Blocking Groups and Annotations for REGISTRY Sequences

BLOCKING GROUPS WITH SHORTCUTS

Shortcut	Blocking Group
Aac	2-amino-2-oxoethyl
2Abz	2-aminobenzoyl
4Abz	4-aminobenzoyl
Ac	acetyl
AcO	acetyloxy
Acm	(acetylamino)methyl
Acr	3-(9-acridinyl)
Adc	tricyclo[3.3.1.1(3,7)]dec-1-yloxy
Aet	2-aminoethyl
All	propenyl
Amoc	(9-anthracenylmethoxy)carbonyl
Aoc	(1,1-dimethylpropoxy)carbonyl
Azoc	[1-methyl-1-[4-(phenylazo)phenyl]ethoxy]carbonyl
Bac	bromoacetyl
Bam	(benzoylamino)methyl
Beoc	(2-bromoethoxy)carbonyl
Bhoc	(diphenylmethoxy)carbonyl
Bic	(5-benzisoxazolylmethoxy)carbonyl
Bmv	1-methyl-3-oxo-3-phenyl-1-propenyl
Bnps	(3-bromo-2-nitrophenyl)thio
BOC	(1,1-dimethylethoxy)carbonyl
Bocae	[[(1,1-dimethylethoxy)carbonyl]amino]ethyl
lBop	2-(phenylmethoxy)phenoxy
Bpoc	(1-[1,1'-biphenyl]-4-yl-1-methylethoxy)carbonyl
Br	bromo
Bs	(4-bromophenyl)sulfonyl
Bt	1H-benzotriazol-1-yl
BTC	[(phenylmethyl)thio]carbonyl
Btm	[(phenylmethyl)thio]methyl
i-Bu	2-methylpropyl
t-Bu	1,1-dimethylethyl
Bum	[(2-methyl-1-oxopropyl)amino]methyl
i-BuO	2-methyl-1-oxopropyl
Bz	benzoyl
2BZ	[(2-bromophenyl)methoxy]carbonyl
4BZ	[(4-bromophenyl)methoxy]carbonyl
Bza	1H-benzimidazol-2-yl
Bzh	diphenylmethyl
Bzl	phenylmethyl
Cac	carboxyacetyl
Cbm	aminocarbonyl
Cbs	(4-chlorophenyl)sulfonyl
CBz	(phenylmethoxy)carbonyl
Cdf	chlorodifluoroacetyl
Ceoc	(2-chloroethoxy)carbonyl

BLOCKING GROUPS WITH SHORTCUTS

Shortcut	Blocking Group
CF3	trifluoromethyl
Chb	(5-chloro-2-hydroxyphenyl)phenylmethylene
Chc	cyclohexylcarbonyl
Chp	cycloheptyl
Chx	cyclohexyl
Chxa	cyclohexylacetyl
Cl	chloro
2-6Clb	(2,6-dichlorophenyl)methyl
Cm	carboxymethyl
Cpc	cyclopentylcarbonyl
Cpe	cyclopentyl
Cpm	cyclopropylmethyl
2CZ	[(2-chlorophenyl)methoxy]carbonyl
4CZ	[(4-chlorophenyl)methoxy]carbonyl
Dbpoc	(2,2-dibromopropoxy)carbonyl
2-4DCZ	[(2,4-dichlorophenyl)methoxy]carbonyl
2-6DCZ	[(2,6-dichlorophenyl)methoxy]carbonyl
Ddz	[1-(3,5-dimethoxyphenyl)-1-methylethoxy]carbonyl
De	2-(diethylamino)ethyl
Dec	1-oxodecyl
Dip	[2-methyl-1-(1-methylethyl)propoxy]carbonyl
Dmoc	[(dimethylamino)oxy]carbonyl
DMB	(3,4-dimethylphenyl)methyl
Dmt	bis(4-methoxyphenyl)methyl
DNP	2,4-dinitrophenyl
DNPS	(2,4-dinitrophenyl)thio
Dpp	diphenoxyphosphinyl
Eac	(ethylamino)carbonyl
Eoc	ethoxycarbonyl
Et	ethyl
F	fluoro
For	formyl
Ft	(1,3-dihydro-1,3-dioxo-2H-isoindol-2-yl)methyl
Glt	4-carboxy-1-oxobutyl
Hex	1-oxohexyl
I	iodo
Ioc	(2-methylpropoxy)carbonyl
Ipa	7-methyl-1-oxooctyl
Ips	(4-iodophenyl)sulfonyl
Kpc	(6-oxo-2-piperidinyl)carbonyl
MOB	(4-methoxyphenyl)methyl
MOS	(4-methoxyphenyl)sulfonyl
Mac	4-methyl-7-amino-coumaryl
Mal	3-carboxy-1-oxo-2-propenyl
Mbh	bis(4-methoxyphenyl)methyl
Me	methyl
MeOe	2-methoxy-2-oxoethyl
Mhoc	[(1-methylcyclohexyl)oxy]carbonyl

BLOCKING GROUPS WITH SHORTCUTS

Shortcut	Blocking Group
Mmt	(4-methoxyphenyl)diphenylmethyl
Moz	[(4-methoxyphenyl)methoxy]carbonyl
Mpt	dimethylphosphinothioyl
Ms	methylsulfonyl
Msc	[2-(methylsulfonyl)ethoxy]carbonyl
Msi	methylsulfinyl
Msp	4-(methylsulfonyl)phenyl
Mtos	(2,4,6-trimethoxyphenyl)sulfonyl
Mtp	4-(methylthio)phenyl
Mts	(2,4,6-trimethylphenyl)sulfonyl
Mz	[[4-[(4-methoxyphenyl)azo]phenyl]methoxy]carbonyl
N	nitro
N3	azido
Nabs	[4-[(4-hydroxy-1-naphthalenyl)azo]phenyl]sulfonyl
1-Naph	1-naphthalenyl
2-Naph	2-naphthalenyl
Ng	2-methoxy-4-nitrophenyl
Ngu	[imino(nitroamino)methyl]amino
NH2	amino
Nis	(4-nitrophenyl)sulfonyl
Nm	3-nitrophenyl
No	2-nitrophenyl
Np	4-nitrophenyl
Npe	2-nitro-1-phenylethyl
Nps	(2-nitrophenyl)thio
Ns	2-nitro-4-sulfophenyl
O	oxygen
Oct	1-oxooctyl
2OHEt	2-hydroxyethyl
2OHPh	2-hydroxyphenyl
Ole	1-oxo-9-octadecenyl
Pa	1-oxononyl
Pal	1-oxohexadecyl
Pbp	pentabromophenyl
Pcp	pentachlorophenyl
Pfp	pentafluorophenyl
Ph	phenyl
Pht	2-carboxybenzoyl
Pic	4-pyridinylmethyl
2Pip	2-piperidinyl
Pipoc	(1-piperidinyloxy)carbonyl

BLOCKING GROUPS WITH SHORTCUTS

Shortcut	Blocking Group
Pnb	(4-nitrophenyl)methyl
PO2	phosphono
Poc	(cyclopentyloxy)carbonyl
Ppt	diphenylphosphinothioyl
Pr	propyl
i-Pr	1-methylethyl
Ptc	(phenylamino)thioxomethyl
Py	2-pyridinyl
3Py	3-pyridinyl
4Py	4-pyridinyl
Pz	[[4-(phenylazo)phenyl]methoxy]carbonyl
Q	quinolinyl
QC	5-chloro-8-quinolinyl
Qu	8-quinolinyl
Qxc	2-quinoxalinylcarbonyl
Sbz	2-sulfobenzoyl
Scm	(carboxymethyl)thio
SO3H	sulfo
Su	2,5-dioxo-1-pyrrolidinyl
Suc	3-carboxy-1-oxopropyl
Tac	[[(4-methylphenyl)sulfonyl]amino]carbonyl
Tbs	(1,1-dimethylethyl)dimethylsilyl
TBZ	phenylthioxomethyl
Tcboc	(2,2,2-trichloro-1,1-dimethylethoxy)carbonyl
Tce	2,2,2-trichloroethyl
Tcp	2,4,5-trichlorophenyl
Tec	[2-[(4-methylphenyl)sulfonyl]ethoxy]carbonyl
Teoc	(2,2,2-trichloroethoxy)carbonyl
Tfe	2,2,2-trifluoroethyl
Tfp	2,2,3,3-tetrafluoro-1-oxopropyl
Tmb	(2,4,6-trimethylphenyl)methyl
TNP	2,4,6-trinitrophenyl
Tos	(4-methylphenyl)sulfonyl
Tosa	[(4-methylphenyl)sulfonyl]amino
Trit	triphenylmethyl
Trs	(triphenylmethyl)thio
5Urd	5'-uridylyl
Vi	ethenyl
Xan	9H-xanthen-9-yl
Za	[(phenylmethoxy)carbonyl]amino
Zae	[[(phenylmethoxy)carbonyl]amino]ethyl
ZNO2	[(4-nitrophenyl)methoxy]carbonyl
Zoa	[[(phenylmethoxy)carbonyl]oxy]acetyl

BLOCKING GROUPS WITH SHORTCUTS

Alphabetized by blocking group name

Blocking Group Name	Shortcut
acetyl	Ac
(acetylamino)methyl	Acm
acetyloxy	AcO
3-(9-acridinyl)	Acr
amino	NH2
2-aminobenzoyl	2Abz
aminocarbonyl	Cbm
2-aminoethyl	Aet
2-amino-2-oxoethyl	Aac
(9-anthracenylmethoxy)carbonyl	Amoc
azido	N3
1H-benzimidazol-2-yl	Bza
1H-benzotriazol-1-yl	Bt
(5-benzisoxazolylmethoxy)carbonyl	Bic
benzoyl	Bz
(benzoylamino)methyl	Bam
(1-[1,1'-biphenyl]-4-yl-1-methylethoxy)carbonyl	Bpoc
bis(4-methoxyphenyl)methyl	Dmt
bromo	Br
bromoacetyl	Bac
(2-bromoethoxy)carbonyl	Beoc
(3-bromo-2-nitrophenyl)thio	Bnps
[(2-bromophenyl)methoxy]carbonyl	2BZ
(4-bromophenyl)sulfonyl	Bs
carboxyacetyl	Cac
2-carboxybenzoyl	Pht
carboxymethyl	Cm
(carboxymethyl)thio	Scm
4-carboxy-1-oxobutyl	Glt
3-carboxy-1-oxo-2-propenyl	Mal
3-carboxy-1-oxopropyl	Suc
chloro	Cl
chlorodifluoroacetyl	Cdf
(2-chloroethoxy)carbonyl	Ceoc
(5-chloro-2-hydroxyphenyl)phenylmethylene	Chb
[(2-chlorophenyl)methoxy]carbonyl	2CZ
(4-chlorophenyl)sulfonyl	Cbs
5-chloro-8-quinolinyl	QC
cycloheptyl	Chp
cyclohexyl	Chx
cyclohexylacetyl	Chxa
cyclohexylcarbonyl	Chc

BLOCKING GROUPS WITH SHORTCUTS

Blocking Group Name	Shortcut
cyclopentyl	Cpe
cyclopentylcarbonyl	Cpc
(cyclopentyloxy)carbonyl	Poc
cyclopropylmethyl	Cpm
(2,2-dibromopropoxy)carbonyl	Dbpoc
[(2,4-dichlorophenyl)methoxy]carbonyl	2-4DCZ
2-(diethylamino)ethyl	De
(1,3-dihydro-1,3-dioxo-2H-isoindol-2-yl)methyl	Ft
[1-(3,5-dimethoxyphenyl)-1-methylethoxy]carbonyl	Ddz
[(dimethylamino)oxy]carbonyl	Dmoc
[[(1,1-dimethylethoxy)carbonyl]amino]ethyl	Bocae
(1,1-dimethylethoxy)carbonyl	BOC
1,1-dimethylethyl	t-Bu
(1,1-dimethylethyl)dimethylsilyl	Tbs
(3,4-dimethylphenyl)methyl	DMB
dimethylphosphinothioyl	Mpt
(1,1-dimethylpropoxy)carbonyl	Aoc
2,4-dinitrophenyl	DNP
(2,4-dinitrophenyl)thio	DNPS
2,5-dioxo-1-pyrrolidinyl	Su
diphenoxyphosphinyl	Dpp
(diphenylmethoxy)carbonyl	Bhoc
diphenylmethyl	Bzh
diphenylphosphinothioyl	Ppt
ethenyl	Vi
ethoxycarbonyl	Eoc
ethyl	Et
(ethylamino)carbonyl	Eac
fluoro	F
formyl	For
9H-xanthen-9-yl	Xan
2-hydroxyethyl	2OHEt
[4-[(4-hydroxy-1-naphthalenyl)azo]phenyl]sulfonyl	Nabs
2-hydroxyphenyl	2OHPh
[imino(nitroamino)methyl]amino	Ngu
iodo	I
(4-iodophenyl)sulfonyl	Ips
2-methoxy-4-nitrophenyl	Ng
2-methoxy-2-oxoethyl	MeOe

BLOCKING GROUPS WITH SHORTCUTS

Blocking Group Name	Shortcut
[[4-[(4-methoxyphenyl)azo]phenyl]methoxy]carbonyl	Mz
(4-methoxyphenyl)diphenylmethyl	Mmt
[(4-methoxyphenyl)methoxy]carbonyl	Moz
(4-methoxyphenyl)methyl	MOB
(4-methoxyphenyl)sulfonyl	OS
methyl	Me
4-methyl-7-amino-coumaryl	Mac
[(1-methylcyclohexyl)oxy]carbonyl	Mhoc
1-methylethyl	i-Pr
[2-methyl-1-(1-methylethyl)propoxy]carbonyl	Dip
7-methyl-1-oxooctyl	Ipa
1-methyl-3-oxo-3-phenyl-1-propenyl	mv
2-methyl-1-oxopropyl	i-BuO
[(2-methyl-1-oxopropyl)amino]methyl	Bum
[1-methyl-1-[4-(phenylazo)phenyl]ethoxy]carbonyl	Azoc
[(4-methylphenyl)sulfonyl]amino	Tosa
[[(4-methylphenyl)sulfonyl]amino]carbonyl	Tac
(4-methylphenyl)sulfonyl	Tos
[2-[(4-methylphenyl)sulfonyl]ethoxy]carbonyl	Tec
(2-methylpropoxy)carbonyl	Ioc
(2-methylpropoxy)methyl	iBom
2-methylpropyl	i-Bu
methylsulfinyl	Msi
methylsulfonyl	Ms
[2-(methylsulfonyl)ethoxy]carbonyl	Msc
4-(methylsulfonyl)phenyl	Msp
4-(methylthio)phenyl	Mtp
1-naphthalenyl	1-Naph
nitro	N
2-nitro-1-phenylethyl	Npe
[(4-nitrophenyl)methoxy]carbonyl	ZNO2
(4-nitrophenyl)methyl	Pnb
2-nitrophenyl	No
3-nitrophenyl	Nm
4-nitrophenyl	Np
(4-nitrophenyl)sulfonyl	Nis
(2-nitrophenyl)thio	Nps
2-nitro-4-sulfophenyl	Ns
1-oxodecyl	Dec
1-oxohexadecyl	Pal
1-oxohexyl	Hex
1-oxononyl	Pa
1-oxo-9-octadecenyl	Ole

BLOCKING GROUPS WITH SHORTCUTS

Blocking Group Name	Shortcut
1-oxooctyl	Oct
(6-oxo-2-piperidinyl)carbonyl	Kpc
oxygen	O
pentabromophenyl	Pbp
pentachlorophenyl	Pcp
pentafluorophenyl	Pfp
phenyl	Ph
(phenylamino)thioxomethyl	Ptc
[[4-(phenylazo)phenyl]methoxy]carbonyl	Pz
[(phenylmethoxy)carbonyl]amino	Za
[[(phenylmethoxy)carbonyl]amino]ethyl	Zae
(phenylmethoxy)carbonyl	CBz
[[(phenylmethoxy)carbonyl]oxy]acetyl	Zoa
2-(phenylmethoxy)phenoxy	Bop
phenylmethyl	Bzl
[(phenylmethyl)thio]carbonyl	BTC
[(phenylmethyl)thio]methyl	Btm
phenylthioxomethyl	TBZ
phosphono	PO2
2-piperidinyl	2Pip
(1-piperidinyloxy)carbonyl	Pipoc
propenyl	All
propyl	Pr
2-pyridinyl	Py
4-pyridinylmethyl	Pic
quinolinyl	Q
8-quinolinyl	Qu
2-quinoxalinylcarbonyl	Qxc
sulfo	SO3H
2-sulfobenzoyl	Sbz
2,2,3,3-tetrafluoro-1-oxopropyl	Tfp
(2,2,2-trichloro-1,1-dimethylethoxy)carbonyl	Tcboc
(2,2,2-trichloroethoxy)carbonyl	Teoc
2,2,2-trichloroethyl	Tce
2,4,5-trichlorophenyl	Tcp
tricyclo[3.3.1.1(3,7)]dec-1-yloxy	Adc
trifluoroacetyl	Tfa
2,2,2-trifluoroethyl	Tfe
trifluoromethyl	CF3

BLOCKING GROUPS WITHOUT SHORTCUTS

acetylamino
2-amino-3-(4-hydroxyphenyl)propyl
aminoiminomethyl
3-[(4-azidophenyl)azo]
benzenemethoxy
4-benzoylbenzoyl
bis(2-chloroethyl)amino
bis(ethylamino)methylene
3-bromo-1-oxo-propyl
(butylamino)carbonyl
(butylamino)iminoethyl
1-(carboxymethyl)cyclohexyl
(3-carboxy-4-nitrophenyl)thio
3-carboxypropyl
[(chloroacetyl)amino]methyl
1-[(cyclohexylamino)carbonyl]-2-methylpropyl
(cyclohexyloxy)carbonyl
1,3-dihydro-1,3-dioxo-2H-H-isoindol-2-yl
10,11-dihydro-5H-dibenzo[a,d]cyclohepten-5-yl
[(4,5-dimethoxy-2-nitrophenyl)methoxy]carbonyl
2-(dimethylamino)ethyl
[5-(dimethylamino)-1-naphthalenyl]sulfonyl
[[[(1,1-dimethylethoxy)carbonyl]amino]oxy]acetyl
[[[(1,1-dimethylethoxy)carbonyl]amino]oxy]-1-oxopropyl
(1,1-dimethylethyl)thio
2,2-dimethyl-1-oxopropyl
(1,6-dioxoheptyl)amino
3-ethoxy-2-(ethoxycarbonyl)-3-oxopropyl
(ethylamino)iminoethyl
(ethylthio)thioxomethyl
(9H-fluoren-9-ylmethoxy)carbonyl
2,2,3,3,4,4,4-heptafluoro-1-oxobutyl
(3a,4,5,6,7,7a-hexahydro-3a,7a-dihydroxy-1H-benzimidazol-2-yl)
1,3,3a,4,7,7a-hexahydro-1,3-dioxo-4,7-methano-2H-isoindol-2-yl
3-hydroxy-8-methyl-1-oxononyl
3-hydroxy-1-oxodecyl
3-hydroxy-1-oxotetradecyl
3-hydroxy-1-oxotridecyl
3-(4-hydroxyphenyl)-1-oxopropyl
1-iminoethyl
imino(nitroamino)methyl
imino[(2,2,2-trifluoroethyl)amino]methyl
(1-mercaptocyclohexyl)acetyl
(4-methoxy-2,6-dimethylphenyl)sulfonyl
4-methoxy-1,4-dioxobutyl
4-methoxy-4-oxobutyl
[(4-methoxyphenyl)methyl]sulfinyl
(4-methoxy-2,3,6-trimethylphenyl)sulfonyl
[(1-methylethyl)amino]carbonyl
3-methyl-1-oxobutyl
2-methyl-1-oxo-2-propenyl
(4-methylphenyl)methyl
[(4-methylphenyl)methyl]sulfonyl
4-(methylsulfonyl)-1-oxobutyl

4-(methylthio)-1-oxobutyl
3-(methylthio)propyl
4-(4-nitrophenoxy)-4-oxobutyl
(3-nitro-2-pyridinyl)thio
2-oxo-2-phenylethyl
1-oxo-3-[(phenylmethyl)thio]propyl
1-oxo-3-phenyl-2-propenyl
1-oxo-3-phenylpropyl
1-oxo-3-[4-(sulfooxy)phenyl]propyl
(pentamethylphenyl)sulfonyl
[(phenylacetyl)amino]methyl
4-(phenylazo)benzoyl
4-(phenylazo)phenyl
(phenylmethoxy)methyl
3-phenyl-2-oxaziridinyl
1-pyrenyl
2-pyridinylcarbonyl
(4-pyridinylmethoxy)carbonyl
(4-pyridinyloxy)carbonyl
tetrahydro-2H-pyran-2-yl
1,4,5,6-tetrahydro-2-(nitroamino)-4-pyrimidinyl

ANNOTATION FOR CHEMICALLY MODIFIED NUCLEIC ACIDS

This section defines the symbols and terms which are used to annotate chemically modified sequences of nucleic acids. The chemical annotation data appear in the NTE (Note) field in the Registry File. The NTE data for chemically modified nucleic acid sequences may consist of the following types of data: global terms, strand-specific terms, type of modification, location, and description.

Global terms provide a broad classification for the entire nucleic acid sequence. No location is specified for global terms. All chemically modified sequences or strands have the global term of “modified”. Strand number refers to the number of strands in a multistranded complex. The strands are ordered from largest to smallest. Alphabetical order of the sequence residues is used to break a tie. The complete list of global terms and their definitions appears in Table 1.

Strand-specific terms appear in Table 2. These terms have a strand number associated with them, but no position number is specified.

Type of modification is a general term which describes the chemical modification which has occurred in the sequence. The complete list of terms and their definitions appears in Table 3.

Location identifies the nucleoside or linkage in the nucleic acid sequence where the chemical modification has occurred. The sequence is displayed with the nucleoside having a free 5'-hydroxy group on the left and is numbered from left to right. The location of phosphate esters or linkages in the sequence is identified by citing the locants for the nucleosides to which the phosphate is linked, moving from left to right. Unprimed numerical locants refer to the purine or pyrimidine base of a nucleoside and primed numerical locants refer to the sugar moiety. The Greek letter .alpha. is used for the methyl group at the 5 position on thymidine, and N refers to the amino group in adenosine, cytidine or guanosine. P refers to the phosphate linkage. When the location of the chemical modification is not known, the question mark (“?”) appears.

Description terms define the chemical modification made to the nucleoside or linkage. Description terms include symbols for modified nucleosides (Table 4 and 5), terms for chemical groups or chemical modifications (Table 6), generic terms (Table 7), uncommon linkages (Table 8), isotopes (Table 9), stereoisomers (Table 10), or metals (Table 11).

TABLE 1

GLOBAL TERMS

Term in NTE	Definition
modified	Used with the records of all sequences that have been chemically modified.
singlestranded	Used for nucleic acid sequences consisting of one strand.
doublestranded	Used as default value for DNA sequences and all other nucleic acid sequences consisting of two strands.
multistranded (#)	Used when the number of strands is greater than two; the number of strands appears in parentheses.

TABLE 2

STRAND-SPECIFIC TERMS

Term in NTE	Definition
homopolymer	Used when the nucleic acid sequence is replicated an indeterminate number of times. A strand number is associated with “homopolymer” in records for multistranded sequences. Note that the term “5'-phosphate” is used with the term “homopolymer.”
copolymer	Used when a polymer is derived from two or more strands. Strand numbers are associated with “copolymer” to indicate which strands are included. Each strand has the description “5'-phosphate”.
linear	Used for linear sequences.
cyclic	Used to indicate that the 5'-end of the sequence is chemically bonded to its 3'-end, i.e. the sequence is cyclic. A strand number is associated with “cyclic” in records for multistranded sequences. Note that the term “5'-phosphate” is present when “cyclic” is used.

TABLE 3

TYPE OF MODIFICATION TERMS

Term in NTE	Definition
modified base	Used for nucleosides which have been modified by substitution, esterification or replacement. Included are derivatives in which "substitution" of the purine or pyrimidine has resulted in a new ring system (fused, spiro, or bridged) and those in which carbon or nitrogen in the purine or pyrimidine ring has been replaced by another atom. The symbol used to represent the nucleoside in the sequence is a, c, g, t or u. Valid description terms are listed in Tables 4, 6 and 7.
uncommon base	Used for unusual nucleosides which cannot be conveniently described by modifying the normal nucleosides (a, c, g, t or u) as described above. Included are those residues in which a part or all of the purine or pyrimidine ring has been removed and those which contain completely new ring systems unrelated to purine or pyrimidine. Also included are residues which contain an unusual sugar moiety such as a six-carbon sugar. The symbol used to represent the nucleoside in the sequence is x.
DNA-containing	Used to indicate that an RNA sequence contains one or more DNA residues. Description symbols from Table 5 are used to indicate the DNA residues.
RNA-containing	Used to indicate that a DNA sequence contains one or more RNA residues. Description symbols from Table 5 are used to indicate the RNA residues.
modified link	Used to indicate phosphate linkages which have been modified by substitution, esterification or replacement. The location of the modification is indicated by citing the locants for the two adjacent nucleosides.
uncommon link	Used when the normal phosphate linkage has been replaced or lengthened. The location of the "uncommon link" is indicated by citing the locants for the two adjacent nucleosides. Valid description terms are listed in Table 8.
stereoisomer	Used when one or more sugar residues have unusual stereo such as .alpha.-D-erythro- or .beta.-D-xylo-. A complete list of valid description terms is given in Table 10.

metal complex	Used to indicate that the sequence is coordinately complexed with a metal; the element symbol of the metal appears in the description. The nucleoside or nucleosides which are bound to the metal are indicated in the Location. The question mark (“?”) appears when the site of the bonding is unknown. Valid metal element symbols and names are listed in Table 11.
complex	Used when the nucleic acid sequence is associated with a nonmetallic, non-nucleic acid substance. The term “unavailable” is used in the description.
labeled	Used to indicate labeling of any atom in the sequence, substituents or esters. Valid description terms for isotopes are given in Table 9.
covalent bridge	Used to indicate the presence of a bridge of chain and/or ring atoms between two strands. The strand number and location are used to indicate the point of attachment to each strand. The term “unavailable” is used in the description.

TABLE 4

DESCRIPTION TERMS FOR MODIFIED NUCLEOSIDES

Symbol in NTE	Modified Nucleoside	Sequence Symbol
p	pseudouridine	u
i	inosine	i
xan	xanthosine	g
hu	dihydrouridine	u
cm	2'-O-methylcytidine	c
pm	2'-O-methylpseudouridine	u
gm	2'-O-methylguanosine	g
um	2'-O-methyluridine	u
am	2'-O-methyladenosine	a
im	2'-O-methylinosine	i
mla	1-methyladenosine	a
mlp	1-methylpseudouridine	u
mlg	1-methylguanosine	g
mli	1-methylinosine	i
m2a	2-methyladenosine	a
m2g	N-methylguanosine	g
m3c	3-methylcytidine	c
m5c	5-methylcytidine	c
m5u	5-methyluridine (thymidine in a ribonucleotide)	u
m6a	N6-methyladenosine	a
m7g	7-methylguanosine	g
m22g	N,N-dimethylguanosine	g
m26a	N,N-dimethyladenosine	a
ac4c	4-acetylcytidine	c
ac2g	N-acetylguanosine	g
s2t	2-thiothymidine	t
s4t	4-thiothymidine	t
s2c	2-thiocytidine	c
s2u	2-thiouridine	u
s4u	4-thiouridine	u
s6g	6-thioguanosine	g
ib2g	N-(2-methyl-1-oxopropyl)guanosine (N-isobutyrylguanosine)	g
bz6a	N-benzoyladenosine	a
bz4c	N-benzoylcytidine	c
an4c	N-(4-methoxybenzoyl)cytidine (N-p-anisoylcytidine)	c
c7a	7-deazaadenosine	a
m227g	2,2,7-trimethylguanosine	g
m7i	7-methylinosine	i
s6i	6-thioinosine	i
c7i	7-deazainosine	i
c7g	7-deazaguanosine	g

TABLE 5

DESCRIPTION SYMBOLS IN MIXED DNA-RNA SEQUENCES

A sequence is classified as a DNA when 50% or more of the residues contain 2'-deoxy sugars. The other residues are identified as modified nucleosides and described using description symbols in this table. The term “RNA-containing” appears in the type of modification field.

A sequence is classified as an RNA only when more than half of the residues contain .beta.-D-ribofuranosyl sugar moieties. The other residues are identified as modified nucleosides and described using description symbols in this table. The term “DNA-containing” appears in the type of modification field.

Symbol in NTE	Definition	Sequence Symbol
da	These symbols are used in the description to	a
dc	indicate DNA residues in a sequence which is	c
dg	predominately RNA or PNA. (DNA-containing	g
dt	appears in the type of modification field.)	t
du		u
di		i
ra	These symbols are used in the description to	a
rc	indicate RNA residues in a sequence which is	c
rg	predominately DNA or PNA. (RNA-containing	g
ru	appears in the type of modification field.)	u
ri		i
pa	These symbols are used in the description to	a
pc	indicate PNA residues in a sequence which is	c
pg	predominately DNA or RNA. (PNA-containing	g
pt	appears in the type of modification field.)	t
pu		u
pi		i

TABLE 6

DESCRIPTION TERMS FOR CHEMICAL GROUPS AND MODIFICATIONS

A locant indicating the position on the modified nucleoside base, sugar or linkage precedes the term.

Term in NTE	Definition
ac	acetyl
an	anisoyl (4-methoxybenzoyl)
br	bromo
bz	benzoyl
bzl	benzyl (phenylmethyl)
cl	chloro
(2-clph)	(2-chlorophenyl)
(3-clph)	(3-chlorophenyl)
(4-clph)	(4-chlorophenyl)
dmt	dimethoxytrityl [bis(4-methoxyphenyl)phenylmethyl]
dns	dansyl [[5-(dimethylamino-1-naphthalenyl]sulfonyl]
et	ethyl
fl	fluoro
ib	isobutyryl (2-methyl-1-oxopropyl)
io	iodo
me	methyl
mmt	monomethoxytrityl [(4-methoxyphenyl)diphenylmethyl]
mo	methoxy
nh2	amino
oh	hydroxy
ph	phenyl
sh	mercapto
thp	(tetrahydro-2H-pyran-2-yl] ~
tos	tosyl [(4-methylphenyl)sulfonyl]
tr	trityl (triphenylmethyl)
deamino	Removal of the amino group from a nucleoside base.
deoxo	Removal of the keto group from a nucleoside base or the double bonded oxygen from the phosphate group.
deoxy	Removal of a hydroxy group from a sugar or from the phosphate group.
thio	Replacement of any oxygen implied in the sequence by sulfur. May be subsequently esterified or substituted.
dithio	Replacement of both oxygens on a phosphate by sulfur.
phosphate	A hydroxy or mercapto group has been esterified by phosphoric acid.

The following descriptive terms for chemical groups and modifications for nucleotides are added to REGISTRY/ZREGISTRY as of mid-January, 2005; however, the backfile will not be updated.

Term in NTE	**Definition**
boc	t-butyloxycarbonyl
bu	butyl
ibu	isobutyl
sbu	sec-butyl
tbu	tert-butyl
cbz	benzyloxycarbonyl
cho	formyl
dnp	2,4-dinitrophenyl
fmoc	9H-fluoren-9-ylmethoxycarbonyl
pr	propyl
ipr	isopropyl
tms	trimethylsilyl
moe	2'-O-(2-methoxyethyl)
aza	aza
deaza	deaza
ethenyl	ethenyl
2-propenyl	2-propenyl
1-propynyl	1-propynyl
biotin-linked	biotin-linked
cyanine dye-linked	cyanine dye-linked
digoxigenin-linked	digoxigenin-linked
fluorescein-linked	fluorescein-linked
rhodamine-linked	rhodamine-linked
steroid-linked	steroid-linked
porphyrin-linked	porphyrin-linked
psoralen-linked	psoralen-linked
photoadduct	photoadduct
glycosylated	glycosylated
phosphonate	phosphonate
phosphorothioate	phosphorothioate
modified phosphate	modified phosphate

TABLE 7

GENERIC DESCRIPTION TERMS

When the chemical modification cannot be described with terms from Table 4 or Table 6, the generic terms in Table 7 are used to describe the modification. A locant indicating the position on the modified nucleoside base, sugar or linkage precedes the term.

Term in NTE	**Definition**
substituted	Replacement of hydrogen on C, N, O, or S in any of the nucleosides by substituents not shown in Table 4 or Table 6 Included are the hydrogens on the 3'- and 5'-hydroxy groups and the tautomeric forms of the nucleoside keto groups. Substitution may also take place on a phosphate linkage provided a hydroxy or oxo group has been removed.
phosphoramidate	A hydroxy or mercapto group has been esterified by phosphoramidic acid which is usually N-substituted.
ester	Used when esters of hydroxy groups or phosphate linkages have been formed by acids or alcohols not shown above, i.e., phosphoric and phosphoramidic acids and acyl, aryl and alkyl groups in Table 6.
modified adenosine modified cytidine modified guanosine modified thymidine modified uridine	Used when normal nucleoside base has been modified by removing a carbon or nitrogen from the ring and replacing it with another atom or by “unusual” substitution which results in the formation of a new fused, bridged, or spiro ring system. Nucleosides of this type are represented by a, c, g, t or u in the sequence.
thymidine dimer	Used when one or more bonds have been formed, usually by irradiation, between the pyrimidine rings of two adjacent thymidines. The normal symbol “t” is used in the sequence and the position of the bonded thymidines is indicated in the Location.
unavailable	This term is used when none of the other descriptions apply. It is also used with the Type of Modification Term “uncommon base” and with “uncommon link” when the linkage cannot be described with a term from Table 8.

TABLE 8

DESCRIPTION TERMS FOR UNCOMMON LINKAGES

Uncommon linkages that are multiple phosphate entities are defined in the description using the term “phosphate” with the appropriate numerical prefix, e.g. “triphosphate” or “tetraphosphate”. When nucleosides are not linked at the normal 3' and 5' (3'->5') positions by the phosphate group, terms such as “(2'->5')” or “(3'->3')” are used. Uncommon linkages which cannot be described by the terms in Table 8 receive the term “unavailable” in the description.

Multiphosphates	**Uncommon locant sets**
diphosphate	(2'->2')
triphosphate	(2'->3')
tetraphosphate	(2'->5')
pentaphoaphate	(3'->2')
hexaphosphate	(3'->3')
heptaphosphate	(5'->2')
octaphosphate	(5'->3')
nonaphosphate	(5'->5')
decaphosphate	

TABLE 9

DESCRIPTION TERMS FOR ISOTOPES

The term “labeled” is used in the type of modification for isotopically labeled nucleic acid sequences. The specific nucleosides or linkages which have been labeled are specified in the location. The question mark (“?”) appears when the site of the labelling is unknown. The description indicates the specific isotope, e.g., N15, P32, H2, etc. Valid description terms for isotopes are listed in Table 9. The specific site of the labelling is specified if possible.

Other isotope values and isotopes of other elements will be added to this list as needed.

Hydrogen	**Carbon**	**Nitrogen**	**Oxygen**	**Phosphorus**	**Sulfur**
H2	C10	N12	O15	P29	S32
H3	C11	N13	O17	P30	S33
	C13	N15	O18	P32	S34
	C14	N16		P33	S35
	C15				S36
					S37

TABLE 10

DESCRIPTION TERMS FOR STEREOISOMERS

.alpha.-D-arabino	.beta.-D-lyxo	.alpha.-L-threo
.beta.-D-arabino	.alpha.-L-lyxo	.beta.-L-threo
.alpha.-L-arabino	.beta.-L-lyxo	.alpha.-D-xylo
.beta.-L-arabino	.alpha.-D-ribo	.beta.-D-xylo
.alpha.-D-erythro	.alpha.-L-ribo	.alpha.-L-xylo
.alpha.-L-erythro	.beta.-L-ribo	.beta.-L-xylo
.beta.-L-erythro	.alpha.-D-threo	R
.alpha.-D-lyxo	.beta.-D-threo	S

TABLE 11

DESCRIPTION TERMS FOR METALS

Symbol in NTE	Metal	Symbol in NTE	Metal	Symbol in NTE	Metal
Ac	actinium	Ge	germanium	Pr	praseodymium
Ag	silver	Hf	hafnium	Pt	platinum
Al	aluminum	Hg	mercury	Pu	plutonium
Am	americium	Ho	holmium	Ra	radium
Au	gold	In	indium	Rb	rubidium
Ba	barium	Ir	iridium	Re	rhenium
Be	beryllium	K	potassium	Rh	rhodium
Bi	bismuth	La	lanthanum	Ru	ruthenium
Bk	berkelium	Li	lithium	Sb	antimony
Ca	calcium	Lr	lawrencium	Sc	scandium
Cd	cadmium	Lu	lutetium	Sm	samarium
Ce	cerium	Md	mendelevium	Sn	tin
Cf	californium	Mg	magnesium	Sr	strontium
Cm	curium	Mn	manganese	Ta	tantalum
Co	cobalt	Mo	molybdenum	Tb	terbium
Cr	chromium	Na	sodium	Tc	technetium
Cs	cesium	Nb	niobium	Th	thorium
Cu	copper	Nd	neodymium	Ti	titanium
Dy	dysprosium	Ni	nickel	Tl	thallium
Er	erbium	No	nobelium	Tm	thulium
Es	einsteinium	Np	neptunium	U	uranium
Eu	europium	Os	osmium	V	vanadium
Fe	iron	Pa	protactinium	W	tungsten
Fm	fermium	Pb	lead	Y	yttrium
Fr	francium	Pd	palladium	Yb	ytterbium
Ga	gallium	Pm	promethium	Zn	zinc
Gd	gadolinium	Po	polonium	Zr	zirconium