stating the number of neutralizing anions. The location of the charge may be specified by a superscripted '*N*' locant placed before the 'ium' ending.

Examples:

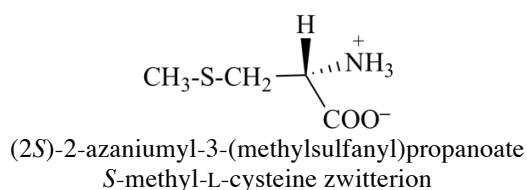
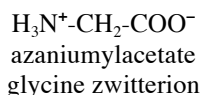


**P-103.2.4.4** Zwitterionic amino acids are named in two ways:

(1) by prefixing the name of the cationic substituent group into that of the anionic parent (see P-74.1.3) and using CIP stereodescriptors to describe the configuration;

(2) by adding the term 'zwitterion' to the name of the amino acid.

Examples:



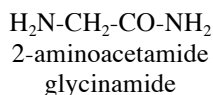
**P-103.2.5** Acyl groups

Acyl groups derived from organic acids, for example  $\text{H}_2\text{N-CHR-CO-}$ , are named by changing the ending 'ine' (or 'an' in tryptophan) into 'yl', for example alanyl, valyl, tryptophyl. 'Cysteinyl' is used instead of 'cysteyl'; 'cystyl' is derived from 'cystine' (see P-103.1.1.2).

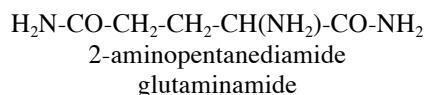
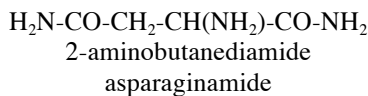




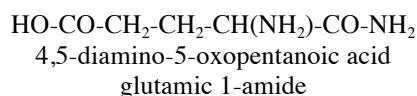
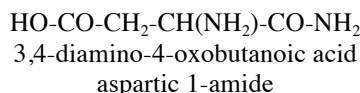
Examples:



The 4-amide of aspartic acid has its own name, asparagine, and the 5-amide of glutamic acid is glutamine (see Table 10.4). Their 1-amides are named as follows:

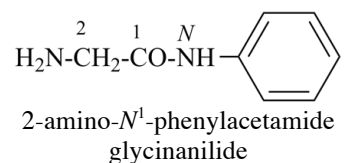


The 1-amides of aspartic acid and glutamic acid are named as follows:



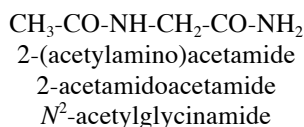
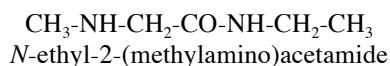
Names of anilides are formed by *N*-substitution of the amide group by the phenyl group or a substituted phenyl group. The ending ‘anilide’, in place of ‘amide’, may also be used.

Example:



Substitution on nitrogen atoms in amides of amino acids is expressed systematically by the methods described for amides (P-66.1.1.3) and amines (P-62.2.2.1).

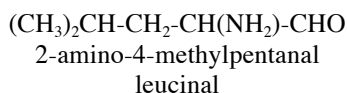
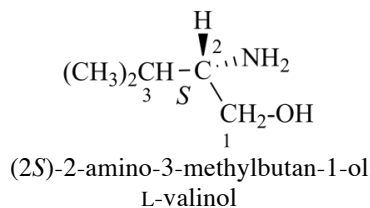
Examples:

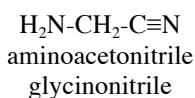
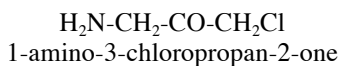


### P-103.2.8 Alcohols, aldehydes, ketones, and nitriles

Alcohols, aldehydes, ketones and nitriles corresponding to amino acids with retained trivial names are named systematically by using the principles, rules and conventions of substitutive nomenclature. The endings ‘ol’, ‘al’, ‘one’ and ‘onitrile’ are added to retained names, with elision of the final letter ‘e’ and may be used to designate a change in the characteristic group of an amino acid. Ketones must be named using systematic IUPAC substitutive nomenclature with ‘*R*’ and ‘*S*’ stereodescriptors when needed.

Examples:





### P-103.3 NOMENCLATURE OF PEPTIDES

Nomenclature of peptides is highly specialized and well documented in ref. 18. Nomenclature of cyclic peptides is under study by the IUPAC Chemical Nomenclature and Structure Representation Division.

#### P-103.3.1 Definitions

#### P-103.3.2 Names of peptides

#### P-103.3.3 Symbols of peptides

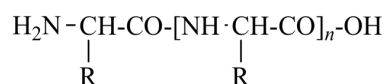
#### P-103.3.4 Indication of configuration in peptides

#### P-103.3.5 Modification of named peptides

#### P-103.3.6 Cyclic peptides

### P-103.3.1 Definitions

Peptides are amides derived from two or more amino carboxylic acid molecules (the same or different) by formation of a covalent bond from the carbonyl carbon of one to the nitrogen atom of another with formal loss of water. The term is usually applied to structures formed from  $\alpha$ -amino carboxylic acids, but it includes those derived from any amino carboxylic acid. In the following example, 'R' may be any organyl group, commonly but not necessarily one found in 'common' amino acids (see Table 10.4):



The amide bonds in peptides are called 'peptide bonds'. Peptides bonds formed between 'C-1' of one amino acid and 'N-2' of another amino acid are called 'eupeptide bonds'. Those formed between an amino group of one amino acid and a carboxy group of another amino acid, which are not 'eupeptide' bonds, are called 'isopeptide bonds'.

### P-103.3.2 Names of peptides

To name peptides, the names of acyl groups ending in 'yl' (see P-103.2.5) are used. Thus if the amino acids glycine,  $\text{H}_2\text{N}-\text{CH}_2-\text{COOH}$ , and alanine,  $\text{H}_2\text{N}-\text{CH}(\text{CH}_3)-\text{COOH}$ , condense so that glycine acylates alanine, the dipeptide formed,  $\text{H}_2\text{N}-\text{CH}_2-\text{CO}-\text{NH}-\text{CH}(\text{CH}_3)-\text{COOH}$ , is named glycylalanine. If they condense in the reverse order, the product  $\text{H}_2\text{N}-\text{CH}(\text{CH}_3)-\text{CO}-\text{NH}-\text{CH}_2-\text{COOH}$  is named alanylglycine. Higher peptides are named similarly, e.g. alanylleucyltryptophan.

### P-103.3.3 Symbols of peptides

The peptide glycylglycylglycine is symbolized as Gly-Gly-Gly. This involves modifying the symbols Gly for glycine,  $\text{H}_2\text{N}-\text{CH}_2-\text{COOH}$ , by adding hyphens to it, in three ways:

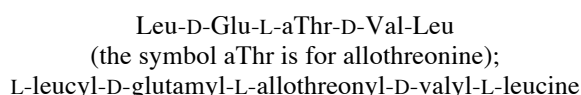
- (a) Gly- =  $\text{H}_2\text{N}-\text{CH}_2-\text{CO}-$
- (b) -Gly =  $-\text{HN}-\text{CH}_2-\text{COOH}$
- (c) -Gly- =  $-\text{HN}-\text{CH}_2-\text{CO}-$

Thus the hyphen, which represents the peptide bond, removes an -OH group from the -COOH group of the amino acid when written on the right of the symbol, and a hydrogen atom, when written on the left of the symbol.

### P-103.3.4 Indication of configuration in peptides

The stereodescriptor 'L' is not indicated in the names nor in the symbolic representation of peptides composed of amino acids listed in Table 10.4. In contrast, the stereodescriptor 'D' is indicated at the front of the acyl group or name of each component having that configuration.

Example:

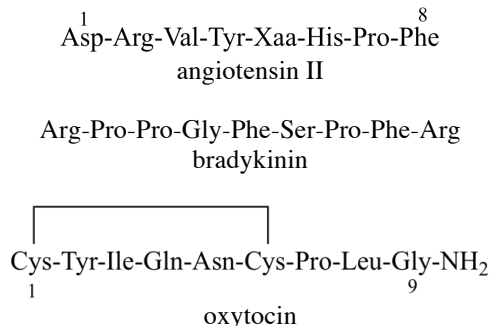


The symbol 'DL' is used to indicate a racemic mixture when one chirality center is present as indicated in P-103.1.3.1 but is not allowed in peptides, as its presence indicates the presence of diastereoisomers in unknown proportions. The italicized prefix *ambo* is used to indicate that both stereoisomers are present, for example, the result of the acylation of

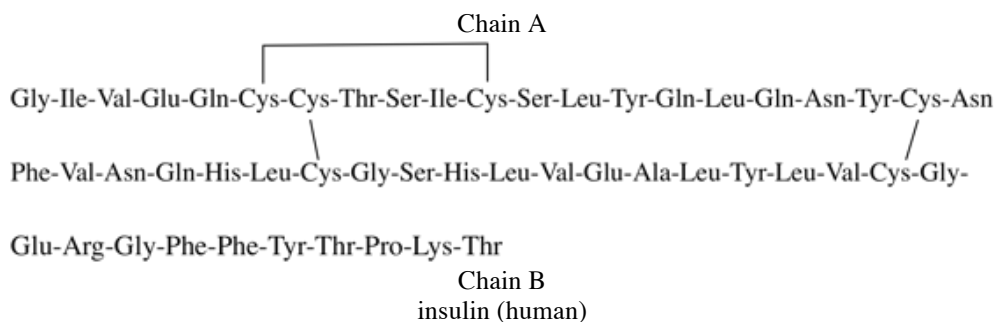
L-leucine by DL-alanine is *ambo*-alanylleucine or *ambo*-Ala-Leu. A residue of unknown configuration is indicated by the prefix  $\xi$  (Greek letter xi). The enantiomer of a named peptide is specified by the prefix *ent* (for *enantio*, see P-101.8.1), giving *ent*-bradykinin from bradykinin.

### P-103.3.5 Modification of named peptides

It is often convenient to specify the structure of a peptide by reference to a named sequence to which it is a variant. The recommendations that follow allow this, but they apply only to modification of the sequence involving normal amide links between residues. The retained names angiotensin II, bradykinin, oxytocin, and insulin (human) are used to illustrate these recommendations.



Specification of a sequence may require the species as well as the peptide to be named (see insulin below). If so, the name of the species must be attached, in parentheses, to the peptide whenever a modifying prefix is present.



P-103.3.5.1 Replacement of residues

P-103.3.5.2 Extension of the peptide chain

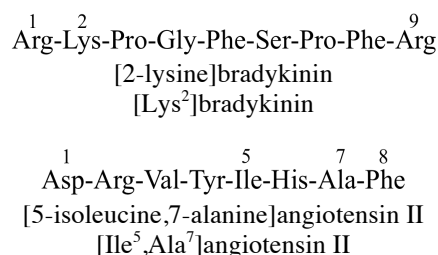
P-103.3.5.3 Insertion of residues

P-103.3.5.4 Removal of residues

#### P-103.3.5.1 Replacement of residues

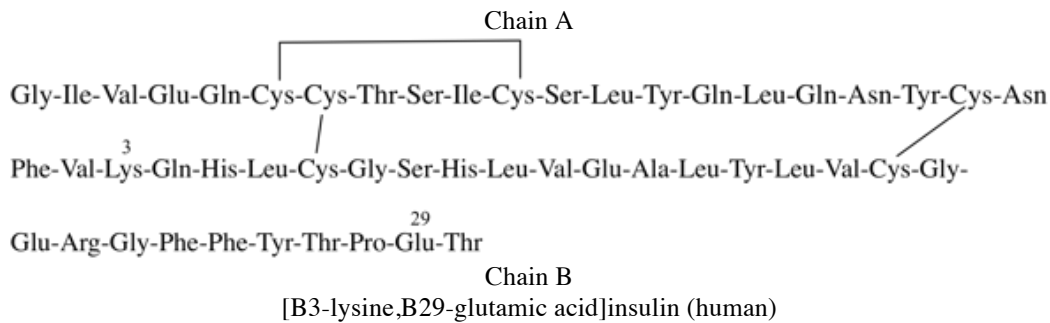
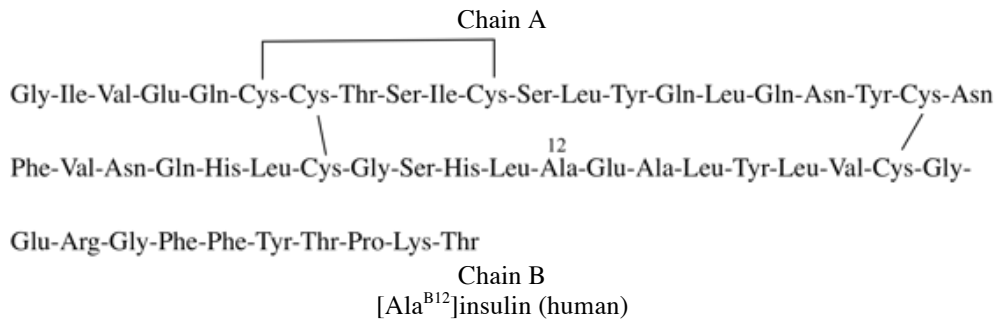
In a peptide of trivial name bradykinin, if the  $q$ th amino acid residue, starting from the 'N-terminal' end of the chain is replaced by the amino acid 'Xaa', the name of the modified peptide is [ $q$ -aminoacid]bradykinin and the abbreviated form is [Xaa $^q$ ]bradykinin. In a full name, the replacement amino acid is designated by its residue name, not the name of its acyl group (e.g. alanine, not alanyl). In the abbreviated form, the amino acid residues are designated by three letter symbols. In the abbreviated form, the position of replacement is indicated by a superscript.

Examples:



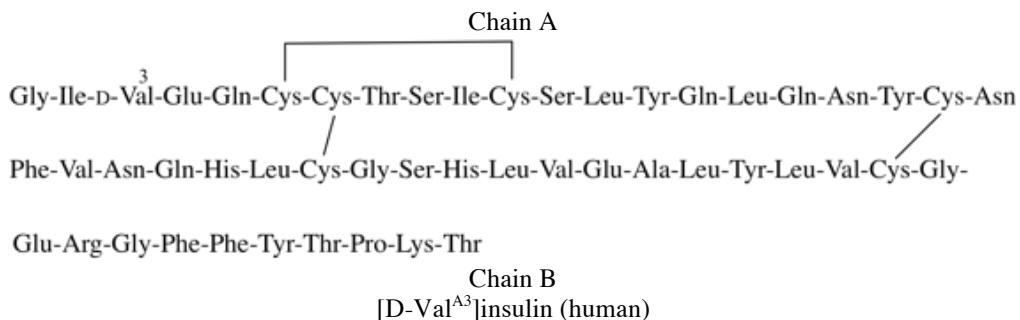
**P-103.3.5.1.1** The specification of a sequence requires the species as well as the peptide to be named. If so, the name of the species must be attached, in parentheses, to the name of the peptide whenever a modifying prefix is present.

Examples:



**P-103.3.5.1.2** The replacement of an amino acid residue by its enantiomer is expressed as follows. The replacement of L-proline in position 3 by D-proline results in [3-D-proline]bradykinin with the abbreviation [D-Pro<sup>3</sup>]bradykinin. A mixture of this with bradykinin gives [3-*ambo*-proline]bradykinin or [*ambo*-Pro<sup>3</sup>]bradykinin.

Example:

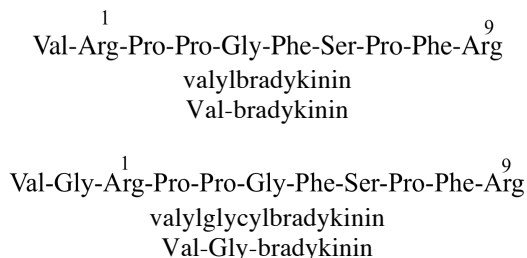


**P-103.3.5.2** Extension of the peptide chain

The compounds obtained by the extension of a peptide at either the N-terminus or the C-terminus are designated by the kind of names and abbreviations shown below.

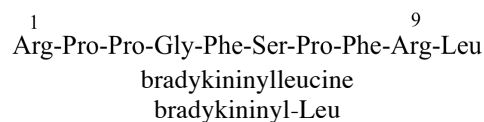
**P-103.3.5.2.1** Extension at the 'N-terminus'

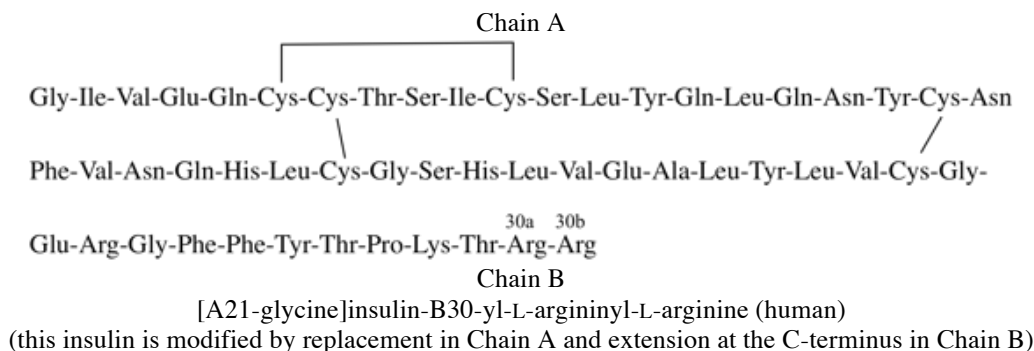
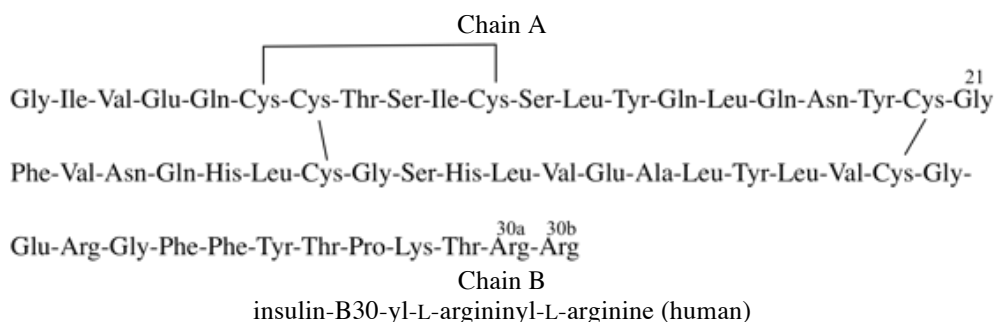
Examples:



**P-103.3.5.2.2** Extension at the 'C-terminus'

Examples:

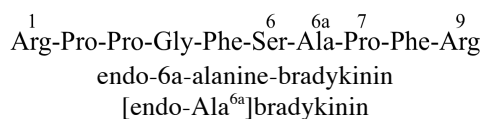




### P-103.3.5.3 Insertion of residues

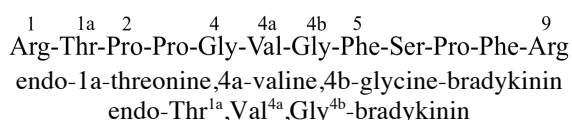
In peptide nomenclature, the prefix 'endo' (nonitalic) is used to denote the insertion of an amino acid residue in a well identified position in the peptide. For example, the name endo-6a-alanine-bradykinin, or [endo-Ala<sup>6a</sup>]bradykinin, means that the amino acid residue 'alanyl' has been inserted between the positions 6 and 7 in the structure of bradykinin. The prefix 'endo' is not to be confused with the recommended stereodescriptor 'endo' (written in italics) described in P-93.5.2.2.1.

Example:



Multiple insertions, and insertion of a maximum of two residues, together in the same place in the chain are shown by a logical extension of this recommendation. Thus the insertion into the peptide bradykinin of threonine between residues 1 and 2, and of valine and glycine (in that order) between residues 4 and 5 is shown by the name 'endo-1a-threonine,4a-valine,4b-glycine-bradykinin' or 'endo-Thr<sup>1a</sup>,Val<sup>4a</sup>,Gly<sup>4b</sup>-bradykinin'.

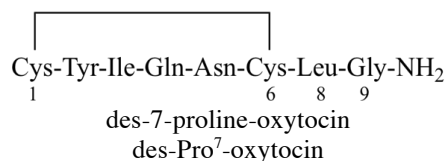
Example:

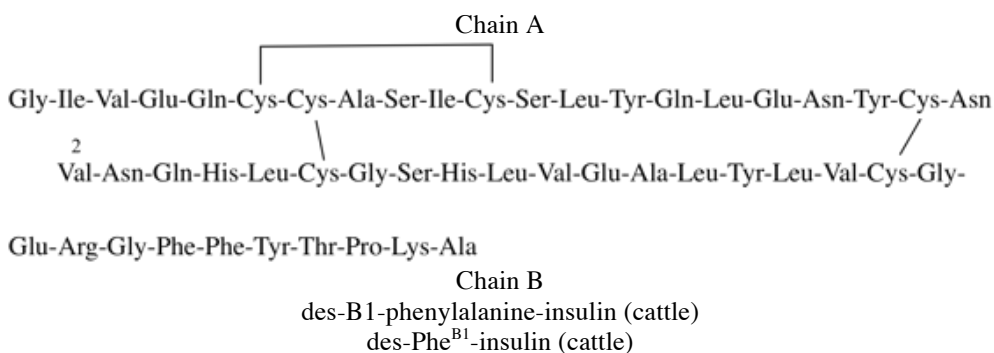


### P-103.3.5.4 Removal of residues

The subtractive prefix 'des', in peptide nomenclature, is used to denote the removal of an amino acid residue from any position in a peptide structure. For example, the name des-8-phenylalanine-bradykinin means that the amino-acid residue 'phenylalanyl', located in position 8 of the peptide bradykinin, has been removed; or in the abbreviated form des-Phe<sup>8</sup>-bradykinin. In the modification of parent structures described in Section P-101, the prefix 'des' is used to indicate the removal of a terminal ring in steroids with the addition of the appropriate number of hydrogen atoms at each junction with the adjacent ring (see P-101.3.6).

Examples:





### P-103.3.6 Cyclic peptides

Cyclic peptides include rings generated from a peptide (acyclic peptide) by formation of a peptide or ester bond, by a disulfide link, or by a new carbon-carbon, carbon-nitrogen, carbon-oxygen, or carbon-sulfur bond (excluding esters and amides). Cyclic peptides in which the ring consists entirely of amino acid residues with eupeptide bonds are called 'homodetic cyclic peptides'. Those formed of eupeptide and isopeptide bonds are called 'heterodetic cyclic peptides'. This aspect of peptide nomenclature is under study by the IUPAC Chemical Nomenclature and Structure Representation Division.

### P-104 CYCLITOLS

- P-104.0 Introduction
- P-104.1 Definitions
- P-104.2 Name construction
- P-104.3 Derivatives of cyclitols

#### P-104.0 INTRODUCTION

The nomenclature of cyclitols is described in the document entitled 'Nomenclature of Cyclitols, Recommendations 1973' (ref. 39).

#### P-104.1 DEFINITIONS

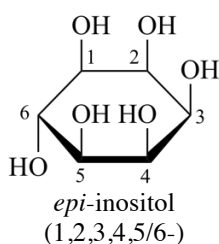
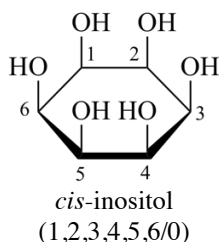
Cyclitols are cycloalkanes in which three or more ring atoms are each substituted with one hydroxy group. Inositols, cyclohexane-1,2,3,4,5,6-hexols, are a specific group of cyclitols.

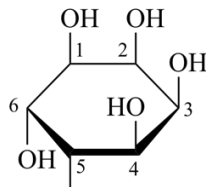
Inositols have retained names and, together with their *O*-alkyl, *O*-aryl, alkanooate/carboxylate esters, and amino derivatives (where NH<sub>2</sub> replaces OH) employ the stereodescriptors 'D' and 'L' to describe configurations. Other names of cyclitols are systematic substitutive names whose configurations are described by CIP stereodescriptors.

#### P-104.2 NAME CONSTRUCTION

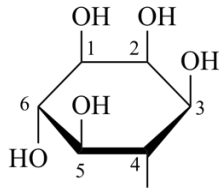
Various methods are recommended for naming cyclitols.

**P-104.2.1** Stereoisomeric inositols are described by adding italicized prefixes at the front of the name 'inositol'. Positional numbers described in the following method are indicated in parentheses. Names denoted by the prefixes are preferred.

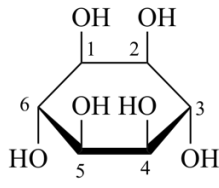




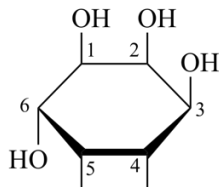
*allo*-inositol  
(1,2,3,4/5,6-)



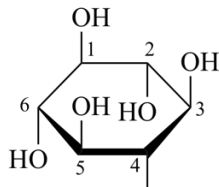
*myo*-inositol  
(1,2,3,5/4,6-)



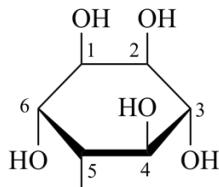
*muco*-inositol  
(1,2,4,5/3,6-)



*neo*-inositol  
(1,2,3/4,5,6-)

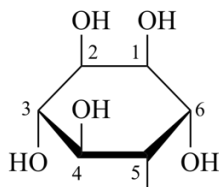


*scyllo*-inositol  
(1,3,5/2,4,6-)



1*L*-*chiro*-inositol  
(1,2,4/3,5,6-)

[formerly *L*-*chiro*-inositol or  
(-)-inositol]



1*D*-*chiro*-inositol  
(1,2,4/3,5,6-)

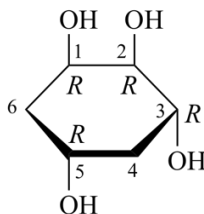
[formerly *D*-*chiro*-inositol or (+)-inositol]



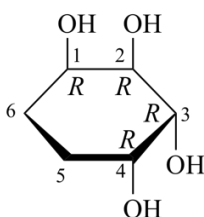
The absolute configuration of cyclitols is denoted by 'D' and 'L' and is determined in the following way. For the planar ring representation where the hydroxy group numbered '1' is above the plane of the ring, the configuration 'L' corresponds to a clockwise numbering, and the configuration 'D' corresponds to an anticlockwise numbering, as illustrated by the two enantiomeric *chiro*-inositols above. The stereodescriptors 'D' and 'L' followed by a hyphen are placed before the name of the compound and are preceded by the locant of the defining center, i.e. '1' as shown above.

**P-104.2.2** Cyclitols, with the exception of inositols, are named systematically from cyclohexane as the parent using the CIP method and its Sequence Rules for describing stereoisomers. This method is preferred to the method of positional numbers described in P-104.2.3.

Examples:



(1*R*,2*R*,3*R*,5*R*)-cyclohexane-1,2,3,5-tetrol



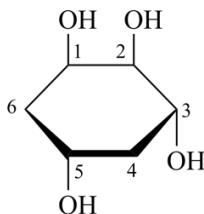
(1*R*,2*R*,3*R*,4*R*)-cyclohexane-1,2,3,4-tetrol

**P-104.2.3** Locants are assigned to hydroxy groups in cyclitols, and thus the direction of numbering is described, with reference to the steric relations and nature of the substituents attached to the ring. The substituents lying above the plane of the ring constitute a set, and those lying below another set. Lowest locants are related to one set of the substituents according to the following criteria, which are applied successively until a decision is reached:

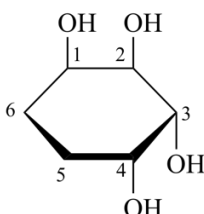
- to the substituents considered as a numerical series, without regard to configuration;
- if one set of the substituents is more numerous than the other, to the more numerous;
- if the sets are equally numerous and one of them can be denoted by lower numbers, to that set;
- to substituents other than unmodified hydroxy groups;
- to the substituent first cited in alphanumerical order;
- to those designations that lead to an 'L' rather than a 'D' configuration, as determined by the method of P-104.2.1 above (applies to *meso* compounds only).

The positional numbers are described by a fractional expression in which the numerator is the set of substituents with the lowest locants, arranged in ascending order, and the denominator is the other set.

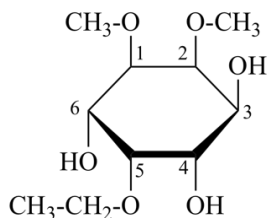
Examples:



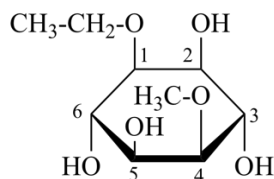
1L-1,2/3,5-cyclohexanetetrol [criteria (a) and (c)]  
(1*R*,2*R*,3*R*,5*R*)-cyclohexane-1,2,3,5-tetrol



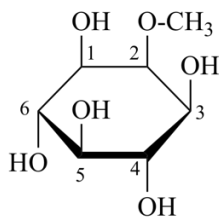
1L-1,2/3,4-cyclohexanetetrol [criterion (a)]  
(1*R*,2*R*,3*R*,4*R*)-cyclohexane-1,2,3,4-tetrol



1L-5-*O*-ethyl-1,2-di-*O*-methyl-*neo*-inositol [criterion (d)]



1L-1-*O*-ethyl-4-*O*-methyl-*muco*-inositol [criteria (d) and (e)]



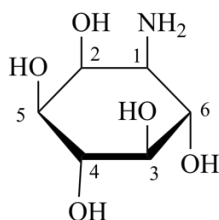
2-*O*-methyl-*myo*-inositol [criteria (b) and (f)]

### P-104.3 DERIVATIVES OF CYCLITOLS

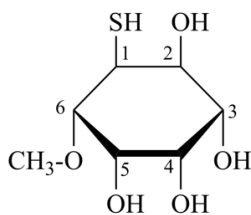
#### P-104.3.1 Derivatives of inositols

Inositols are modified in the same way as carbohydrates to generate names for derivatives. There is no limit to *O*-substitution by alkyl (aryl) groups (see P-102.5.6.1). Hydroxy groups can be exchanged for amino groups using the ‘deoxy’ operation (see P-102.5.4). When characteristic groups that are senior to hydroxy groups are put in the place of a hydroxy group, fully substitutive names must be constructed. However, esters are named as alkanooates/carboxylates. The numbering of the inositol remains unchanged and the configuration is expressed by an ‘L’ or ‘D’ stereodescriptor. Systematic names may have different numbering from the corresponding inositol name (see examples 2 and 5).

Examples:



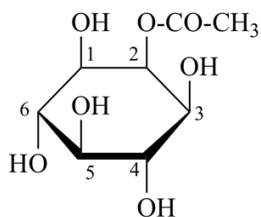
1D-1-amino-1-deoxy-*myo*-inositol



1L-1-deoxy-6-*O*-methyl-1-sulfanyl-*allo*-inositol

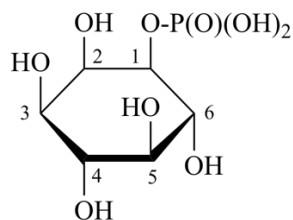
(not 1L-6-*O*-methyl-1-thio-*allo*-inositol)

(1*S*,2*R*,3*S*,4*S*,5*S*,6*S*)-5-methoxy-6-sulfanylcyclohexane-1,2,3,4-tetrol

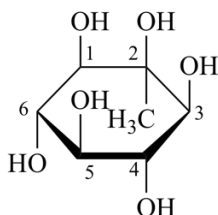


*myo*-inositol 2-acetate

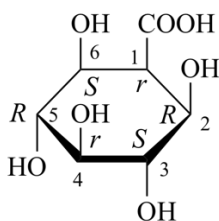
(see P-102.5.6.1.1 for naming carbohydrate esters)



1D-*myo*-inositol 1-(dihydrogen phosphate)



2-*C*-methyl-*myo*-inositol  
(1*s*,2*R*,3*S*,4*s*,5*R*,6*S*)-1-methylcyclohexane-1,2,3,4,5,6-hexol

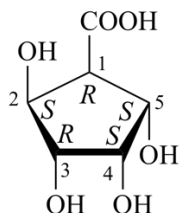


(1*r*,2*R*,3*S*,4*r*,5*R*,6*S*)-2,3,4,5,6-pentahydroxycyclohexane-1-carboxylic acid  
(not 2-carboxy-2-deoxy-*myo*-inositol;  
COOH is senior to OH)

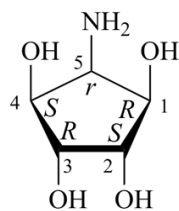
### P-104.3.2 Derivatives of cyclitols other than inositols

Names for derivatives of cyclitols other than inositols are all constructed by applying the principles, rules, and conventions of substitutive nomenclature described in Chapters P-1 through P-9.

Examples:



(1*R*,2*S*,3*R*,4*S*,5*S*)-2,3,4,5-tetrahydroxycyclopentane-1-carboxylic acid



(1*R*,2*S*,3*R*,4*S*,5*r*)-5-aminocyclopentane-1,2,3,4-tetrol

## P-105 NUCLEOSIDES

P-105.0 Introduction

P-105.1 Retained names of nucleosides

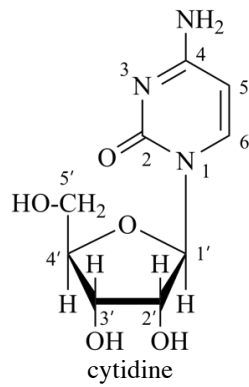
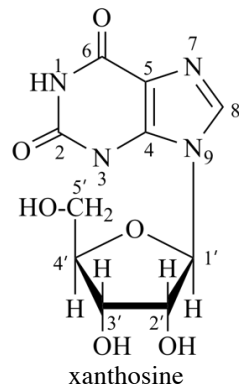
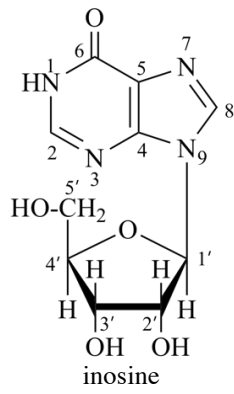
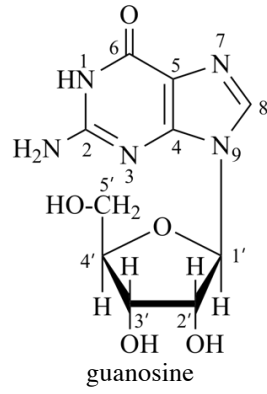
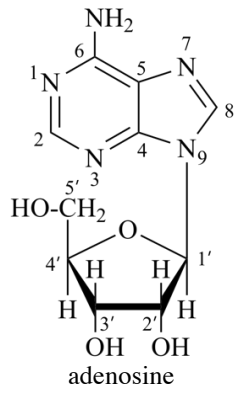
P-105.2 Substitution on nucleosides

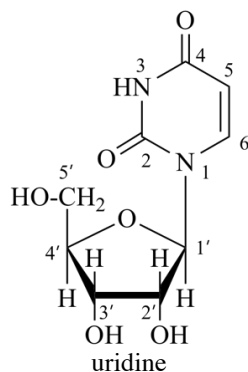
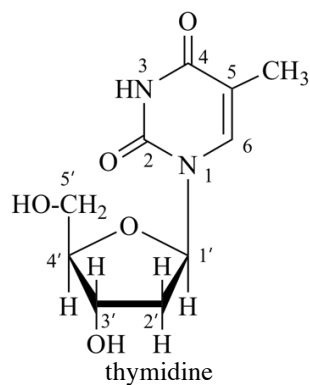
### P-105.0 INTRODUCTION

The nomenclature of nucleosides is exemplified in the document entitled 'Abbreviations and Symbols for Nucleic Acids, Polynucleotides and their Constituents' (ref. 47). The procedures for naming derivatives given in this Section are adapted from Section P-102 for modifications on the carbohydrate moiety and from the general rules of substitution of organic compounds.

### P-105.1 RETAINED NAMES OF NUCLEOSIDES

The following names are retained.

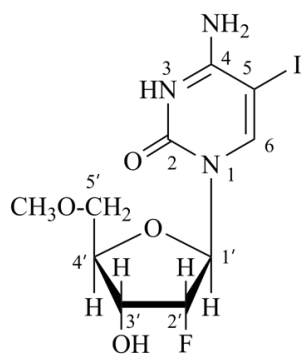
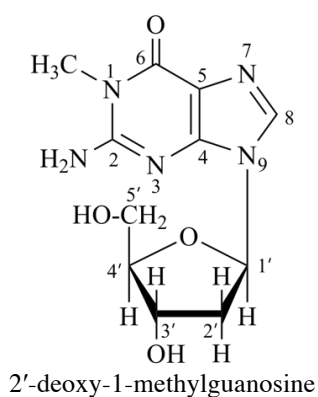




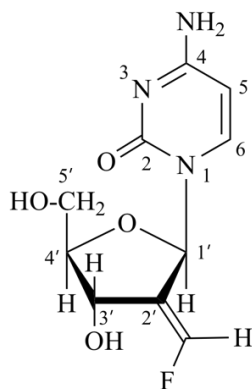
## P-105.2 SUBSTITUTION ON NUCLEOSIDES

**P-105.2.1** Nucleosides having retained names can be fully substituted on the purine or pyrimidine ring. Replacement of oxo groups of nucleosides is described by functional replacement prefixes. The ribofuranosyl component may be modified as prescribed for carbohydrates (see P-102.5).

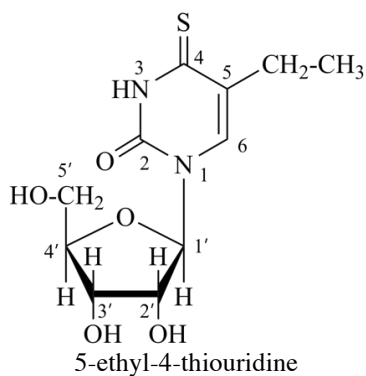
Examples:



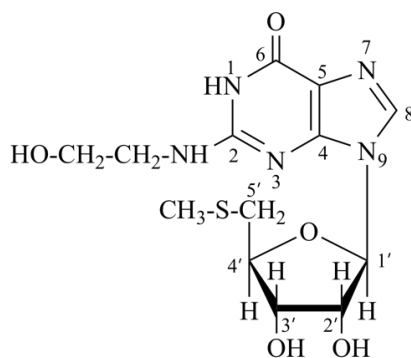
(the replacement of a hydroxy group by a fluorine atom at the same position is allowed)  
 4-amino-1-[(2*R*,3*R*,4*R*,5*R*)-3-fluoro-4-hydroxy-5-(methoxymethyl)oxolan-2-yl]-5-iodopyrimidin-2(1*H*)-one



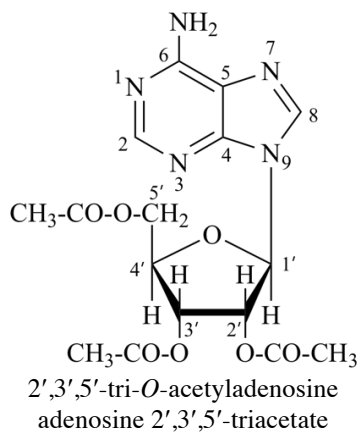
(2'*E*)-2'-deoxy-2'-(fluoromethylidene)cytidine  
4-amino-1-[(2*R*,3*E*,4*S*,5*R*)-3-(fluoromethylidene)-4-hydroxy-5-(hydroxymethyl)oxolan-2-yl]pyrimidin-2(1*H*)-one



5-ethyl-4-thiouridine



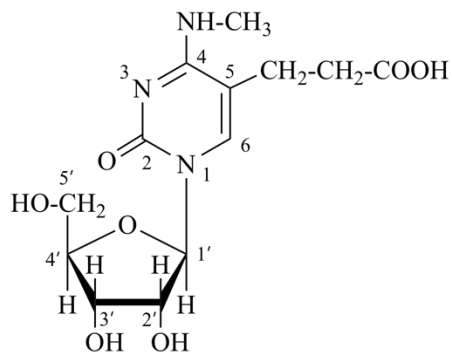
*N*<sup>2</sup>-(2-hydroxyethyl)-5'-*S*-methyl-5'-thioguanosine



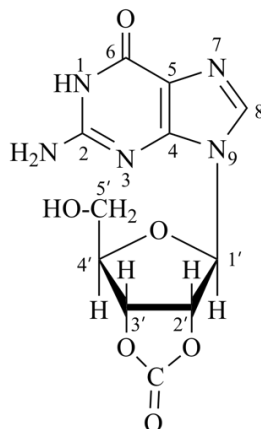
2',3',5'-tri-*O*-acetyladenosine  
adenosine 2',3',5'-triacetate

**P-105.2.2** In the presence of a characteristic group of higher priority than pseudoketone, normal substitutive nomenclature principles are applied.

Examples:



3-[4-(methylamino)-2-oxo-1-β-D-ribofuranosyl]-1,2-dihydropyrimidin-5-yl]propanoic acid



2',3'-dideoxyguanosine-2',3'-diyl carbonate (see P-101.7.4)  
guanosine cyclic-2',3'-carbonate

## P-106 NUCLEOTIDES

P-106.0 Introduction

P-106.1 Retained names

P-106.2 Nucleoside diphosphates and triphosphates

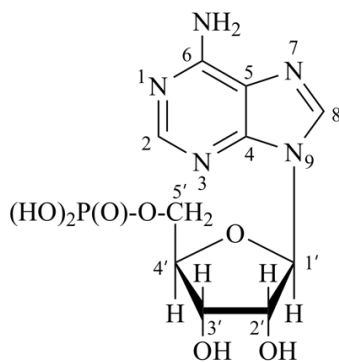
P-106.3 Derivatives of nucleotides

### P-106.0 INTRODUCTION

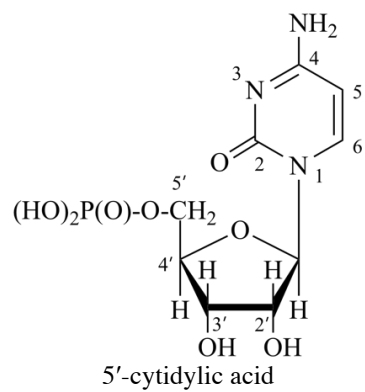
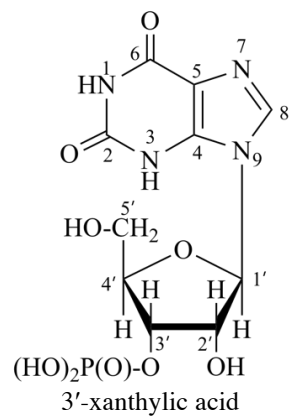
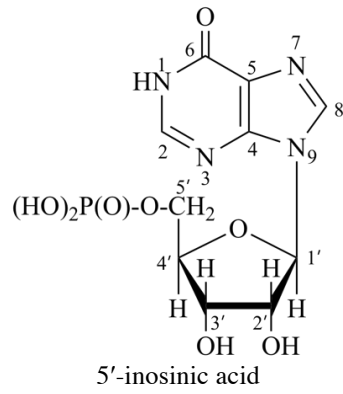
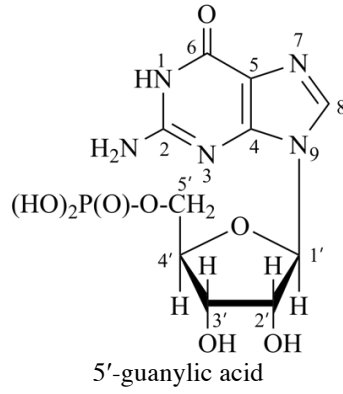
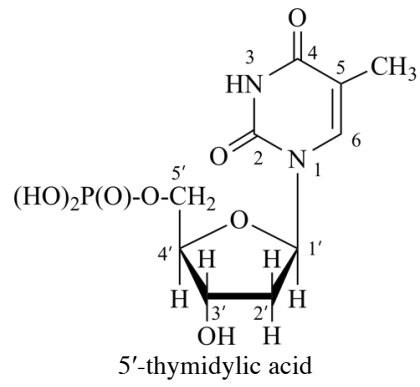
Names of nucleotides are exemplified in the document entitled 'Abbreviations and Symbols for Nucleic Acids, Polynucleotides and their Constituents' (ref. 47). The procedures for naming of derivatives given in this section are adapted from Section P-102 for modifications on the carbohydrate moiety and from the general rules for substitution of organic compounds.

### P-106.1 RETAINED NAMES

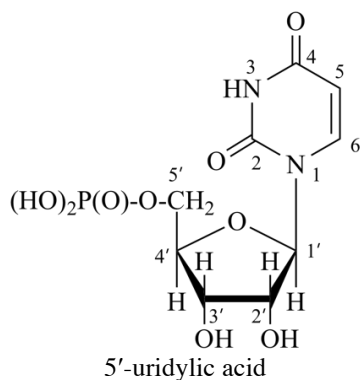
The following are traditional names for esters of nucleosides with phosphoric acid. The primed locant of the ribosyl component is cited to locate the position of the phosphate group.



5'-adenylic acid



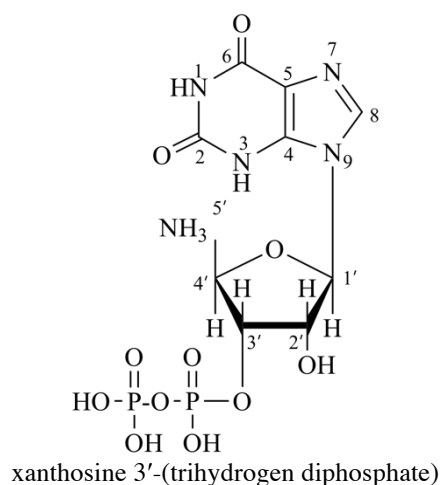
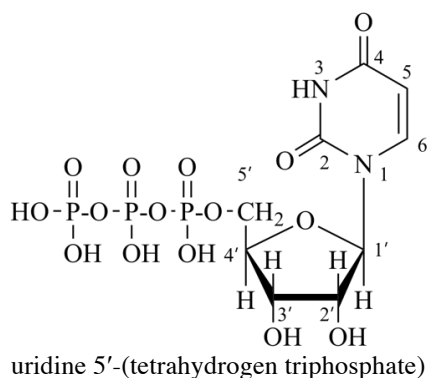




### P-106.2 NUCLEOSIDE DIPHOSPHATES AND TRIPHOSPHATES

Diphosphate, triphosphate, etc. esters of nucleosides are named by citing a phrase such as diphosphate, after the name of the nucleoside. The presence of hydrogen atoms on the diphosphate, triphosphate, etc. component of the molecule is indicated by the words 'hydrogen', 'dihydrogen', etc. Parentheses are used to avoid ambiguity.

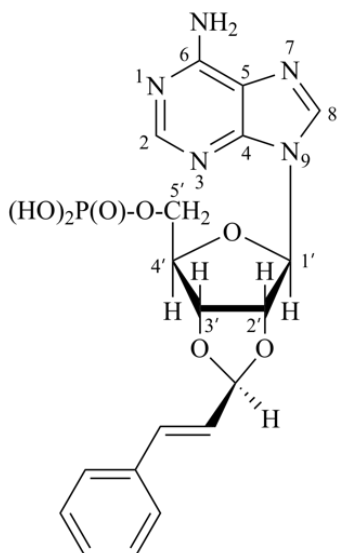
Examples:



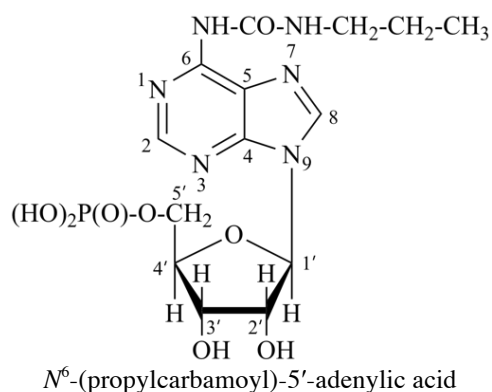
### P-106.3 DERIVATIVES OF NUCLEOTIDES

**P-106.3.1** Derivatives of nucleotides having retained names are named in the same manner as the corresponding nucleoside, i.e., they can be fully substituted on the purine or pyrimidine ring and the ribofuranosyl component may be modified as prescribed for carbohydrates (see P-102.5). The 2- and 3-deoxy modifications of the ribose component are also used.

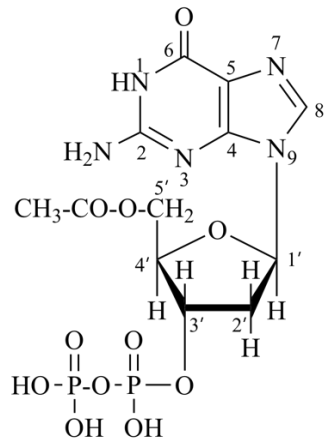
Examples:



2',3'-O-[(1*S*,2*E*)-3-phenylprop-2-ene-1,1-diyl]-5'-adenylic acid



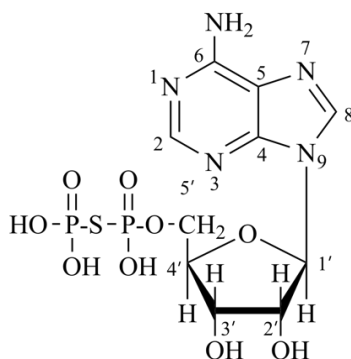
*N*<sup>6</sup>-(propylcarbamoyl)-5'-adenylic acid



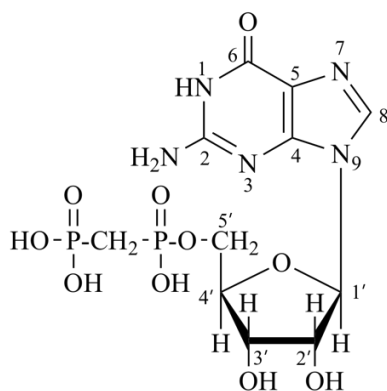
5'-*O*-acetyl-2'-deoxyguanosine 3'-(trihydrogen diphosphate)  
2'-deoxyguanosine 5'-acetate 3'-(trihydrogen diphosphate)

**P-106.3.2** Analogues of nucleoside di- and polyphosphates can be named by the functional replacement techniques applicable to di- and polyphosphoric acids (see P-67.2).

Examples:



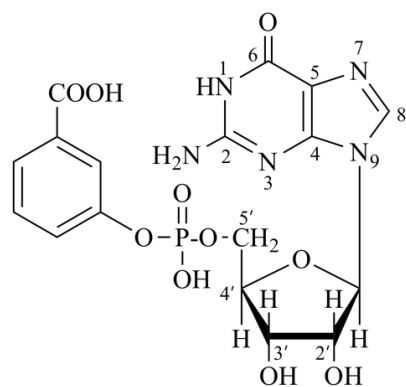
adenosine 5'-(trihydrogen 2-thiodiphosphate)



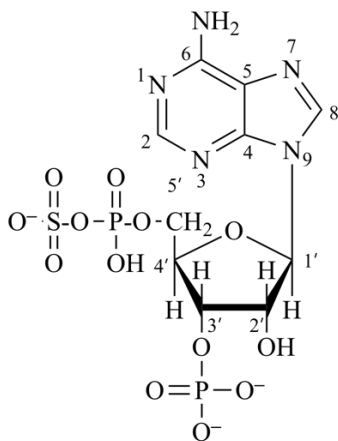
guanosine 5'-[hydrogen (phosphonomethyl)phosphonate]  
 guanosine 5'-(trihydrogen methylenediphosphate)  
 guanosine 5'-(trihydrogen 2-carbadiphosphate)  
 (this name preserves the integrity of the nucleotide name; see P-101.4.3)

**P-106.3.3** In the presence of a characteristic group of higher priority than the phosphoric acid residue, normal substitutive nomenclature principles are applied. Substitutive prefix names are derived from the traditional names for the nucleoside monophosphates by replacing the 'ic acid' ending with 'yl', for example, adenylyl and cytidylyl. Note that the substituent prefix name from inosinic acid is an exception; it is named inosinylyl so that the ending is like the other substituent prefix names derived from the nucleotide monophosphates.

Examples:



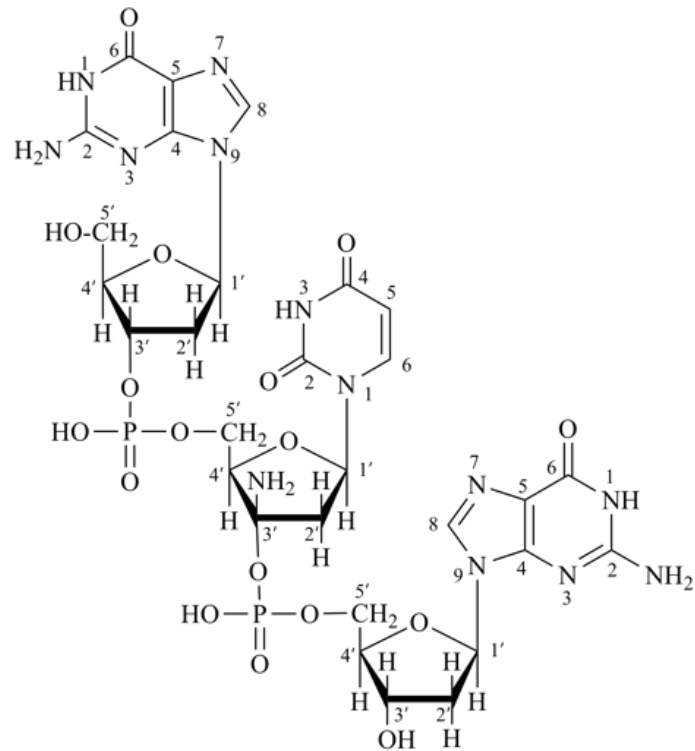
3-(5'-guanylyloxy)benzoic acid



3'-O-phosphonato-5'-adenylyl sulfate  
 3'-phospho-5'-adenylyl sulfate

**P-106.3.4** Oligonucleotides are named using the prefix names derived from the traditional names for the nucleotides.

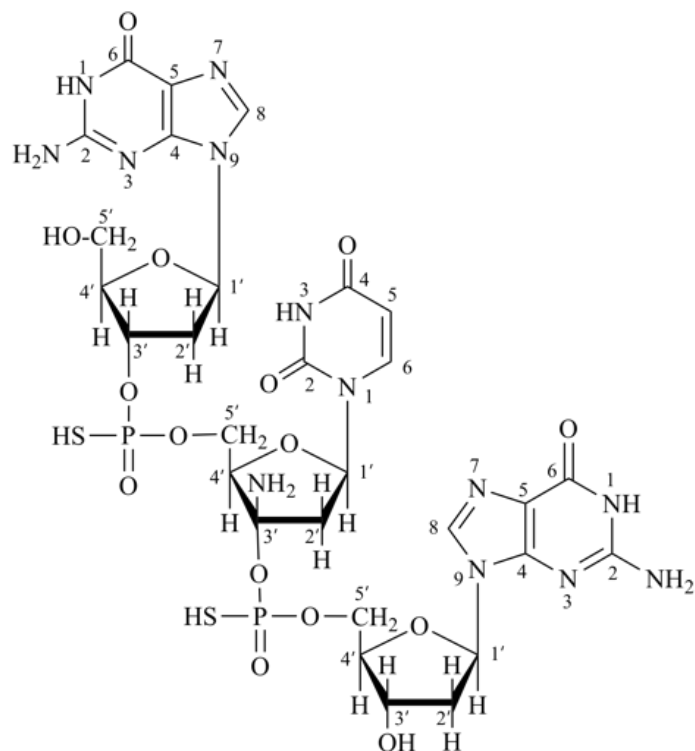
Example:



2'-deoxyguanylyl-(3'→5')-2'-deoxyuridylyl-(3'→5')-2'-deoxyguanosine

**P-106.3.5** When a phosphorothioic acid is used (HS-P in place of HO-P), the prefix *P*-thio is added at the front of the name of the nucleotide.

Example:



2'-deoxy-*P*-thioguananylyl-(3'→5')-2'-deoxy-*P*-thiouridylyl-(3'→5')-2'-deoxyguanosine

## P-107 LIPIDS

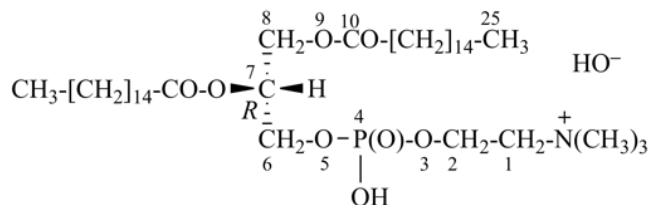
- P-107.0 Introduction
- P-107.1 Definitions
- P-107.2 Glycerides
- P-107.3 Phosphatidic acids
- P-107.4 Glycolipids

### P-107.0 INTRODUCTION





Example:

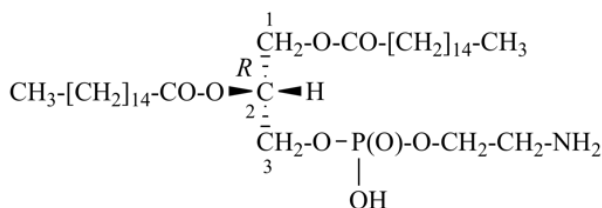


(7*R*)-7-(hexadecanoyloxy)-4-hydroxy-*N,N,N*-trimethyl-4,10-dioxo-3,5,9-trioxa-4λ<sup>5</sup>-phosphapentacosan-1-aminium hydroxide

### P-107.3.5 Phosphatidylethanolamine

The term ‘phosphatidylethanolamines’ [more correctly designated as ‘phosphatidyl(amino)ethanols’] is used to describe the acyl derivatives of phosphatidic acids whose phosphorus acid component is esterified with 2-aminoethanol. Semisystematic names for specific compounds are formed in accordance with the principles, rules, and conventions of substitutive nomenclature.

Example:

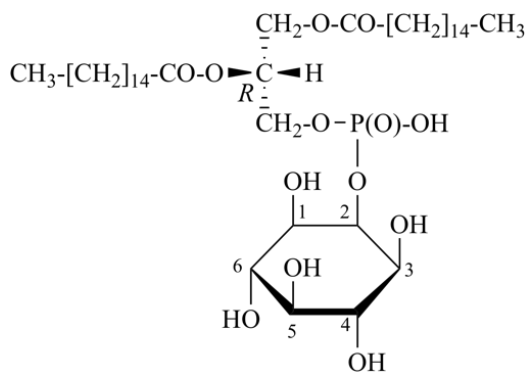


1,2-di(hexadecanoyl)-*sn*-glycero-3-phosphoethanolamine  
(2*R*)-3-[[2-aminoethoxy]hydroxyphosphoryl]oxy}propane-1,2-diyl di(hexadecanoate)

### P-107.3.6 Phosphatidylinositols

The term ‘phosphatidylinositols’ is used to describe the acyl derivatives of phosphatidic acids whose phosphorus acid component is esterified with an inositol molecule. Semisystematic names for specific compounds are formed in accordance with the principles, rules, and conventions of substitutive nomenclature.

Example:



*L*-*myo*-inositol 2-[(2*R*)-2,3-bis(hexadecanoyloxy)propyl hydrogen phosphate]  
(2*R*)-3-[(hydroxy{[(1*s*,2*R*,3*R*,4*s*,5*S*,6*S*)-2,3,4,5,6-pentahydroxycyclohexyl]oxy}phosphoryl]oxy]propane-1,2-diyl di(hexadecanoate)

## P-107.4 GLYCOLIPIDS

### P-107.4.1 Definitions

The term ‘glycolipid’ designates any compound containing one or more monosaccharide residues bound by a glycosidic linkage to a hydrophobic moiety such as an acyl glycerol, a sphingoid (a long chain aliphatic amino alcohol), a ceramide (an *N*-acyl-sphingoid), or a prenylphosphate.

Glycoglycerolipids are glycolipids containing one or more glycerol residues.

Glycosphingolipids designate lipids containing at least one monosaccharide residue and either a sphingoid or a ceramide.

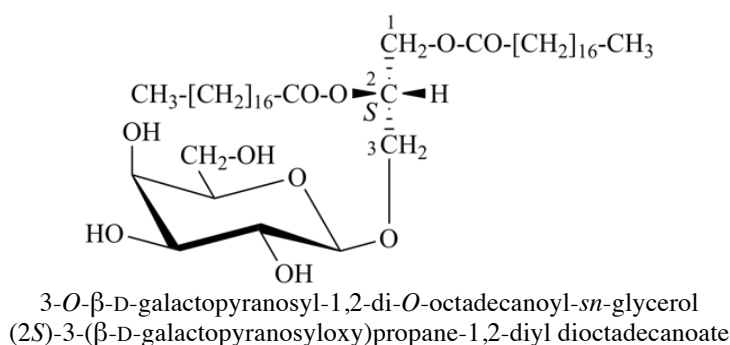
The term ‘glycophosphatidylinositol’ designates glycolipids which contain saccharides glycosidically linked to the inositol moiety of phosphatidylinositols.

Specific compounds are named systematically.

### P-107.4.2 Glycoglycerolipids

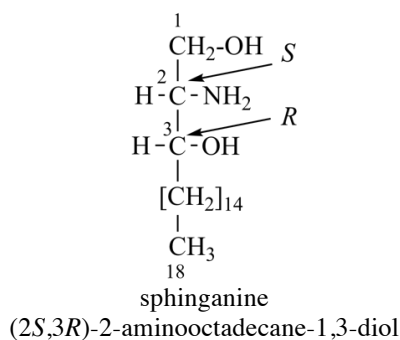
Specific compounds are named on the basis of the parent glycerol, whose configuration is specifically numbered as indicated in P-107.3.2]

Example:



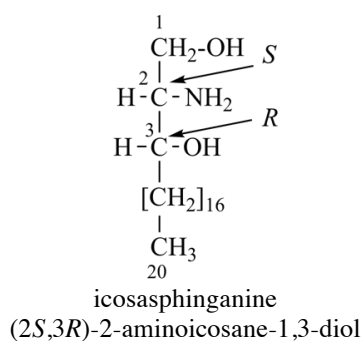
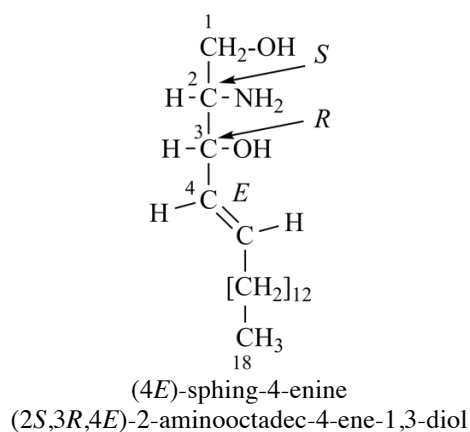
### P-107.4.3 Glycosphingolipids

**P-107.4.3.1** Names are formed by using the retained name 'sphinganine' for the aliphatic amino alcohol having the described absolute configuration. The retained name 'sphinganine' is preferred to the systematic name (2*S*,3*R*)-2-aminooctadecane-1,3-diol.

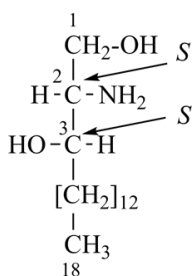


The retained name sphinganine is used to generate the names of unsaturated and *N*-substituted derivatives, as well as *O*-substituted derivatives. Other derivatives, such as hydroxy, oxo, and amino derivatives, as well as isomers with different chain length or other diastereoisomers, are named systematically in accordance with the principles, rules, and conventions of substitutive nomenclature.

Examples:





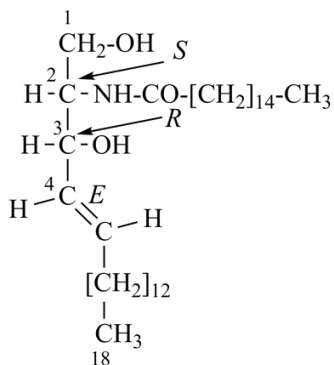


(2*S*,3*S*)-2-amino-octadecane-1,3-diol

### P-107.4.3.2 Ceramides

Ceramides are *N*-acylsphingoids.

Example:

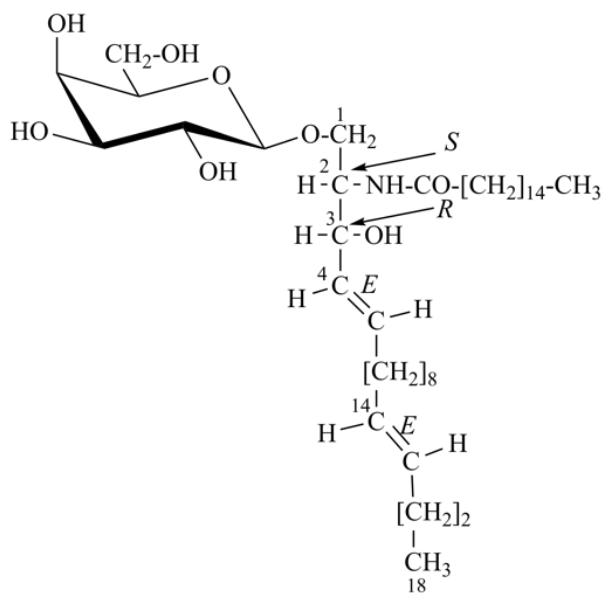


(4*E*)-*N*-hexadecanoylsphing-4-ene  
*N*-[(2*S*,3*R*,4*E*)-1,3-dihydroxyoctadec-4-en-2-yl]hexadecanamide

### P-107.4.3.3 Neutral glycosphingolipids

A neutral glycosphingolipid is a carbohydrate-containing derivative of a sphingoid or ceramide. It is understood that the carbohydrate residue is attached by a glycosidic linkage to 1-*O*-. Preferred systematic names must include all locants.

Example:



(4*E*,14*E*)-1-*O*-(β-*D*-galactopyranosyl)-*N*-hexadecanoylsphinga-4,14-dienine  
*N*-[(2*S*,3*R*,4*E*,14*E*)-1-(β-*D*-galactopyranosyloxy)-3-hydroxyoctadeca-4,14-dien-2-yl]hexadecanamide