



Dept. of Computer Science and Engineering
University of Rajshahi
www.ru.ac.bd

Dr. Shamim Ahmad

PTM Database

Example: Hemoglobin

UniProtKB results

UniProtKB consists of two sections:

- Reviewed (Swiss-Prot) - Manually annotated**
Records with information extracted from literature and curator-evaluated computational analysis.
- Unreviewed (TrEMBL) - Computationally analyzed**
Records that await full manual annotation.

The UniProt Knowledgebase (UniProtKB) is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation. In addition to capturing the core data mandatory for each UniProtKB entry (namely, the amino acid sequence, protein name or description, taxonomic data and status information), as much annotation information as possible is added.

Filter by:

Entry name	Protein name	Gene name	Organism	Length
P19892	HBG2_HUMAN	Hemoglobin subunit gamma-2	Homo sapiens (Human)	147
P19891	HBG1_HUMAN	Hemoglobin subunit gamma-1	Homo sapiens (Human)	147
P12088	HBG1_MOUSE	Hemoglobin subunit beta-1	Mus musculus (Mouse)	147
P12091	HBG1_RAT	Hemoglobin subunit beta-1	Rattus norvegicus (Rat)	147
P12089	HBG2_MOUSE	Hemoglobin subunit beta-2	Mus musculus (Mouse)	147

UniProtKB - P69891 (HBG1_HUMAN)

Display

Protein: Hemoglobin subunit gamma-1

Gene: HBG1

Organism: Homo sapiens (Human)

Status: Reviewed - Annotation score: ★★★★★ - Experimental evidence at protein level

Function:

Gamma chains make up the fetal hemoglobin F, in combination with alpha chains.

Caution:

The modification form of Lys 142 is subject of controversy and could be the artefactual result of sample handling. [View population...](#)

Sites:

UniProt entry for Hemoglobin subunit gamma-1 (P69891). The 'Subcellular location' section shows a diagram of a cell with the cytosol highlighted in yellow. The 'GO - Cellular component' section lists 'cytosol' as a location, with a source of 'BiochemProt'.

UniProt entry for Hemoglobin subunit gamma-1 (P69891). The 'GO - Cellular component' section lists 'cytosol' and 'hemoglobin complex' as locations, with sources 'BiochemProt' and 'GO_Central' respectively. A tooltip for 'hemoglobin complex' explains that this subsection of the PTM / Processing section specifies the position and type of each modified residue.

UniProt entry for Hemoglobin subunit gamma-1 (P69891). The 'PTM / Processing' section includes a table for 'Molecule processing' and 'Amino acid modifications'. The 'Molecule processing' table shows a single entry for 'Chain' with a length of 146. The 'Amino acid modifications' table lists 146 modifications, including phosphorylation at various positions.

Position(s)	Description	Action	Display as view	Length
2-147	Hemoglobin subunit gamma-1	↳ A01 BLAST		146
2	N-acetylserine in form N6-PT	⊞ By similarity		1
13	Phosphoserine	⊞ Combined sources		1
49	Phosphoserine	⊞ Combined sources		1
53	Phosphoserine	⊞ Combined sources		1
60	N6-acetyllysine	⊞ By similarity		1
63	N6-acetyllysine	⊞ By similarity		1
94	S-nitrosocysteine	⊞ By similarity		1
140	Phosphoserine	⊞ Combined sources		1

UniProt entry for Hemoglobin subunit gamma-1 (P69891). The 'PTM / Processing' section includes a table for 'Molecule processing' and 'Amino acid modifications'. A tooltip for 'Amino acid modifications' explains that this subsection of the PTM / Processing section specifies the position and type of each modified residue, excluding both glycosylated and postion cross-linked residues.

Position(s)	Description	Action	Display as view	Length
2-147	Hemoglobin subunit gamma-1	↳ A01 BLAST		146
2	N-acetylserine in form N6-PT	⊞ By similarity		1
13	Phosphoserine	⊞ Combined sources		1
49	Phosphoserine	⊞ Combined sources		1
53	Phosphoserine	⊞ Combined sources		1
60	N6-acetyllysine	⊞ By similarity		1
63	N6-acetyllysine	⊞ By similarity		1
94	S-nitrosocysteine	⊞ By similarity		1
140	Phosphoserine	⊞ Combined sources		1

UniProt

Modified residue

This subsection of the PTM (Post-translational modification) section specifies the position and type of each modified residue excluding lipids, glycans and protein cross-links.

Common modifications include phosphorylation, methylation, acetylation, amidation, formation of pyroglutamate, carboxylation, isomerization, hydroxylation, sulfation, flavin-binding, cysteine oxidation and nitrosylation.

We describe the chemical nature of the modified residue using a controlled vocabulary (see the document [Controlled vocabulary of post-translational modifications \(PTM\)](#)).

We provide additional information concerning the modification, such as:

1. the **form of the protein** which undergoes the modification. This may be either a specific **isoform**, a particular **processed or modified form** of the protein, or a specific **sequence variant**.
Examples: P16150 (pathogen), P04710 (processed form), P04871 (sequence variant).
2. the **enzyme** which carries out the modification (by [EC](#)). For proteins of infectious organisms, such as viruses, phages and bacteria, we also indicate whether the modification is carried out by a **host protein**.
Example: P02279 (proliferating cell nuclear antigen) (viral protein modified by host protein).
3. information on the **frequency** of the modification or the **relationship with another feature** ('partial', 'alternate', 'transient'). The term '**partial**' indicates that not all protein molecules are modified, '**alternate**' means that the same amino acid can be modified in more than one site, and '**transient**' is applied to exceptional cases of otherwise stable modifications. For partial modifications, we do not propagate this comment to homologous proteins and we do not specify the fraction of proteins modified, as this may depend on the experimental conditions.
Examples: P16150 (partial modification), Q01653 (alternate modification), Q01654 (transient modification).

Unknown sites bearing unknown modifications

Here is an example of a feature where the identity of the amino acid is unknown (an X is shown at this position in the sequence) and the only information concerning the modification is that the N-terminus is blocked (P00707 (Blocked amino end (X))).

1. Phosphorylation

Phosphorylation refers to the transfer of a gamma phosphate to an amino acid. It is a key mechanism for signaling in both eukaryotic and prokaryotic cells. It can occur on a number of cytoplasmic and nuclear residues, i.e. on the hydroxyl group of serine, threonine or tyrosine, on the nitrogen of arginine, histidine or lysine, on the carboxyl group of aspartate, or on the sulphydryl group of cysteine.

Related keyword: Phosphoprotein

Phosphorylation is frequent on serine, threonine, and tyrosine from eukaryotic proteins, serine phosphorylation being the most common. Phosphorylation of histidine and aspartate is known to occur as part of the **two-component signaling** in prokaryotes and has also been described in eukaryotes (mainly fungi and plants).
Example: Q01550

Since phosphorylation (phosphoserine, phosphothreonine and phosphotyrosine) is a reversible modification, phosphorylation sites are never annotated as 'partial'!
Example: Q01629, P03682, P04689, P06433

Histidine can be phosphorylated on either of its two nitrogen atoms. We refer to this phosphorylation, when phosphorylation occurs on the nitrogen atom that is closest to the alpha-carbon and 'Ser-phosphothreonine', when it occurs on the most distal one. When the exact position of phosphate attachment on histidine is not known we simply use the term 'Triphosphohistidine'.
Examples: P16400, P10900, P16076, P02102, P04623

Note that phosphorylation of aspartate follows a specific syntax:
Example: P16482

We annotate experimentally determined phosphorylation sites and transfer this information to related isoforms as described. We do not annotate predicted phosphorylation sites. When transferring information regarding phosphorylation, we do not usually specify the kinase responsible as the orthologous entry, except when the modification is part of a precise, well-studied transduction pathway. For example: Q01653, P01653, P01654 for the phosphorylation of histidine.

Modified residue

unknown regarding phosphorylation, we do not usually specify the kinase responsible as the orthologous entry, except when the modification is part of a precise, well-studied transduction pathway. For example: Q01653, P01653, P01654 for the phosphorylation of histidine.

2. Methylation

Cytoplasmic and nuclear proteins can be enzymatically modified in several ways by the addition of methyl groups from S-adenosylmethionine. Methylation reactions occurring on carboxyl groups can be reversible and modulate the activity of the target protein, while those on nitrogen atoms at the N-terminus and on side-chains are usually irreversible.

Related keyword: Methylation

Carboxyl methylation

Carboxyl methylation can occur either on a C-terminal cysteine, leucine or lysine residue, or on the side chain of a glutamate residue (or glutamine, after deamidation). It can affect protein-protein interactions and protein function.
Examples: P17775, P02694

Cysteine carboxymethylation frequently occurs after proteolysis of the CMT (Cys - alpha, beta - any residue) sequence and proteolytic cleavage of the C-A bond.
Example: P14888

In prokaryotes, glutamate methyl ester formation plays a major role in chemotactic signal transduction. We indicate whether the glutamate methyl ester is formed either from glutamate or from glutamine.
Example: P01616

Nitrogen methylation

Nitrogen methylation can occur on the N-terminus of a polypeptide chain or on a phenylalanine, isoleucine, leucine, methionine, lysine, proline or alanine residue. It can also occur on the side chain of lysine, arginine or histidine residues. In eukaryotes, arginine and lysine methylation have been found mainly on histones and play an important role in signal transduction processes, nuclear transport and regulation of transcription.
Examples: P04706, P08973

UniProt

Display

PEPTIDYL LYSYLASE [EC:3.4.21.10] (P06801)

There are 2 potential isoforms mapped to this entry. [View](#) | [Help](#) | [Show all 6 data tables](#)

Entry	Entry name	Protein name	Date names	Length	Annotations
ESP8094	ESP8094_HUMAN	Hemoglobin subunit gamma-2	HBG2	80	Annotation score: ●●○○○
AA047943	AA047943_HUMAN	Hemoglobin subunit gamma-1	HBG1 HBG1	30	Annotation score: ●○○○○

Natural variant

Position	Disruption	Alters	Origin of life	Length
14 - G → M	Macdonald	1 Publication	Corresponds to variant	1
16 - R → S	Macdonald	1 Publication	Corresponds to variant	1
16 - E → K	Beck	3 Publications	Corresponds to variant	1
16 - G → E	Beck	3 Publications	Corresponds to variant	1
17 - E → G	Beck	2 Publications	Corresponds to variant	1
17 - E → Q	Beck	1 Publication	Corresponds to variant	1
17 - T → R	Beck	1 Publication	Corresponds to variant	1

Uniprot: Hemoglobin subunit gamma 1

Display

Natural variant: [2] A → T in Barents
 UniProt: P69891
 Natural variant: 132, N → E in Barents
 UniProt: P69891

Feature viewer

Feature table

Sequence databases

Select the link: 891035 Genomic DNA Translation: AA034293.1
 GenBank: 891937 Genomic DNA Translation: AA034953.1
 RefSeq: U05513 Genomic DNA Translation: CA023771.1
 GenBank: U05514 Genomic DNA Translation: CA023772.1
 CCDB: J08776 Genomic DNA Translation: AA032653.1
 UniProt: U01117 Genomic DNA Translation: AA016322.1
 AF130068 mRNA Translation: AA032623.1
 CH471954 Genomic DNA Translation: G4094984.1
 BC023993 mRNA Translation: AA110913.1
 BC026715 mRNA Translation: AA120719.1
 AF481523 Genomic DNA Translation: AA139545.1
 CCDS: CCDS37754.1
 RefSeq: R43602.004664
 RefSeq: NP_069550.2, NM_069550.2
 UniGene: U1.782189

Genome annotation databases

Ensembl: ENSLTO00000209897:ENSP00000274761:ENST00000212804
 GenBank: 3067

Feature table (left sidebar):

- Function
- Names & Synonyms
- Coordinates & Location
- Pathology & Biotech
- PTM / Post-translational
- Expression
- Interaction
- Protein levels
- Gene & Transcripts
- Sequence variants (17)
- Similar proteins

Uniprot: Hemoglobin subunit gamma 1

Display

Keywords - Coding sequence diversity: Polymorphism

Feature viewer

Feature table

Similar proteins

Protein	Similar protein	Species	Score	Length	Source
P69891	Hemoglobin subunit gamma 1	HUMAN	●●●●●	147	Uniprot_P69891
	UF000011007	D		148	
	UF000004972	D		92	

Cross-references

Web resources

Human hemoglobin variants and thalassemias

Sequence databases

Select the link: 891035 Genomic DNA Translation: AA034293.1
 GenBank: 891937 Genomic DNA Translation: AA034953.1
 RefSeq: U05513 Genomic DNA Translation: CA023771.1

Feature table (left sidebar):

- Function
- Names & Synonyms
- Coordinates & Location
- Pathology & Biotech
- PTM / Post-translational
- Expression
- Interaction
- Protein levels
- Gene & Transcripts
- Sequence variants (17)
- Similar proteins