PSI-MOD

Annotation rules

Version 03.04.2025

Annotation rules for [Term]

- Mandatory keys
  - id:
  - name:
  - def:
  - is_a:

- Optional keys
  - comment
  - subset:
  - synonym:
  - xref:
  - relationship:

[Typedef]
descriptors

Annotation rules for [Term]

- id:
  - PSI-MOD identifier : MOD:nnnnn
    - MOD:01888

- [Typedef]
  - Descriptors

  - =>Add descriptors for references in def (Pubmed, RESID, ChEBI(?), others)

Annotation rules for [Term]

- Name:
  - full exact name non ambiguous , usage of common name ok (leaf terms)
    - N4-methyl-L-asparagine and not N-methyl-asparagine
  - Name is final product (not a reaction description)
    - methylated asparagine and not arginine methylation
  - Can include a precision about origin residue, in parentheses, to remove ambiguity
    - 2-pyrrolidone-5-carboxylic acid (Glu)

Annotation rules for [Term]

- def:
  - Explanation of a conversion from source to MOD with references
    - «A protein modification that effectively converts a source residue to MOD.» [Pubmed:123456, RESID:AA0123, ~~UniMod:22#C~~]
    - A protein modification that crosslinks two cysteine residues by formation of a chain of two or more bonded sulfur atoms." [PubMed:18688235]
    - A protein modification that effectively crosslinks an N-formyl-L-methionine residue and an L-histidine residue to form N-[(L-histidin-1'-yl)methyl]-L-methionine." [PubMed:19622680, RESID:AA0566#FMET]
    - A protein modification that effectively converts four L-cysteine residues and a four-iron four-sulfur cluster to tetrakis-L-cysteinyl tetrairon tetrasulfide." [PubMed:3351918, PubMed:7803404, PubMed:7819196, PubMed:932007, RESID:AA0140]
      - comment: Cross-link 4.

Do not need unimod ref as available in xref and source

See if one can encode the litterature reference in a different way

Annotation rules for [Term]

- Issues with def:
  - Multiple, identical defs
    - def: "modification from DeltaMass" [DeltaMass:0]

Annotation rules for [Term]

- comment:
  - Comment with some informational syntax
    - Cross-link 3. => cross-link that involves 3 residues
  - Comment free text
    - This modification occurs uniquely in translation initiation factor eIF-5A [JSG]. Not a goal to annotate functional aspects: keep the few, but do not advocate to extend this usage
  - Comment free text
    - From DeltaMass: Average Mass: 354 Formula:C 16 H 26 O 4 N 3 S 1 (formula incorrect, N and O reversed) Monoisotopic Mass Change:354.172 Average Mass Change:354.471 References:PE Sciex.
  - If multiple residues involved, complete xref: Origin with a list of residue
    - xref: Origin: "C, Q"

Annotation rules for [Term]

- subset:
  - We have psi-mod-slim (subset of most frequently occurred MODS): consistency / maintainance?
    - PSI-MOD-slim

id: MOD:00117
name: S-(L-isoglutamyl)-L-cysteine
def: "A protein modification that effectively crosslinks an L-cysteine residue and an L-glutamine residue by a thioester bond with the formation of S-(L-isoglutamyl)-L-cysteine and the release of ammonia." [ChEBI:22021, DeltaMass:0, PubMed:6838833, RESID:AA0108]
comment: Cross-link 2; DeltaMass calculates the mass difference from glutamic acid rather than glutamine.
subset: PSI-MOD-slim
synonym: "(2S)-2-amino-5-[(2R)-2-amino-2-carboxyethyl]sulfanyl-5-oxopentanoic acid" EXACT RESID-systematic []
synonym: "(S,R)-2-amino-4-[S-(2-amino-2-carboxyethyl)thiocarboxy]butanoic acid" EXACT RESID-alternate []
synonym: "2-amino-5-(2-amino-2-carboxyethyl)thio-5-oxopentanoic acid" EXACT RESID-alternate []
synonym: "CROSSLNK Isoglutamyl cysteine thioester (Cys-Gln)" EXACT UniProt-feature []
synonym: "gamma-(S-cysteinyl)glutamic acid" EXACT RESID-alternate []
synonym: "S-(L-isoglutamyl)-L-cysteine" EXACT RESID-name []
synonym: "S-gamma-glutamyl (crosslinked to cysteine)" EXACT DeltaMass-label []
synonym: "XLNK-SCys-5Glu(Gln)" EXACT PSI-MOD-label []
xref: DiffAvg: "-17.03"
xref: DiffFormula: "C 0 H -3 N -1 O 0 S 0"
xref: DiffMono: "-17.026549"
xref: Formula: "C 8 H 10 N 2 O 3 S 1"
xref: MassAvg: "214.24"
xref: MassMono: "214.041213"
xref: Origin: "C, Q"
xref: Source: "natural"
xref: TermSpec: "none"
is_a: MOD:00395 ! thioester crosslinked residues
is_a: MOD:00907 ! modified L-glutamine residue
is_a: MOD:00946 ! crosslinked residues with loss of ammonia

Annotation rules for [Term]

- synonym:
  - Names from other MOD repositories, precision (EXACT, RELATED) for the synonym, repository-term []; repository terms defined in PSI-MOD file header
    - synonym: "(2S)-2-amino-N4-methylbutanediamic acid" EXACT RESID-systematic []
    - synonym: "beta-aspartyl methylamide" EXACT RESID-alternate []
    - synonym: "beta-methylasparagine" RELATED RESID-misnomer []
    - synonym: "Methyl" RELATED PSI-MS-label []
    - synonym: "Methylation" RELATED UniMod-description []
    - synonym: "methyln" EXACT OMSSA-label []
    - synonym: "MOD_RES N4-methylasparagine" EXACT UniProt-feature []
    - synonym: "N(gamma)-methylasparagine" EXACT RESID-alternate []
    - synonym: "N-methylasparagine" EXACT RESID-alternate []
    - synonym: "N4-methyl-L-asparagine" EXACT RESID-name []
    - synonym: "N4-methylated L-asparagine" EXACT PSI-MOD-alternate []
    - synonym: "N4Me1Asn" EXACT PSI-MOD-label []

What is a PSI-MOD-label?

UniMod synonyms are always flagged as RELATED

Annotation rules for [Term]

- xref: DiffAvg: "14.03"
  - 2 digit precision; can be none
- xref: DiffFormula: "C 1 H 2 N 0 O 0 "
  - Elements in alphabetic order, isotopes : (nn)E (13)C; can be none order if different isotopes : (13)C 1 C 3
- xref: DiffMono: "14.015650 "
  - 6 digit precision; can be none
- xref: Formula: "C 6 H 10 N 2 O 2 "
  - Elements in alphabetic order, isotopes : (nn)E (13)C; can be none
- xref: MassAvg: "142.16 "
  - 2 digit precision; can be none; if Formula with isotopes, value is mono at 2 digit precision
- xref: MassMono: "142.074228 "
  - 6 digit precision; can be none
- xref: Origin: "Q "
  - AA in one letter code; X is any; can be none, separated with comma and space ; can also be a MOD:nnnnn id in multiple steps Mods
- xref: Source: "natural "
  - Natural, artifact, hypothetical, none :
  - Priority ? Multiple are possible? One only, priority N, A, H, N, no import of literature refs
- xref: TermSpec: "none«
  - N-term, C-term, none
- xref: smiles: "C(*)(=O)[C@H](COP(O)(=O)O)N*“
  - String of residues, non charged, with * as boundaries

[Term]
id: MOD:00051
name: N-acetyl-L-aspartic acid
def: "A protein modification that effectively converts an L-aspartic acid residue to N-acetyl-L-aspartic acid." [ChEBI:21547, PubMed:1560020, PubMed:2395459, RESID:AA0042]
subset: PSI-MOD-slim
synonym: "(2S)-2-(acetamido)butanedioic acid" EXACT RESID-systematic []
synonym: "2-(acetylamino)butanedioic acid" EXACT RESID-alternate []
synonym: "AcAsp" EXACT PSI-MOD-label []
synonym: "MOD_RES N-acetylaspartate" EXACT UniProt-feature []
synonym: "N-acetyl-L-aspartic acid" EXACT RESID-name []
synonym: "N-acetylated L-aspartic acid" EXACT PSI-MOD-alternate []
xref: DiffAvg: "42.04"
xref: DiffFormula: "C 2 H 2 N 0 O 1"
xref: DiffMono: "42.010565"
xref: Formula: "C 6 H 8 N 1 O 4"
xref: MassAvg: "158.13"
xref: MassMono: "158.045333"
xref: Origin: "D"
xref: Source: "natural"
xref: TermSpec: "N-term"
is_a: MOD:00904 ! modified L-aspartic acid residue
is_a: MOD:01458 ! alpha-amino acetylated residue

Annotation rules for [Term]

- is_a:
- relationship:
  - Rule?
    - derives_from MOD:01464 ! protonated L-methionine (L-methioninium) residue
- Ad minima, 2 is_a lines (one by main branch if possible)

id: MOD:00079
name: N4-methyl-L-asparagine
def: "A protein modification that effectively converts an L-asparagine residue to N4-methyl-L-asparagine." [OMSSA:75, PubMed:11875433, PubMed:2356973, PubMed:3782095, RESID:AA0070, UniMod:34#N]
subset: PSI-MOD-slim
synonym: "(2S)-2-amino-N4-methylbutanediamic acid" EXACT RESID-systematic []
synonym: "beta-aspartyl methylamide" EXACT RESID-alternate []
synonym: "beta-methylasparagine" RELATED RESID-misnomer []
synonym: "Methyl" RELATED PSI-MS-label []
synonym: "Methylation" RELATED UniMod-description []
synonym: "methyln" EXACT OMSSA-label []
synonym: "MOD_RES N4-methylasparagine" EXACT UniProt-feature []
synonym: "N(gamma)-methylasparagine" EXACT RESID-alternate []
synonym: "N-methylasparagine" EXACT RESID-alternate []
synonym: "N4-methyl-L-asparagine" EXACT RESID-name []
synonym: "N4-methylated L-asparagine" EXACT PSI-MOD-alternate []
synonym: "N4Me1Asn" EXACT PSI-MOD-label []
xref: DiffAvg: "14.03"
xref: DiffFormula: "C 1 H 2 N 0 O 0"
xref: DiffMono: "14.015650"
xref: Formula: "C 5 H 8 N 2 O 2"
xref: MassAvg: "128.13"
xref: MassMono: "128.058578"
xref: Origin: "N"
xref: Source: "natural"
xref: TermSpec: "none"
is_a: MOD:00599 ! monomethylated residue
is_a: MOD:00602 ! N-methylated residue
is_a: MOD:00673 ! methylated asparagine

id: MOD:01965
name: 2xC(13),3x(2)H labeled N6-acetyl-L-lysine
def: "A protein modification that effectively converts an L-lysine residue to 2xC(13),3x(2)H labeled N6-acetyl-L-lysine." [PubMed:18688235]
synonym: "Acetate labeling reagent (K) (heavy form, +5amu)" EXACT PSI-MOD-alternate []
synonym: "COFRADIC heavy acetyl 13C2 2H3" EXACT PSI-MOD-alternate []
xref: DiffAvg: "47.04"
xref: DiffFormula: "(13)C 2 (1)H -1 (2)H 3 O 1"
xref: DiffMono: "47.036105"
xref: Formula: "(12)C 6 (13)C 2 (1)H 11 (2)H 3 N 2 O 2"
xref: MassAvg: "175.13"
xref: MassMono: "175.131068"
xref: Origin: "K"
xref: Source: "artifact"
xref: TermSpec: "none"
is_a: MOD:00064 ! N6-acetyl-L-lysine
is_a: MOD:01428 ! (13)C isotope tagged reagent
is_a: MOD:01431 ! (2)H deuterium tagged reagent

Other cases

- Replace comment: Crosslinks n by a precise term to define the number of modified residues

[Term]
id: MOD:00046
name: O-phospho-L-serine
def: "A protein modification that effectively converts an L-serine residue to O-phospho-L-serine." [ChEBI:15811, DeltaMass:0, OMSSA:6, PubMed:12923550, PubMed:4065410, PubMed:8061611, RESID:AA0037, ~~UniMod:21#S~~]
subset: PSI-MOD-slim
synonym: "(2S)-2-amino-3-(phosphonooxy)propanoic acid" EXACT RESID-systematic []
synonym: "2-amino-3-hydroxypropanoic acid 3-phosphate" EXACT RESID-alternate []
synonym: "2-azanyl-3-(phosphonooxy)propanoic acid" EXACT RESID-alternate []
synonym: "ACT_SITE Phosphoserine intermediate" EXACT UniProt-feature []
synonym: "MOD_RES Phosphoserine" EXACT UniProt-feature []
synonym: "O-phospho-L-serine" EXACT RESID-name []
synonym: "O-phosphonoserine" EXACT RESID-alternate []
synonym: "O-phosphorylated L-serine" EXACT PSI-MOD-alternate []
synonym: "O3-phosphoserine" EXACT RESID-alternate []
synonym: "OPSer" EXACT PSI-MOD-label []
synonym: "Phospho" RELATED PSI-MS-label []
synonym: "Phospho Seryl" EXACT DeltaMass-label []
synonym: "Phosphorylation" RELATED UniMod-description []
synonym: "phosphorylations" EXACT OMSSA-label []
synonym: "serine phosphate ester" EXACT RESID-alternate []

xref: DiffAvg: "79.98"
xref: DiffFormula: "C 0 H 1 N 0 O 3 P 1"
xref: DiffMono: "79.966331"
xref: Formula: "C 3 H 6 N 1 O 5 P 1"
xref: MassAvg: "167.06"
xref: MassMono: "166.998359"
xref: Origin: "S"
xref: Source: "natural"
xref: TermSpec: "none"
is_a: MOD:00771 ! residues isobaric at 166.98-167.00 Da
is_a: MOD:00916 ! modified L-serine residue
is_a: MOD:01455 ! O-phosphorylated residue

Double/multiple modification

- Allow for separate steps (useful for biological pathways)
- Allow for direct double/multiple conversion (useful for MS applications)

- Use Source : «MOD:nnnnn»
- Use relationship in addition to is_a

Entry example for direct double modification

[Term]
id: MOD:01609
name: 7'-hydroxy-2'-alpha-mannosyl-L-tryptophan (Trp)
def: "A protein modification that effectively converts an L-tryptophan residue to 7'-hydroxy-2'-alpha-mannosyl-L-tryptophan." [PubMed:19584055, RESID:AA0506]
synonym: "(2S)-2-amino-3-[7-hydroxy-2-(alpha-D-mannopyranosyl)-1H-indol-3-yl]propanoic acid" EXACT RESID-systematic []
synonym: "7'-hydroxy-2'-alpha-mannosyl-L-tryptophan" EXACT RESID-name []
synonym: "CARBOHYD C-linked (Man) hydroxytryptophan" EXACT UniProt-feature []
xref: DiffAvg: "178.14"
xref: DiffFormula: "C 6 H 10 N 0 O 6"
xref: DiffMono: "178.047738"
xref: Formula: "C 17 H 20 N 2 O 7"
xref: MassAvg: "364.35"
xref: MassMono: "364.127051"
xref: Origin: "W"
xref: Source: "natural"
xref: TermSpec: "none"
xref: uniprot.ptm:PTM-0504
is_a: MOD:00918 ! modified L-tryptophan residue
relationship: has_functional_parent MOD:00222 ! 2'-alpha-mannosyl-L-tryptophan
relationship: has_functional_parent MOD:01664 ! 7'-hydroxy-L-tryptophan

Entry example for second step modification

[Term]
id: MOD:01609
name: 7'-hydroxy-2'-alpha-mannosyl-L-tryptophan
def: "A protein modification that effectively converts an 7'-hydroxy-L-tryptophan residue to 7'-hydroxy-2'-alpha-mannosyl-L-tryptophan." [PubMed:123456, RESID:nnnnnn]
synonym: "(2S)-2-amino-3-[7-hydroxy-2-(alpha-D-mannopyranosyl)-1H-indol-3-yl]propanoic acid" EXACT RESID-systematic []
synonym: "7'-hydroxy-2'-alpha-mannosyl-L-tryptophan" EXACT RESID-name []
synonym: "CARBOHYD C-linked (Man) hydroxytryptophan" EXACT UniProt-feature []
xref: DiffAvg: «nnn.nn"
xref: DiffFormula: "C 6 H 10 N 0 O 5"
xref: DiffMono: «nnn.nnnnnn"
xref: Formula: "C 17 H 20 N 2 O 6"
xref: MassAvg: «nnn.nn"
xref: MassMono: «nnn.nnnnnn"
xref: Origin: " MOD:01664"
xref: Source: "natural"
xref: TermSpec: "none"
xref: uniprot.ptm:PTM-0504 ___???
is_a: MOD:00918 ! modified L-tryptophan residue
is_a: MOD:NNNNN ! Monohydroxylated L-tryptophan
is_a: MOD:NNNNN ! Manosylated residue
relationship: has_functional_parent MOD:01664 ! 7'-hydroxy-L-tryptophan

Structure OBO file

- Xrefs to Unimod, Resid, OMSSA
- Ambiguities Unimod 1 -> many : Unimod xref + # + residue (220#A)
- No free AA, no Unimod unapproved mods and xrefs
- Formula alphabetical, isotope full or no (13)C 2 (12)C 2 or (13)C 2 C 2 ?

To Do

- Remove all UniMod references in def lines (redundant with xref: Unimod and Source information; check if Xref Unimod is present first
- Add descriptors for references in def (Pubmed, RESID, ChEBI(?), OMSSA, others)
- Decide order and format of isotopes in Formulae
- Clarify difference between PSI-MOD-alternate and PSI-MOD-label (in synonyms): from Unimod term PSI-MOD-label?
- Get updates of ptmlist via a 2monthly check for new versions on the uniprot ftp; primary entry point is the PTM-nnnn identifier (xref in PSI-MOD), propagate names to the synonym
- Propose a mapping process for Unimod ; check delta masses identities from Unimod with PSI-MOD values
- Verify referencces in def from 2008 on
- Add mechanism to introduce smiles as xref
- Mechanism for double/multiple steps mods: allow each step AND direct (add a (multiple steps) in def?)
- Incorporate explicit ChEBI refs and synonyms, residues entries in ChEBI are exact synonyms, AA are RELATED (or NARROW?)
- Address GitHub issues
- Update the WG charter
- Create the 400 combinatorial AA->AA entries