

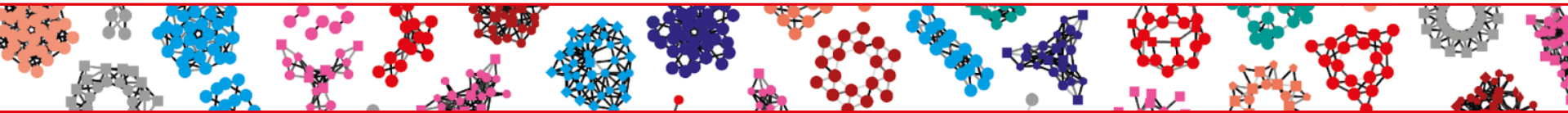
Swiss Institute of  
Bioinformatics

# SPARQLing with GlyConnect

Julien Mariethoz, Proteome Informatics group

Jerven Bolleman, Swiss-Prot group

# Overview



01

• **Biological context**

02

• GlyConnect

03

• Data model

04

• Query examples

# Biological context

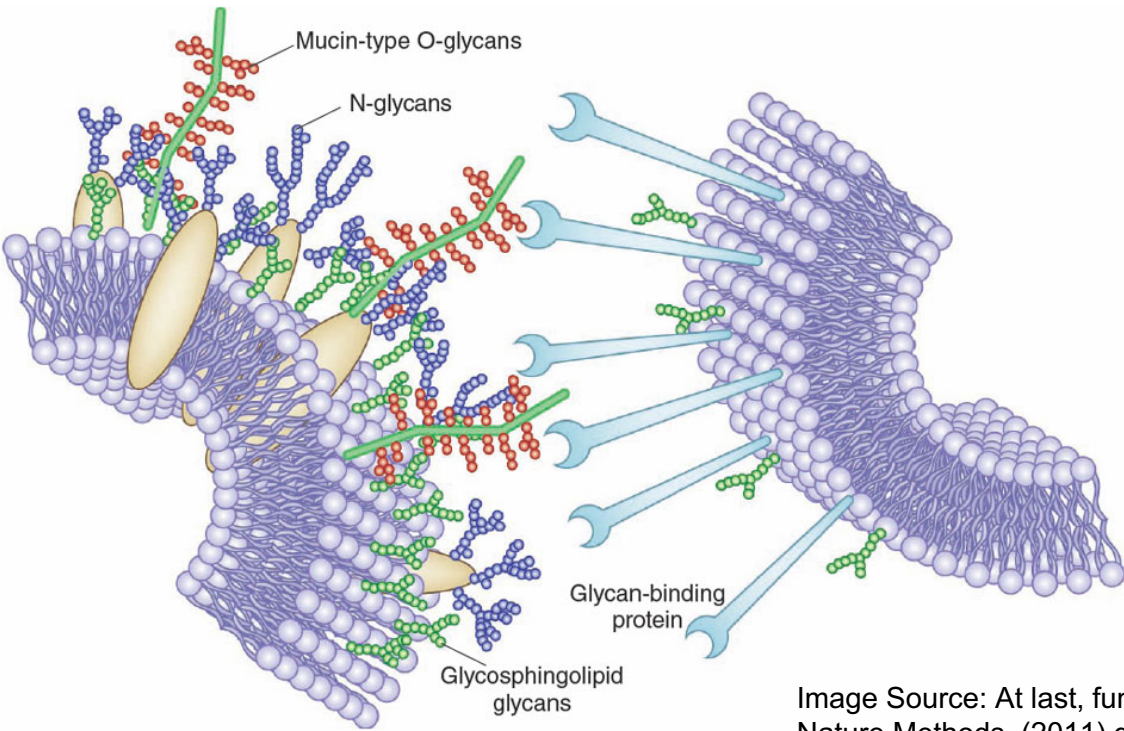
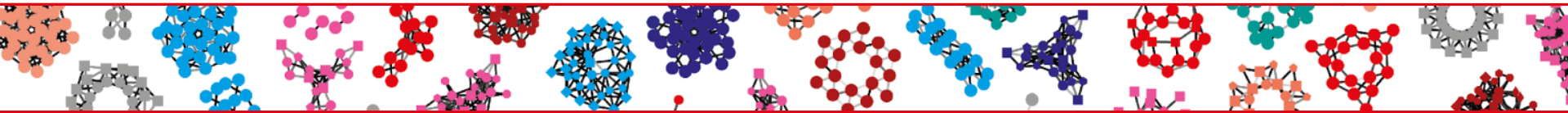


Image Source: At last, functional glycomics.  
Nature Methods (2011) doi:10.1038/nmeth0111-55

# Overview



01

• Biological context

02

• **GlyConnect**

03

• Data model

04

• Query examples

# GlycoProtein data in GlyConnect



1. Glycan (structure or composition) : GlyTouCan
2. Protein : UniProt
3. Glycosylation (site)
4. Reference : Pubmed / doi



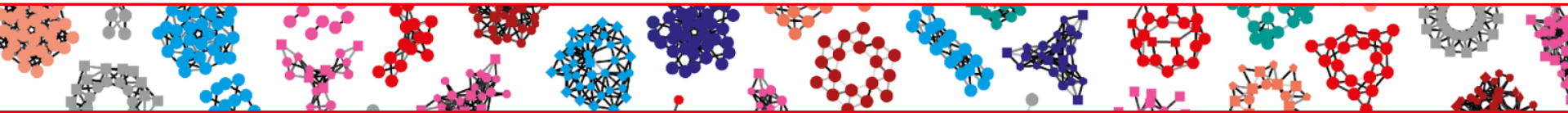
+ Contextual data : tissues, disease, peptides, ...



## GlycoProtein data in GlyConnect - numbers

	<b>Total</b>	<b>Human</b>	<b>RDFized</b>
<b>Glycan (GlyTouCan)</b> 	<b>3'970</b>	<b>2'107</b>	<b>1'134</b>
<b>Protein (UniProt)</b> 	<b>2'631</b>	<b>1'334</b>	<b>2'014</b>
<b>Glycosylation</b>	<b>30'045</b>	<b>19'928</b>	<b>22'146</b>
<b>Reference (Pubmed / doi)</b>	<b>897</b>	<b>382</b>	<b>860</b>

# Overview



01

• Biological context

02

• GlyConnect

03

• **Data model**

04

• Query examples

# Common data model for glyco-conjugate

---

Common model to several glyco related resources :

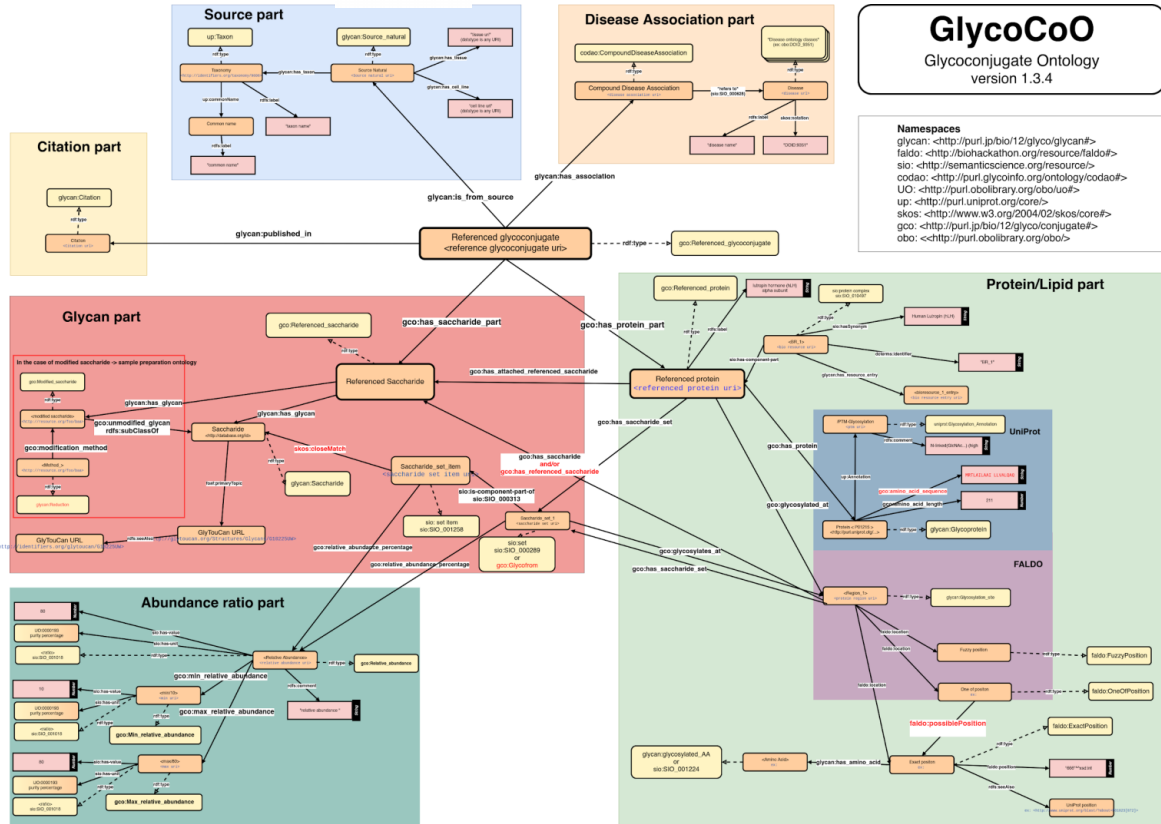
- GlyConnect <http://glyconnect.expasy.org/>
- Glycosmos <https://glycosmos.org/>
- GlycoAbun <https://glyconavi.org/GlycoAbun>
- UnicarbKB <http://unicarbkb.org/>



<https://github.com/glycoinfo/GlycoCoO>



# Data model



**GlycoCoO**  
Glycoconjugate Ontology  
version 1.3.4

**Namespaces**  
glycan: <http://purl.jp/bio/1.2/glycan/>  
aldo: <http://biohackathon.org/resource/aldo/>  
sio: <http://semanticscience.org/resource/>  
codao: <http://purl.glycoinfo.org/ontology/codao/>  
UO: <http://purl.obolibrary.org/obo/uo/>  
up: <http://purl.uniprot.org/core/>  
skos: <http://www.w3.org/2004/02/skos/core/>  
goo: <http://purl.jp/bio/1.2/glycoconjugate/>  
obo: <http://purl.obolibrary.org/obo/>

# GlycoConnect data : glycan structure

<https://glyconnect.expasy.org/browser/structures/3534>



## Structure

Details for this Complex / N-Linked glycan structure

SNFG, Text, Ontol



N-Linked / Complex (avg mass : 2006.8447)

↓ SIFTS list

- Reported glycosite**
- Antipodanin-1 / Homo sapiens**  
UniProt: P18432
- Hemaphysin / Homo sapiens**  
UniProt: P18432
- Human IgE epsilon chain heavy chain H4 / Homo sapiens**  
UniProt: P01577
- Complement factor H / Homo sapiens**  
UniProt: P02746
- Prostaglandin-H2 synthase / Homo sapiens**  
UniProt: P09426
- Leucine-rich alpha 2-glycoprotein / Homo sapiens**  
UniProt: P13029
- Alpha-1-acid glycoprotein 2 / Homo sapiens**  
UniProt: P02746

### Glycan references

GlycoConnect: G361317AL

### Composition

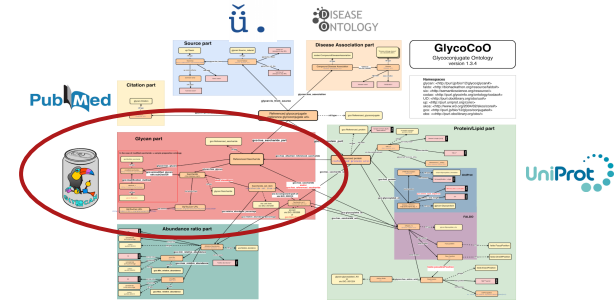
Hex 6 HexNAc 5

### IUPAC

### GlycoCT

```
HEX  
10:10-dGlc-MEX-1:5  
10:10-dAcry:1  
10:10-dGlc-MEX-1:5  
4S:10-dAcry:1  
6B:10-dHex-MEX-1:5  
6B:10-dHex-MEX-1:5  
7B:10-dGlc-MEX-1:5  
8B:10-dGlc-MEX-1:5  
9A:10-dAcry:1  
10B:10-dHex-MEX-1:5  
11B:10-dGlc-MEX-1:5  
12B:10-dAcry:1  
14B:10-dGlc-MEX-1:5  
15B:10-dGlc-MEX-1:5  
16S:10-dAcry:1  
L1N  
1:1.0E12+1.12E  
2:1.0E14+1.10E  
3:1.0E12+1.40  
4:1.0E14+1.10E  
5:1.0E13+1+1.0E  
6:1.0E11+1+1.17E  
7:1.7E(-1+1.1)E  
8:1.7E12+1.0E  
9:1.0E13+1+1.10E  
10:1.0E(-1+1.1)E  
11:1.1E(-1+1.1)E  
12:1.1E(-1+1.1)E  
13:1.1E(-1+1.1)E  
14:1.1E(-1+1.1)E  
15:1.1E(-1+1.1)E  
16:1.1E(-1+1.1)E
```

References (4)



# GlyConnect data : protein

<https://glyconnect.expasy.org/browser/proteins/678>

**Beta-2-glycoprotein 1**  
Taxonomy: Homo sapiens

**Reported glycosite**  
Beta-2-glycoprotein 1 / Homo sapiens  
UniProt

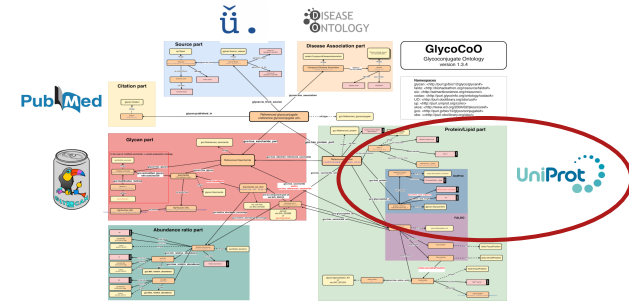
**Protein and gene references**  
UniProtKB: P02749  
RefSeq: NM\_002749  
GeneCards: AP03

**3D structure from PDB**  
Select a PDB entry: NCLZ

**References (9)**  
Large-scale intact glycoprotein identification by Mascot database search (2018)  
Ravi Chand Bollineni, Christian Jeffrey Kozel, Parth Bin Goswami, Jan Haug Aarssen, Bernd Thiele  
PubMed: 29391424 DOI: 10.1038/s41598-018-29331-2  
Status: Unreviewed

Reanalysis of global proteomic and phosphoproteomic data identified a large number of glycoproteins (2018)  
Yingwei Hu, Pankaj Shah, David J. Clark, Minghui An, Hai Zhang  
PubMed: 29749879 DOI: 10.1021/acs.analchem.8b01137  
Status: Unreviewed

Site-Specific Profiling of Serum Glycoproteins Using N-Linked Glycan and Glycosite Analysis Revealing Altered N-glycosylation Sites on Albumin and  $\alpha$ 2B-glycoprotein (2018)  
Shikang Sun, Yingwei Hu, Li Jia, Shadi Toghiani, Ying Liu, Pankaj Shah, Hai Zhang  
PubMed: 29871580 DOI: 10.1021/acs.analchem.8b01581  
Status: Unreviewed



# GlyConnect data : glycosylation site

<https://glyconnect.expasy.org/browser/proteins/678/sites>

GlyConnect Search Browse: Structures Compositions Proteins Tissues Taxonomy Diseases References Help Contact

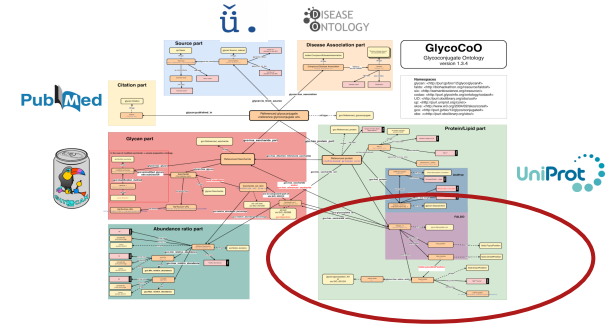
GlyConnect Glycosylation sites

### Beta-2-glycoprotein 1 (Homo sapiens)

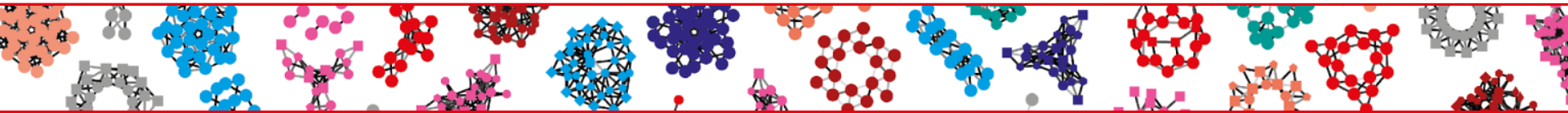
All glycosylation sites

SMU, Text, Output

Site	Core / Type	Cartoon	Composition	Biological associations
Ain-107	N-Linked / Undefined core		Hex5 HexNAc4 NeuAc1	<a href="#">LINKDB (1)</a> <a href="#">PROSITE (1)</a> <a href="#">SOURCE (1)</a> <a href="#">STRUCTURE (1)</a> <a href="#">COMPOSITION (1)</a> <a href="#">PDB (1)</a> <a href="#">REFERENCE (1)</a> <a href="#">SITE (1)</a>
Ain-107	N-Linked / Undefined core		Hex5 HexNAc4 NeuAc2	<a href="#">LINKDB (1)</a> <a href="#">PROSITE (1)</a> <a href="#">SOURCE (1)</a> <a href="#">STRUCTURE (1)</a> <a href="#">COMPOSITION (1)</a> <a href="#">PDB (1)</a> <a href="#">REFERENCE (1)</a> <a href="#">SITE (1)</a>
Ain-107	N-Linked / Undefined core		Hex5 HexNAc4 Hex2 NeuAc1	<a href="#">LINKDB (1)</a> <a href="#">PROSITE (1)</a> <a href="#">SOURCE (1)</a> <a href="#">STRUCTURE (1)</a> <a href="#">COMPOSITION (1)</a> <a href="#">PDB (1)</a> <a href="#">REFERENCE (1)</a> <a href="#">SITE (1)</a>
				<p>Mass spectrometry validated peptide</p> <p>YTTTEPHTNINISNTSPFLNGKDAK (746-756)</p> <p>PROSITE: Ain-107</p> <p>VAR: 038158_1275--N_4630P-V130182</p>
Ain-102	N-Linked / Undefined core		Hex6 HexNAc5 Hex1 NeuAc3	<a href="#">LINKDB (1)</a> <a href="#">PROSITE (1)</a> <a href="#">SOURCE (1)</a> <a href="#">STRUCTURE (1)</a> <a href="#">COMPOSITION (1)</a> <a href="#">PDB (1)</a> <a href="#">REFERENCE (1)</a> <a href="#">SITE (1)</a>
Ain-102	N-Linked / Undefined core		Hex5 HexNAc4 Hex1 NeuAc2	<a href="#">LINKDB (1)</a> <a href="#">PROSITE (1)</a> <a href="#">SOURCE (1)</a> <a href="#">STRUCTURE (1)</a> <a href="#">COMPOSITION (1)</a> <a href="#">PDB (1)</a> <a href="#">REFERENCE (1)</a> <a href="#">SITE (1)</a>
Ain-102	N-Linked / Undefined core		Hex6 HexNAc5 NeuAc3	<a href="#">LINKDB (1)</a> <a href="#">PROSITE (1)</a> <a href="#">SOURCE (1)</a> <a href="#">STRUCTURE (1)</a> <a href="#">COMPOSITION (1)</a> <a href="#">PDB (1)</a> <a href="#">REFERENCE (1)</a> <a href="#">SITE (1)</a>
Ain-102	N-Linked / Undefined core		Hex6 HexNAc5 Hex1 NeuAc1	<a href="#">LINKDB (1)</a> <a href="#">PROSITE (1)</a> <a href="#">SOURCE (1)</a> <a href="#">STRUCTURE (1)</a> <a href="#">COMPOSITION (1)</a> <a href="#">PDB (1)</a> <a href="#">REFERENCE (1)</a> <a href="#">SITE (1)</a>
Ain-102	N-Linked / Undefined core		Hex5 HexNAc4 NeuAc1	<a href="#">LINKDB (1)</a> <a href="#">PROSITE (1)</a> <a href="#">SOURCE (1)</a> <a href="#">STRUCTURE (1)</a> <a href="#">COMPOSITION (1)</a> <a href="#">PDB (1)</a> <a href="#">REFERENCE (1)</a> <a href="#">SITE (1)</a>
Ain-102	N-Linked / Undefined core		Hex6 HexNAc4 Hex1 NeuAc1	<a href="#">LINKDB (1)</a> <a href="#">PROSITE (1)</a> <a href="#">SOURCE (1)</a> <a href="#">STRUCTURE (1)</a> <a href="#">COMPOSITION (1)</a> <a href="#">PDB (1)</a> <a href="#">REFERENCE (1)</a> <a href="#">SITE (1)</a>
Ain-102	N-Linked / Complex		Hex6 HexNAc5	<a href="#">LINKDB (1)</a> <a href="#">PROSITE (1)</a> <a href="#">SOURCE (1)</a> <a href="#">STRUCTURE (1)</a> <a href="#">COMPOSITION (1)</a> <a href="#">PDB (1)</a> <a href="#">REFERENCE (1)</a> <a href="#">SITE (1)</a>
Ain-102	N-Linked / Undefined core		Hex5 HexNAc5 NeuAc1	<a href="#">LINKDB (1)</a> <a href="#">PROSITE (1)</a> <a href="#">SOURCE (1)</a> <a href="#">STRUCTURE (1)</a> <a href="#">COMPOSITION (1)</a> <a href="#">PDB (1)</a> <a href="#">REFERENCE (1)</a> <a href="#">SITE (1)</a>
Ain-	N-Linked / Undefined		Hex6 HexNAc5 NeuAc2	<a href="#">LINKDB (1)</a> <a href="#">PROSITE (1)</a> <a href="#">SOURCE (1)</a> <a href="#">STRUCTURE (1)</a> <a href="#">COMPOSITION (1)</a> <a href="#">PDB (1)</a> <a href="#">REFERENCE (1)</a> <a href="#">SITE (1)</a>



# Overview



01

• GlyConnect

02

• Biological context

03

• Data model

04

• **Query examples**

# GlycoProtein data in GlyConnect

Federated queries with :

- **Glycan databases : use common data model**
- **UniProt : proteins**
- **NextProt : human proteins**
- **Uberon : expression tissues**
- ...

# What are the glycosylations of Beta-2-glycoprotein 1 / P02749?



## Results

### Your query

Add common prefixes

```
1 PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
2 PREFIX faldo:<http://biohackathon.org/resource/faldo#>
3 PREFIX glyco:<http://purl.jp/bio/12/glyco/conjugate#>
4 PREFIX glycan:<http://purl.jp/bio/12/glyco/glycan#>
5 PREFIX foaf:<http://xmlns.com/foaf/0.1/>
6
7 #Select all glycosylations sites of 'Beta-2-glycoprotein 1' (for all species);
8 SELECT DISTINCT ?glycoprotein ?glycoprotein_name ?isoform ?isoform_id ?position ?taxon_name
9 WHERE {
10 ?glycoprotein rdfs:label ?glycoprotein_name .
11 ?glycosite faldo:reference ?isoform .
12 ?glycosite faldo:position ?position .
13 ?specificglycosite faldo:location ?glycosite .
14 ?glycoprotein glyco:glycosylated_at ?specificglycosite .
15 ?refconjugate glyco:has_protein_part ?glycoprotein .
16 ?refconjugate glycan:is_from_source ?source .
17 ?source glycan:has_taxon ?taxon .
18 ?taxon rdfs:label ?taxon_name .
19 BIND(REPLACE(STR(?isoform), "http://purl.uniprot.org/isoforms/", "", "i") AS ?isoform_id)
20 FILTER REGEX(STR(?glycoprotein_name), "Beta-2-glycoprotein 1", "i") .
21 }
22 ORDER BY ?position ?glycoprotein ?isoform
```

Submit Query Cancel



Home Search SPARQL endpoint Browser About Compositions Help Contact

## Results

Sparql XML Spargl JSON CSV Show query Share

glycoprotein	glycoprotein_name	isoform	isoform_id	position	taxon_name
<a href="https://purl.org/glyconnect/referenced_protein/678">https://purl.org/glyconnect/referenced_protein/678</a>	*Beta-2-glycoprotein 1 <sup>read string</sup>	<a href="http://purl.uniprot.org/isoforms/P02749-1">http://purl.uniprot.org/isoforms/P02749-1</a>	*P02749-1 <sup>read string</sup>	*107 <sup>read integer</sup>	*Homo sapiens <sup>read string</sup>
<a href="https://purl.org/glyconnect/referenced_protein/650">https://purl.org/glyconnect/referenced_protein/650</a>	*Beta-2-glycoprotein 1 <sup>read string</sup>	<a href="http://purl.uniprot.org/isoforms/Q01339-1">http://purl.uniprot.org/isoforms/Q01339-1</a>	*Q01339-1 <sup>read string</sup>	*162 <sup>read integer</sup>	*Mus musculus <sup>read string</sup>
<a href="https://purl.org/glyconnect/referenced_protein/678">https://purl.org/glyconnect/referenced_protein/678</a>	*Beta-2-glycoprotein 1 <sup>read string</sup>	<a href="http://purl.uniprot.org/isoforms/P02749-1">http://purl.uniprot.org/isoforms/P02749-1</a>	*P02749-1 <sup>read string</sup>	*162 <sup>read integer</sup>	*Homo sapiens <sup>read string</sup>
<a href="https://purl.org/glyconnect/referenced_protein/678">https://purl.org/glyconnect/referenced_protein/678</a>	*Beta-2-glycoprotein 1 <sup>read string</sup>	<a href="http://purl.uniprot.org/isoforms/P02749-1">http://purl.uniprot.org/isoforms/P02749-1</a>	*P02749-1 <sup>read string</sup>	*183 <sup>read integer</sup>	*Homo sapiens <sup>read string</sup>
<a href="https://purl.org/glyconnect/referenced_protein/678">https://purl.org/glyconnect/referenced_protein/678</a>	*Beta-2-glycoprotein 1 <sup>read string</sup>	<a href="http://purl.uniprot.org/isoforms/P02749-1">http://purl.uniprot.org/isoforms/P02749-1</a>	*P02749-1 <sup>read string</sup>	*253 <sup>read integer</sup>	*Homo sapiens <sup>read string</sup>

# What are the glycosylation sites related to cancer?



Home Search [SPARQL endpoint](#) Browser About Compositions Help Contact

## Results

### Your query

Add common prefixes

```
1 PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
2 PREFIX faldo:<http://biohackathon.org/resource/faldo#>
3 PREFIX glyco:<http://purl.jp/bio/12/glyco/conjugate#>
4 PREFIX glycan:<http://purl.jp/bio/12/glyco/glycan#>
5 PREFIX sio:<http://semanticscience.org/resource/>
6 PREFIX foaf:<http://xmlns.com/foaf/0.1/>
7
8 #Select all glycosylations (glycan, protein, position) with their associated disease where the n
9 SELECT DISTINCT ?isoform ?position ?image ?disease ?disease_name
10 WHERE {
11 ?glycosite faldo:reference ?isoform .
12 ?glycosite faldo:position ?position .
13 ?specific_glycosite faldo:location ?glycosite .
14 ?glycoprotein glyco:glycosylated_at ?specific_glycosite .
15 ?structure glyco:glycosylates_at ?specific_glycosite .
16 ?structure foaf:depiction ?image .
17 ?ref_conjugate glyco:has_protein_part ?glycoprotein .
18 ?ref_conjugate glycan:has_association ?ref_conjugate_disease .
19 ?ref_conjugate_disease sio:SIO_000628 ?disease .
20 ?disease rdfs:label ?disease_name .
21 FILTER REGEX(STR(?disease_name), "cancer", "i") .
22 }
23 ORDER BY ?glycoprotein ?position
```

Submit Query Cancel

## Results

[Sparql XML](#) [Sparql JSON](#) [CSV](#) [Show query](#) [Share](#)

isoform	position	image	disease	disease_name
<a href="https://purl.uniprot.org/isoforms/P01024-1">https://purl.uniprot.org/isoforms/P01024-1</a>	185 <sup>med string</sup>		<a href="https://purl.org/glyconnect/disease/58">https://purl.org/glyconnect/disease/58</a>	"Cancer, breast" <sup>med string</sup>
<a href="https://purl.uniprot.org/isoforms/P01024-1">https://purl.uniprot.org/isoforms/P01024-1</a>	185 <sup>med string</sup>		<a href="https://purl.org/glyconnect/disease/58">https://purl.org/glyconnect/disease/58</a>	"Cancer, breast" <sup>med string</sup>
<a href="https://purl.uniprot.org/isoforms/P01024-1">https://purl.uniprot.org/isoforms/P01024-1</a>	185 <sup>med string</sup>		<a href="https://purl.org/glyconnect/disease/58">https://purl.org/glyconnect/disease/58</a>	"Cancer, breast" <sup>med string</sup>
<a href="https://purl.uniprot.org/isoforms/P01024-1">https://purl.uniprot.org/isoforms/P01024-1</a>	185 <sup>med string</sup>		<a href="https://purl.org/glyconnect/disease/58">https://purl.org/glyconnect/disease/58</a>	"Cancer, breast" <sup>med string</sup>
<a href="https://purl.uniprot.org/isoforms/P01024-1">https://purl.uniprot.org/isoforms/P01024-1</a>	185 <sup>med string</sup>		<a href="https://purl.org/glyconnect/disease/86">https://purl.org/glyconnect/disease/86</a>	"Esophageal cancer" <sup>med string</sup>
<a href="https://purl.uniprot.org/isoforms/P01024-1">https://purl.uniprot.org/isoforms/P01024-1</a>	185 <sup>med string</sup>		<a href="https://purl.org/glyconnect/disease/86">https://purl.org/glyconnect/disease/86</a>	"Esophageal cancer" <sup>med string</sup>
<a href="https://purl.uniprot.org/isoforms/P01024-1">https://purl.uniprot.org/isoforms/P01024-1</a>	185 <sup>med string</sup>		<a href="https://purl.org/glyconnect/disease/86">https://purl.org/glyconnect/disease/86</a>	"Esophageal cancer" <sup>med string</sup>
<a href="https://purl.uniprot.org/isoforms/P01024-1">https://purl.uniprot.org/isoforms/P01024-1</a>	185 <sup>med string</sup>		<a href="https://purl.org/glyconnect/disease/86">https://purl.org/glyconnect/disease/86</a>	"Esophageal cancer" <sup>med string</sup>
<a href="https://purl.uniprot.org/isoforms/P00533-1">https://purl.uniprot.org/isoforms/P00533-1</a>	156 <sup>med string</sup>		<a href="https://purl.org/glyconnect/disease/101">https://purl.org/glyconnect/disease/101</a>	"Prostate cancer" <sup>med string</sup>
<a href="https://purl.uniprot.org/isoforms/P00533-1">https://purl.uniprot.org/isoforms/P00533-1</a>	156 <sup>med string</sup>		<a href="https://purl.org/glyconnect/disease/101">https://purl.org/glyconnect/disease/101</a>	"Prostate cancer" <sup>med string</sup>
<a href="https://purl.uniprot.org/isoforms/P00533-1">https://purl.uniprot.org/isoforms/P00533-1</a>	156 <sup>med string</sup>		<a href="https://purl.org/glyconnect/disease/58">https://purl.org/glyconnect/disease/58</a>	"Cancer, breast" <sup>med string</sup>
<a href="https://purl.uniprot.org/isoforms/P00533-1">https://purl.uniprot.org/isoforms/P00533-1</a>	156 <sup>med string</sup>		<a href="https://purl.org/glyconnect/disease/58">https://purl.org/glyconnect/disease/58</a>	"Cancer, breast" <sup>med string</sup>
<a href="https://purl.uniprot.org/isoforms/P19652-1">https://purl.uniprot.org/isoforms/P19652-1</a>	133 <sup>med string</sup>		<a href="https://purl.org/glyconnect/disease/101">https://purl.org/glyconnect/disease/101</a>	"Prostate cancer" <sup>med string</sup>
<a href="https://purl.uniprot.org/isoforms/P19652-1">https://purl.uniprot.org/isoforms/P19652-1</a>	133 <sup>med string</sup>		<a href="https://purl.org/glyconnect/disease/101">https://purl.org/glyconnect/disease/101</a>	"Prostate cancer" <sup>med string</sup>
<a href="https://purl.uniprot.org/isoforms/P19652-1">https://purl.uniprot.org/isoforms/P19652-1</a>	133 <sup>med string</sup>		<a href="https://purl.org/glyconnect/disease/101">https://purl.org/glyconnect/disease/101</a>	"Prostate cancer" <sup>med string</sup>
<a href="https://purl.uniprot.org/isoforms/P19652-1">https://purl.uniprot.org/isoforms/P19652-1</a>	133 <sup>med string</sup>		<a href="https://purl.org/glyconnect/disease/58">https://purl.org/glyconnect/disease/58</a>	"Cancer, breast" <sup>med string</sup>
<a href="https://purl.uniprot.org/isoforms/P19652-1">https://purl.uniprot.org/isoforms/P19652-1</a>	133 <sup>med string</sup>		<a href="https://purl.org/glyconnect/disease/58">https://purl.org/glyconnect/disease/58</a>	"Cancer, breast" <sup>med string</sup>
<a href="https://purl.uniprot.org/isoforms/P19652-1">https://purl.uniprot.org/isoforms/P19652-1</a>	133 <sup>med string</sup>		<a href="https://purl.org/glyconnect/disease/58">https://purl.org/glyconnect/disease/58</a>	"Cancer, breast" <sup>med string</sup>
<a href="https://purl.uniprot.org/isoforms/P19652-1">https://purl.uniprot.org/isoforms/P19652-1</a>	133 <sup>med string</sup>		<a href="https://purl.org/glyconnect/disease/58">https://purl.org/glyconnect/disease/58</a>	"Cancer, breast" <sup>med string</sup>
<a href="https://purl.uniprot.org/isoforms/P19652-1">https://purl.uniprot.org/isoforms/P19652-1</a>	156 <sup>med string</sup>		<a href="https://purl.org/glyconnect/disease/101">https://purl.org/glyconnect/disease/101</a>	"Prostate cancer" <sup>med string</sup>



# Nextprot variants with NxT pattern of N-Glycosylation

```
6 PREFIX nx: <http://nextprot.org/rdf#>
7
8 SELECT DISTINCT ?glycoprotein ?glycoprotein_name ?isoform ?position ?taxon_name ?entry ?nx_isoform ?nx_isoform_id ?var ?n ?n1 ?n2
9 WHERE{
10   BIND(IRI(REPLACE(STR(?isoform), "http://purl.uniprot.org/isoforms/", "http://nextprot.org/rdf/isoform/NX_")) AS ?isoform)
11   ?glycoprotein rdfs:label ?glycoprotein_name .
12   ?glycosite faldo:reference ?isoform .
13   ?glycosite faldo:position ?position .
14   ?specificglycosite faldo:location ?glycosite .
15   ?glycoprotein glyco:glycosylated_at ?specificglycosite .
16   ?refconjugate glyco:has_protein_part ?glycoprotein .
17   ?refconjugate glycan:is_from_source ?source .
18   ?source glycan:has_taxon ?taxon .
19   ?taxon rdfs:label ?taxon_name .
20
21   FILTER REGEX(STR(?glycoprotein_name), "Beta-2-glycoprotein 1", "i") .
22   SERVICE <https://sparql.nextprot.org/> {
23     ?entry nx:isoform ?nx_isoform. #use isoform from
24     ?entry nx:swissprotPage ?sp.
25     ?nx_isoform nx:variant ?var.
26     ?var nx:variation "N"^^xsd:string. # var -> N
27     ?var nx:start ?position.
28     ?var nx:evidence / nx:quality nx:GOLD. # GOLD annotation
29     ?nx_isoform nx:sequence / nx:chain ?seq.
30     BIND(SUBSTR(STR(?seq), ?position, 1) AS ?n).
31     BIND(SUBSTR(STR(?seq), ?position+1, 1) AS ?n1).
32     BIND(SUBSTR(STR(?seq), ?position+2, 1) AS ?n2).
33     FILTER(
34       (REGEX(?n,"N","i") || REGEX(?n2,"S","i") || REGEX(?n2,"T","i")) # matching pattern NxS or NxT
35       && !REGEX(?n1,"P","i") # matching pattern x!=P
36     ).
37     FILTER( NOT EXISTS { ?entry nx:glycosylationSite ?site} ) . #not referenced as glycosite
38   }
39 }
40 ORDER BY ?position ?glycoprotein ?isoform
```