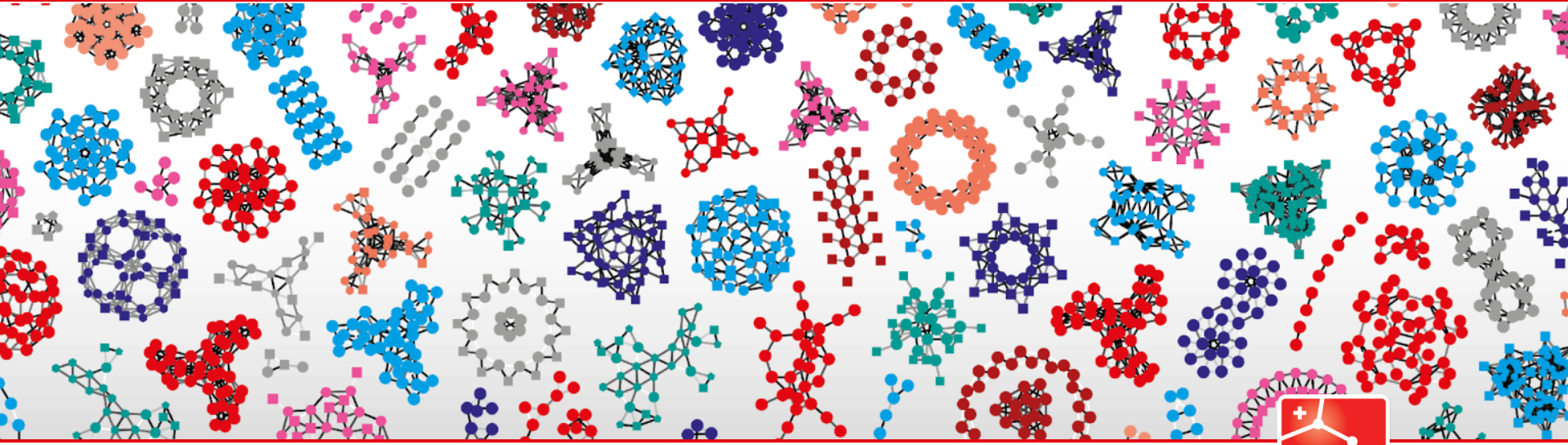


Swiss Institute of Bioinformatics

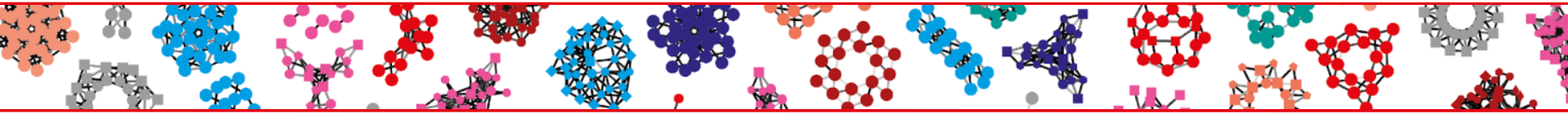
SPARQLing with GlyConnect

Julien Mariethoz, Proteome Informatics group

Jerven Bolleman, Swiss-Prot group

elixir SWITZERLAND

Overview



01 Biological context

02 GlyConnect

03 Data model

04 Query examples

Biological context

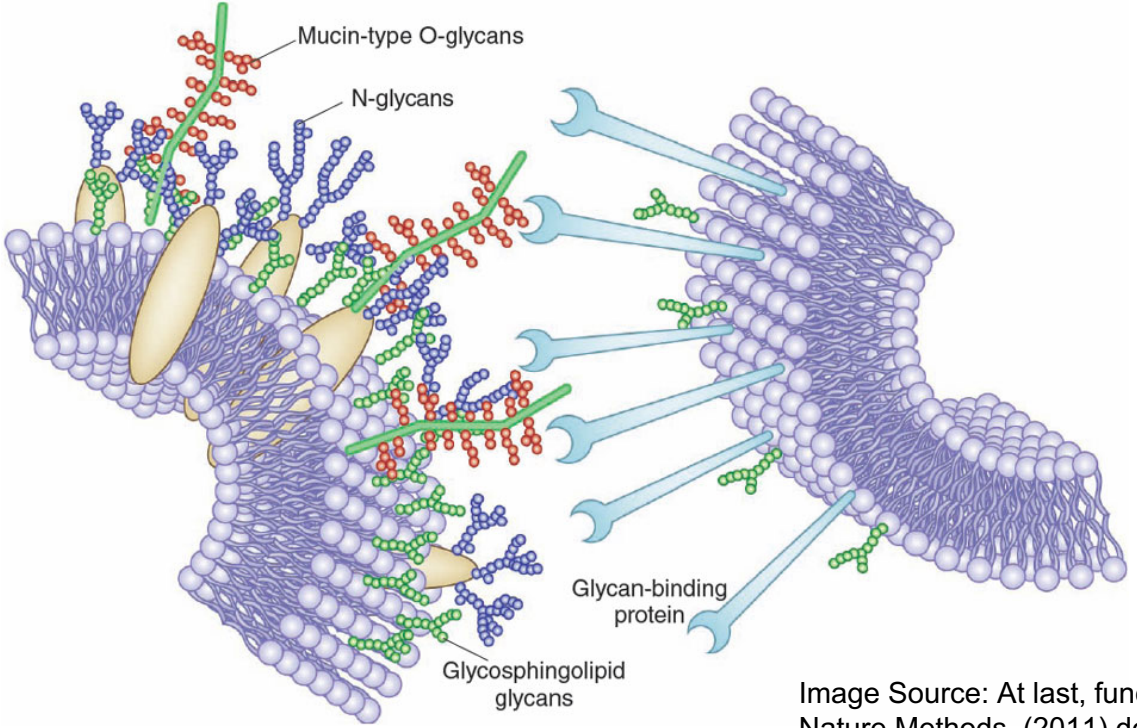
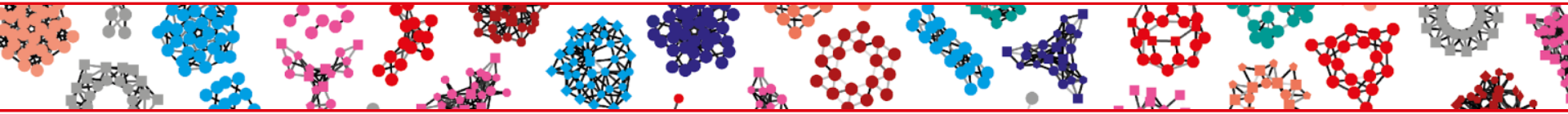


Image Source: At last, functional glycomics.
Nature Methods (2011) doi:10.1038/nmeth0111-55

Overview



01 Biological context

02 GlyConnect

03 Data model

04 Query examples

GlycoProtein data in GlyConnect

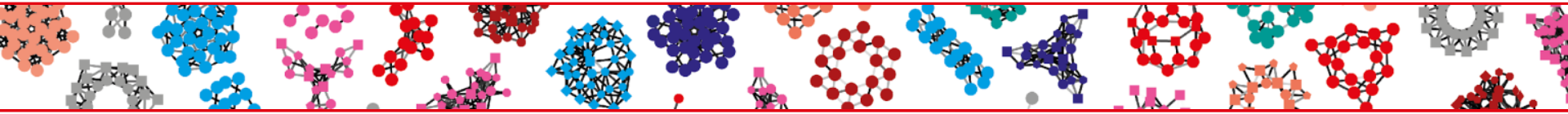
1. Glycan (structure or composition) : GlyTouCan
2. Protein : UniProt
3. Glycosylation (site)
4. Reference : Pubmed / doi

+ Contextual data : tissues, disease, peptides, …

GlycoProtein data in GlyConnect - numbers

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

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 :

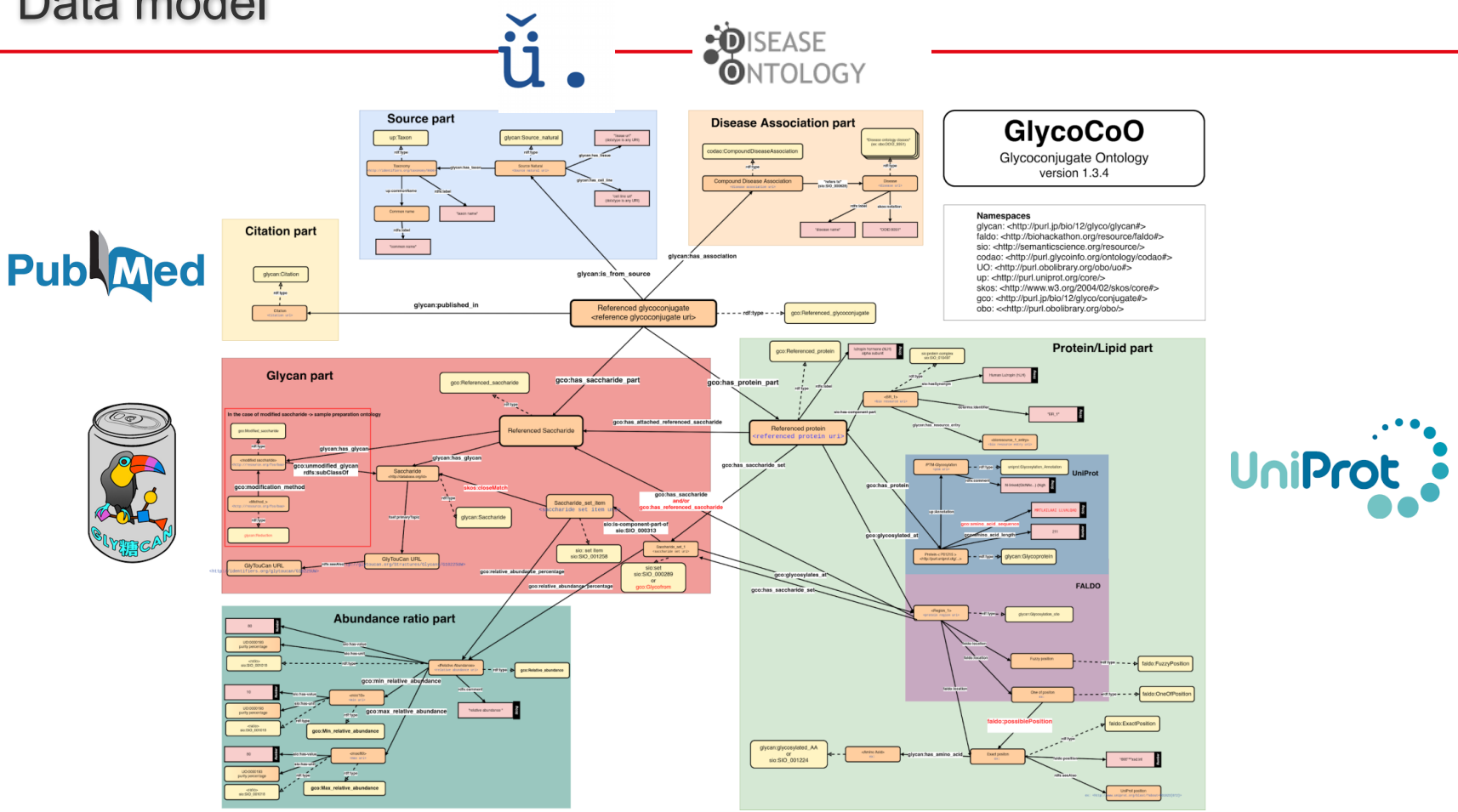
- GlyConnect http://glyconnect.expasy.org/
- Glycosmos https://glycosmos.org/
- GlycoAbun https://glyconavi.org/GlycoAbun
- UnicarbKB http://unicarbkb.org/

}

BioHackathon 2017
in Tokyo/Iwate

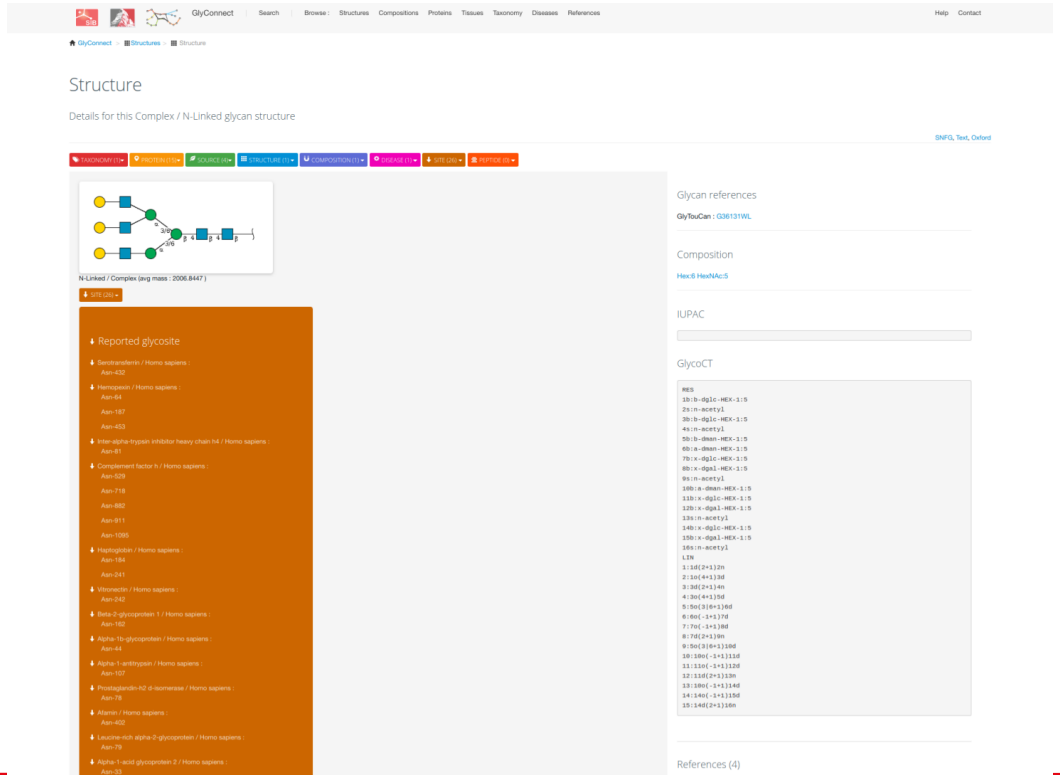
https://github.com/glycoinfo/GlycoCoO

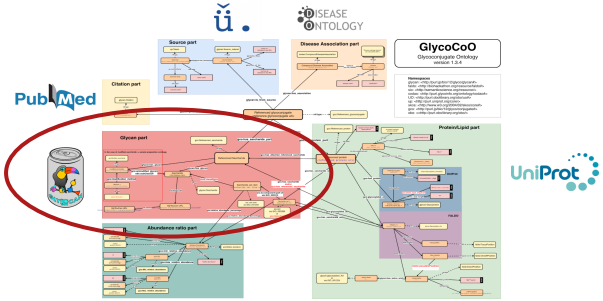
Data model


GlycoCoO
Glycoconjugate Ontology
version 1.3.4
Namespaces
glycan: <http://purl.jp/bio/12/glyco/glycan#>
faldo: <http://biohackathon.org/resource/faldo#>
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#>
gco: <http://purl.jp/bio/12/glyco/conjugate#>
obo: <<http://purl.obolibrary.org/obo/>
Source part
Disease Association part
Citation part
Glycan part
Protein/Lipid part
Abundance ratio part
Referenced glycoconjugate
Referenced Saccharide
Referenced protein
UniProt
FALDO
glycan:is_from_source
glycan:has_association
glycan:published_in
gco:has_saccharide_part
gco:has_protein_part
gco:has_attached_referenced_saccharide
gco:has_saccharide_set
gco:has_protein
gco:glycosylated_at
gco:min_relative_abundance
gco:max_relative_abundance
faldo:possiblePosition

GlyConnect data : glycan structure

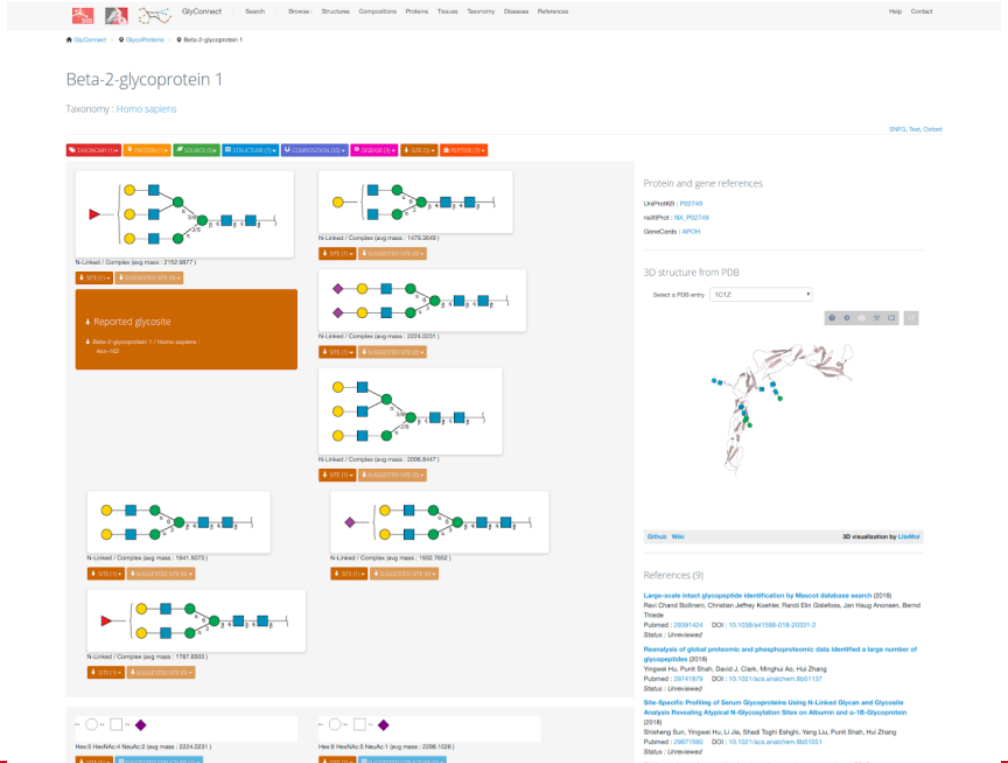
https://glyconnect.expasy.org/browser/structures/3534

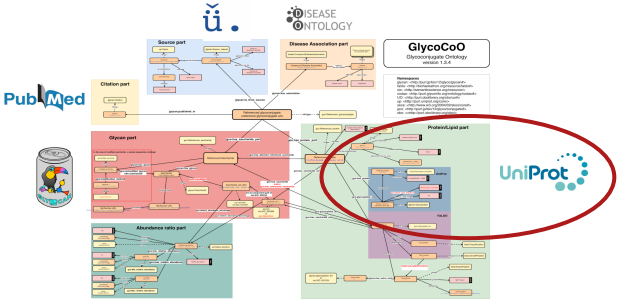




GlyConnect data : protein

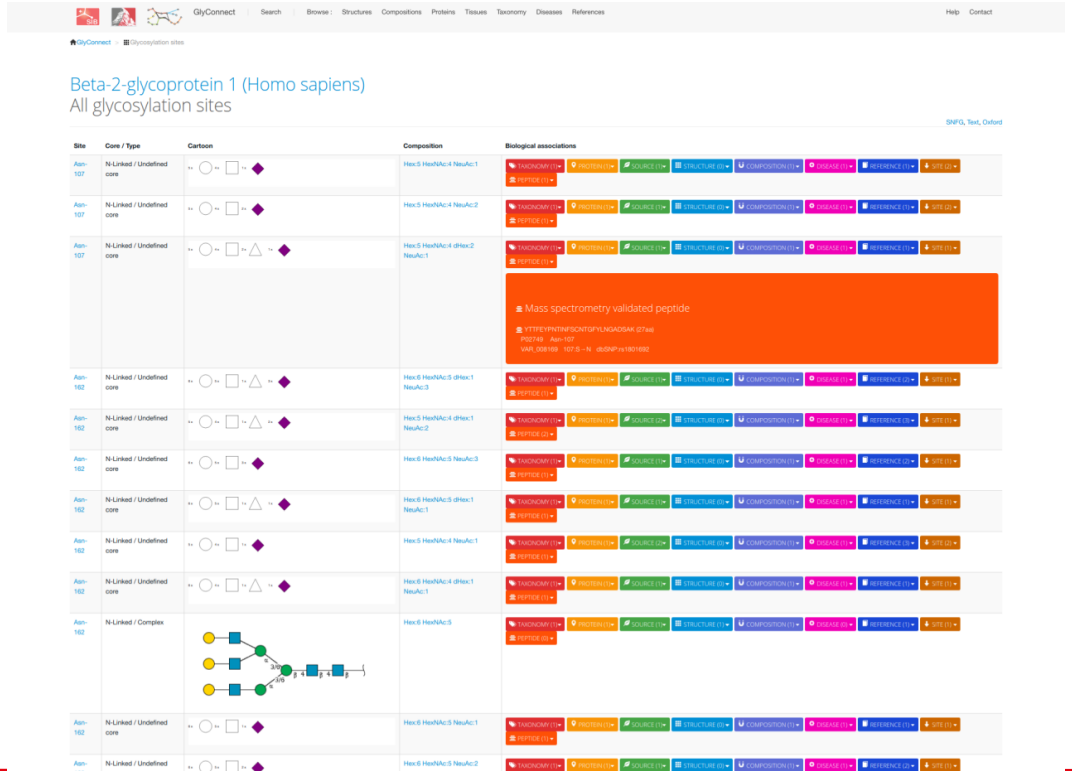
https://glyconnect.expasy.org/browser/proteins/678

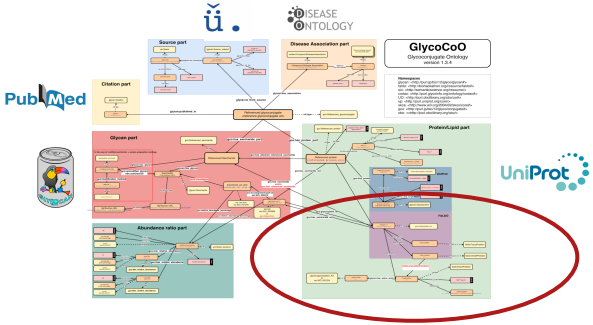




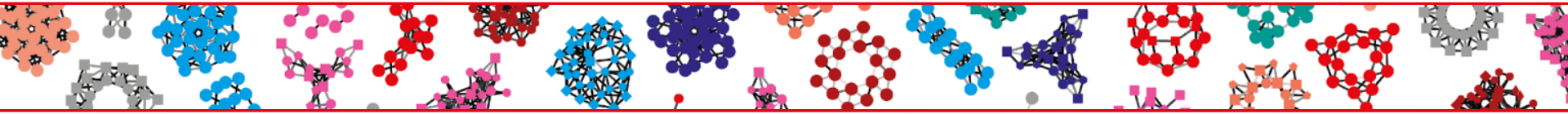
GlyConnect data : glycosylation site

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





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?

GlyConnect

Results

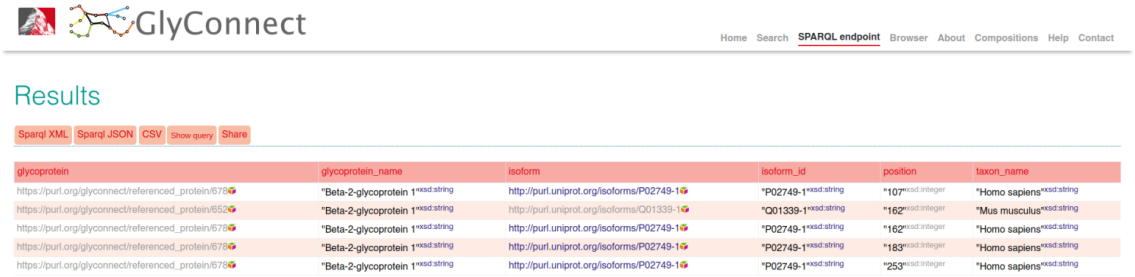
Your query

Add common prefixes

```
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX faldo:<http://biohackathon.org/resource/faldo#>
PREFIX glyco:<http://purl.jp/bio/12/glyco/conjugate#>
PREFIX glycan:<http://purl.jp/bio/12/glyco/glycan#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>

#Select all glycosylations sites of 'Beta-2-glycoprotein 1' (for all species);
SELECT DISTINCT ?glycoprotein ?glycoprotein_name ?isoform ?isoform_id ?position ?taxon_name
WHERE {
?glycoprotein rdfs:label ?glycoprotein_name .
?glycosite faldo:reference ?isoform .
?glycosite faldo:position ?position .
?specificglycosite faldo:location ?glycosite .
?glycoprotein glyco:glycosylated_at ?specificglycosite .
?refconjugate glyco:has_protein_part ?glycoprotein .
?refconjugate glycan:is_from_source ?source .
?source glycan:has_taxon ?taxon .
?taxon rdfs:label ?taxon_name .
BIND(REPLACE(STR(?isoform), "http://purl.uniprot.org/isoforms/", "", "i") AS ?isoform_id)
FILTER REGEX(STR(?glycoprotein_name), "Beta-2-glycoprotein 1", "i") .
}
ORDER BY ?position ?glycoprotein ?isoform
```

Submit Query Cancel



What are the glycosylation sites related to cancer?

GlyConnect

Results

Your query

Add common prefixes

```
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX faldo:<http://biohackathon.org/resource/faldo#>
PREFIX glyco:<http://purl.jp/bio/12/glyco/conjugate#>
PREFIX glycan:<http://purl.jp/bio/12/glyco/glycan#>
PREFIX sio:<http://semanticscience.org/resource/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>

#Select all glycosylations (glycan, protein, position) with their associated disease where the na
SELECT DISTINCT ?isoform ?position ?image ?disease ?disease_name
WHERE {
?glycosite faldo:reference ?isoform .
?glycosite faldo:position ?position .
?specific_glycosite faldo:location ?glycosite .
?glycoprotein glyco:glycosylated_at ?specific_glycosite .
?structure glyco:glycosylates_at ?specific_glycosite .
?structure foaf:depiction ?image .
?ref_conjugate glyco:has_protein_part ?glycoprotein .
?ref_conjugate glycan:has_association ?ref_conjugate_disease .
?ref_conjugate_disease sio:SIO_000628 ?disease .
?disease rdfs:label ?disease_name .
FILTER REGEX(STR(?disease_name), "cancer", "i") .
}
ORDER BY ?glycoprotein ?position
```

Submit Query Cancel

GlyConnect

Home Search SPARQL endpoint Browser About Compositions Help Contact

Results

Sparql XML Sparql JSON CSV Show query Share

isoform	position	image	disease	disease_name
http://purl.uniprot.org/isoforms/P01024-1	"85"xsd:integer		https://purl.org/glyconnect/disease/58	"Cancer, breast"xsd:string
http://purl.uniprot.org/isoforms/P01024-1	"85"xsd:integer		https://purl.org/glyconnect/disease/58	"Cancer, breast"xsd:string
http://purl.uniprot.org/isoforms/P01024-1	"85"xsd:integer		https://purl.org/glyconnect/disease/58	"Cancer, breast"xsd:string
http://purl.uniprot.org/isoforms/P01024-1	"85"xsd:integer		https://purl.org/glyconnect/disease/58	"Cancer, breast"xsd:string
http://purl.uniprot.org/isoforms/P01024-1	"85"xsd:integer		https://purl.org/glyconnect/disease/86	"Esophageal cancer"xsd:string
http://purl.uniprot.org/isoforms/P01024-1	"85"xsd:integer		https://purl.org/glyconnect/disease/86	"Esophageal cancer"xsd:string
http://purl.uniprot.org/isoforms/P01024-1	"85"xsd:integer		https://purl.org/glyconnect/disease/86	"Esophageal cancer"xsd:string
http://purl.uniprot.org/isoforms/P01024-1	"85"xsd:integer		https://purl.org/glyconnect/disease/86	"Esophageal cancer"xsd:string
http://purl.uniprot.org/isoforms/P00533-1	"56"xsd:integer		https://purl.org/glyconnect/disease/101	"Prostate cancer"xsd:string
http://purl.uniprot.org/isoforms/P00533-1	"56"xsd:integer		https://purl.org/glyconnect/disease/101	"Prostate cancer"xsd:string
http://purl.uniprot.org/isoforms/P00533-1	"56"xsd:integer		https://purl.org/glyconnect/disease/58	"Cancer, breast"xsd:string
http://purl.uniprot.org/isoforms/P00533-1	"56"xsd:integer		https://purl.org/glyconnect/disease/58	"Cancer, breast"xsd:string
http://purl.uniprot.org/isoforms/P19652-1	"33"xsd:integer		https://purl.org/glyconnect/disease/101	"Prostate cancer"xsd:string
http://purl.uniprot.org/isoforms/P19652-1	"33"xsd:integer		https://purl.org/glyconnect/disease/101	"Prostate cancer"xsd:string
http://purl.uniprot.org/isoforms/P19652-1	"33"xsd:integer		https://purl.org/glyconnect/disease/101	"Prostate cancer"xsd:string
http://purl.uniprot.org/isoforms/P19652-1	"33"xsd:integer		https://purl.org/glyconnect/disease/101	"Prostate cancer"xsd:string
http://purl.uniprot.org/isoforms/P19652-1	"33"xsd:integer		https://purl.org/glyconnect/disease/58	"Cancer, breast"xsd:string
http://purl.uniprot.org/isoforms/P19652-1	"33"xsd:integer		https://purl.org/glyconnect/disease/58	"Cancer, breast"xsd:string
http://purl.uniprot.org/isoforms/P19652-1	"33"xsd:integer		https://purl.org/glyconnect/disease/58	"Cancer, breast"xsd:string
http://purl.uniprot.org/isoforms/P19652-1	"33"xsd:integer		https://purl.org/glyconnect/disease/58	"Cancer, breast"xsd:string
http://purl.uniprot.org/isoforms/P19652-1	"56"xsd:integer		https://purl.org/glyconnect/disease/101	"Prostate cancer"xsd:string

Nextprot variants with NxT pattern of N-Glycosylation

```
PREFIX nx: <http://nextprot.org/rdf#>

SELECT DISTINCT ?glycoprotein ?glycoprotein_name ?isoform ?position ?taxon_name ?entry ?nx_isoform ?nx_isoform_id ?var ?n ?n1 ?n2
WHERE{
    BIND(IRI(REPLACE(STR(?isoform), "http://purl.uniprot.org/isoforms/", "http://nextprot.org/rdf/isoform/NX_")) AS ?isoform)
    ?glycoprotein rdfs:label ?glycoprotein_name .
    ?glycosite faldo:reference ?isoform .
    ?glycosite faldo:position ?position .
    ?specificglycosite faldo:location ?glycosite .
    ?glycoprotein glyco:glycosylated_at ?specificglycosite .
    ?refconjugate glyco:has_protein_part ?glycoprotein .
    ?refconjugate glycan:is_from_source ?source .
    ?source glycan:has_taxon ?taxon .
    ?taxon rdfs:label ?taxon_name .

    FILTER REGEX(STR(?glycoprotein_name), "Beta-2-glycoprotein 1", "i") .
    SERVICE <https://sparql.nextprot.org/> {
        ?entry nx:isoform ?nx_isoform. #use isoform from
        ?entry nx:swissprotPage ?sp.
        ?nx_isoform nx:variant ?var.
        ?var nx:variation "N"^^xsd:string. # var -> N
        ?var nx:start ?position.
        ?var nx:evidence / nx:quality nx:GOLD. # GOLD annotation
        ?nx_isoform nx:sequence / nx:chain ?seq.
        BIND(SUBSTR(STR(?seq), ?position, 1) AS ?n).
        BIND(SUBSTR(STR(?seq), ?position+1, 1) AS ?n1).
        BIND(SUBSTR(STR(?seq), ?position+2, 1) AS ?n2).
        FILTER(
          (REGEX(?n,"N","i") || REGEX(?n2,"S","i") || REGEX(?n2,"T","i")) # matching pattern NxS or NxT
          && !REGEX(?n1,"P","i") # matching pattern x!=P
        ).
       FILTER( NOT EXISTS { ?entry nx:glycosylationSite ?site}) . #not referenced as glycosite
    }
}
ORDER BY ?position ?glycoprotein ?isoform
```