

Practical introduction to SPARQL for biologists and informaticians

Using the real world UniProt and neXtProt
databases as illustrative examples



Swiss Institute of
Bioinformatics

Jerven Bolleman (Swiss-Prot)
Daniel Teixeira (CALIPHO)
Pierre-André Michel (CALIPHO)
Alain Gateau (CALIPHO)

Start sparql-playground for neXtProt

- Download the zip file:
<https://github.com/calipho-sib/sparql-playground/archive/1.5.0.zip>
- Make sure you have Java 1.7 or Java 1.8
You can check with `java -version`
- Unzip the file and execute the start script:
 - `start-nextprot.bat` for Windows users
 - `start-nextprot.sh` for Unix users
- Open your browser on: <http://localhost:7777>

neXtProt and SPARQL

SPARQL as the advanced search system

- Integrated in UI
- More than 100 sample queries
- Extension of SNORQL as a toolbox for users to work out queries
- Help service
- Persistence of user queries

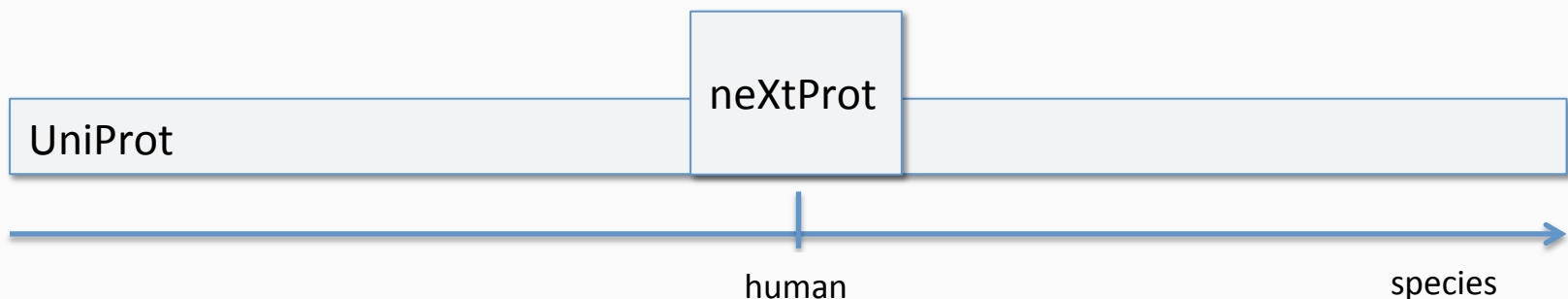
neXtProt content

Specificity

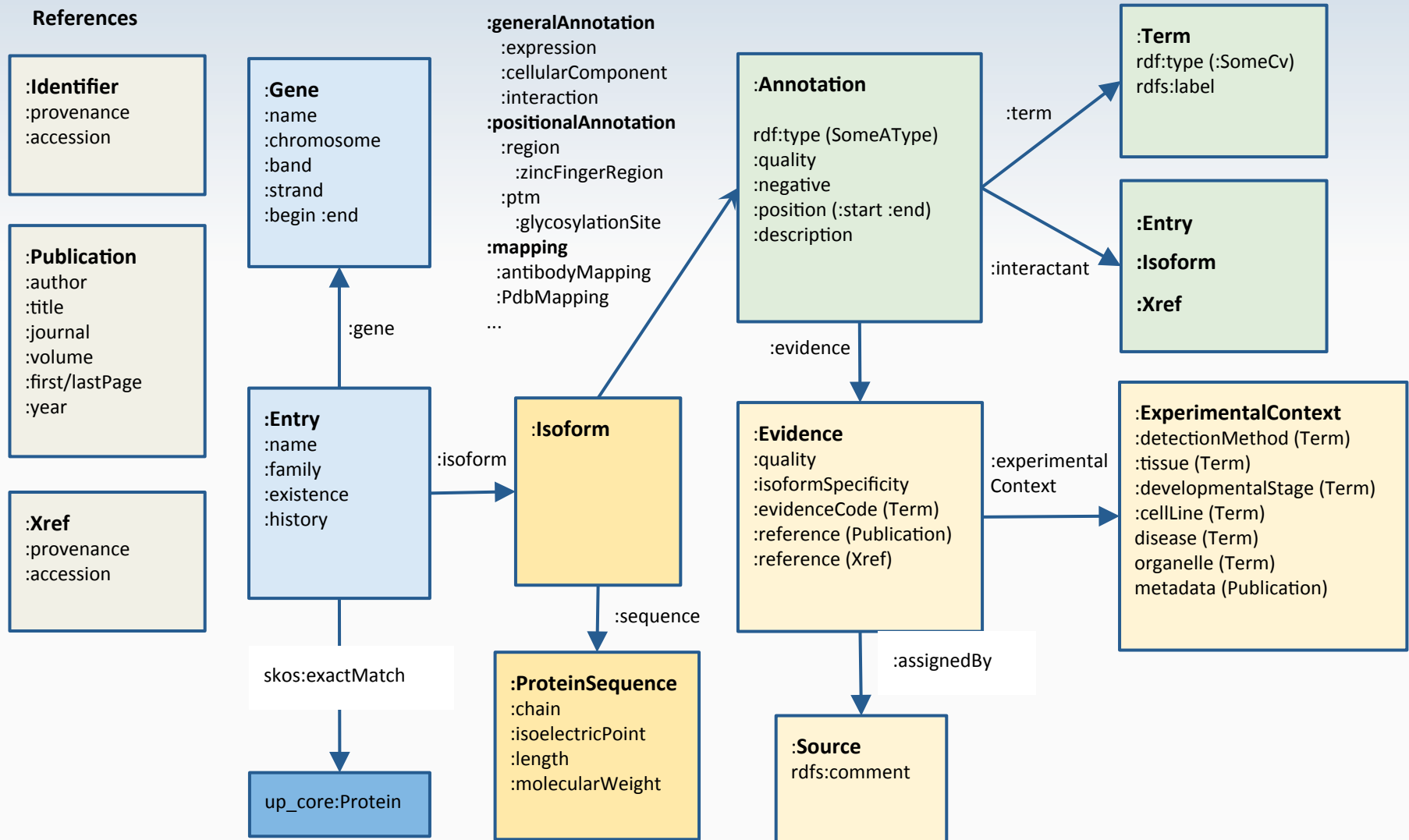
- Human centric
- Isoform centric

Data sources

- UniProt
- proteomics: PeptideAtlas, SRM atlas
- localization: DKFZ, DYP, GO
- variants: UniProtKB, dbSNP via Ensembl, Cosmic, neXtProt
- expression: BGee, HPA
- function: full set of human GO annotations
- interactions: IntAct silver quality interactions



neXtProt - RDF model overview



neXtProt – data model browser

neXtProt SnorQL Resources ▾ Help About Contact us

History (40138)

Identifier (838348)

Induction (3909)

InitiatorMethionine (3266)

InteractingRegion (3524)

InteractionInfo (24165)

IntramembraneRegion (511)

Isoform (41980)

IsoformAnnotationSpecificityQualifier (3)

LargeScalePublication (622)

LipidationSite (1639)

MappingPosition (2731976)

MaturationPeptide (1357)

MatureProtein (43942)

MeshAnatomyCv (1748)

MeshCv (4663)

MetalBindingSite (13301)

Miscellaneous (4010)

Isoform

Values ▾ 41980

Specific information of a protein isoform.

`:BinaryInteraction` `:interactant`

`:Entry` `:isoform`

:Isoform as range

Example

```
:isoform a rdf:Property ;  
  rdfs:domain :Entry ;  
  rdfs:range :Isoform .
```

:Isoform as domain

Example

```
:crossLink a rdf:Property ;  
  rdfs:domain :Isoform ;  
  rdfs:range :CrossLink .
```

`:activeSite` `:ActiveSite` 6514

`:allergen` `:Allergen` 12

`:antibodyMapping` `:AntibodyMapping` 35884

`:betaStrand` `:BetaStrand` 116268

`:binaryInteraction` `:BinaryInteraction` 415002

`:bindingSite` `:BindingSite` 10344

`:calciumBindingRegion` `:CalciumBindingRegion` 1036

`:catalyticActivity` `:CatalyticActivity` 7928

`:caution` `:Caution` 2912

`:cellularComponent` `:GoCellularComponent` 129467

`:cellularComponent` `:SubcellularLocation` 84267

`:cellularComponent` `:SubcellularLocationNote` 11130

`:cleavageSite` `:CleavageSite` 1002

`:cofactor` `:Cofactor` 3707

`:coiledCoilRegion` `:CoiledCoilRegion` 6567

`:compositionallyBiasedRegion` `:CompositionallyBiased` 16

`:crossLink` `:CrossLink` 26010

`:detectedExpression` `:ExpressionProfile` 4701164

`:developmentalStageInfo` `:DevelopmentalStageInfo` 16

`:disease` `:Disease` 29707

`:disulfideBond` `:DisulfideBond` 31380

`:dnaBindingRegion` `:DnaBindingRegion` 1376

`:domain` `:Domain` 41427


Example

Looking for

“phosphorylated proteins located in the
Golgi apparatus”

- Search the existing sample queries
- Understand the query best matching your needs
- Modify it or get help
- Run your query
- Save it for reuse (not in SPARQL playground)

neXtProt – searching samples queries

 neXtProt SnorQL

Resources ▾ Help About Contact us

Login

SPARQL endpoint: <https://api.nextprot.org/sparql>

PREFIX ...

```
select distinct ?entry where {
  ?entry :isoform ?iso.
  ?iso :keyword / :term cv:KW-0597.
  ?iso :cellularComponent /:term /:childOf cv:SL-0086.
}
```

Proteins phosphorylated and located in the cytoplasm

In this query we use the keyword "Phosphorylation" (KW-0597) and the UniProt subcellular location ontology term "Cytoplasm" (SL-0886). Using the "childOf" allows to select for subcellular locations that are, in that ontology, children of cytoplasm like for example "Cell cortex".

html ▾

Go

Reset

eg. peroxisome, liver

Q

Found a bug? [Improve this query!](#)

Tags ▾

Filter sparql examples

Clear filter

My queries

3D structure

active site

alternative splicing

annotation quality

antibody

carbohydrate

cellular component

cellular process

CHEMBL

chromosomal location

classification

coiled coil

chromosomal location

disease

phosphorylated and located in the

ation PTM subcellular location

t are located in both the cytoplasm and

location

h 7 transmembrane regions

nsmembrane

hly expressed in brain but not expressed

ated in mitochondrion and that lack a

location

ose genes are on chromosome 13 and

are associated with a disease

chromosomal location disease


NXQ_00007 - Proteins associated with diseases that are associated with cardiovascular aspects

disease

NXQ_00008 - Proteins whose genes are less than 50000 bp away from the location of the gene coding for protein p53

chromosomal location gene

neXtProt – selecting a sample query

 neXtProt SnorQL Resources ▾ Help About Contact us Login

SPARQL endpoint: <https://api.nextprot.org/sparql>

PREFIX ...

```
select distinct ?entry where {  
  ?entry :isoform ?iso.  
  ?iso :keyword / :term cv:KW-0597.  
  ?iso :cellularComponent /:term /:childOf cv:SL-0086.  
}
```

Proteins phosphorylated and located in the cytoplasm

In this query we use the keyword "Phosphorylation" (KW-0597) and the UniProt subcellular location ontology term "Cytoplasm" (SL-0886). Using the "childOf" allows to select for subcellular locations that are, in that ontology, children of cytoplasm like for example "Cell cortex".

html ▾ Go Reset

eg. peroxisome, liver Q

Found a bug? [Improve this query!](#)

subcellular location ▾ Filter sparql examples

NXQ_00001 - Proteins phosphorylated and located in the cytoplasm

cellular component phosphorylation PTM subcellular location

NXQ_00002 - Proteins that are located in both the cytoplasm and in the nucleus

cellular component subcellular location

NXQ_00005 - Proteins located in mitochondrion and that lack a transit peptide

cellular component subcellular location

NXQ_00010 - Proteins that are glycosylated and are not located in the membrane

carbohydrate cellular component glycosylation PTM subcellular location

NXQ_00016 - Proteins with a mature chain of less than 1000 amino acids which are secreted and do not contain cysteines in the mature chain

cellular component processing sequence subcellular location

NXQ_00017 - Proteins larger than 1000 amino acids that are located in the nucleus and expressed in the nervous system

cellular component expression sequence subcellular location

NXQ_00018 - Proteins that are acetylated and methylated and located in the nucleus

cellular component PTM subcellular location

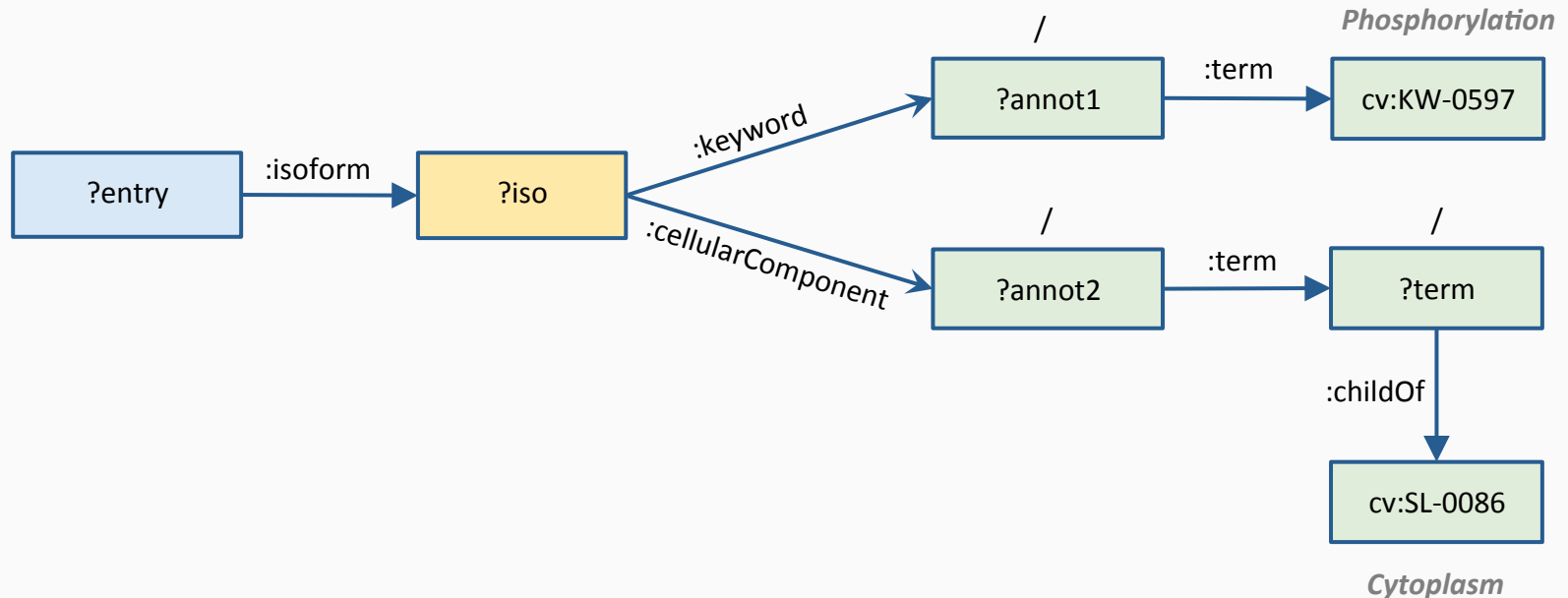
NXQ_00021 - Proteins with at least one HPA antibody that are

neXtProt – understanding the query

SPARQL endpoint: <https://api.nextprot.org/sparql>

PREFIX ...

```
select distinct ?entry where {  
  ?entry :isoform ?iso.  
  ?iso :keyword / :term cv:KW-0597.  
  ?iso :cellularComponent /:term /:childOf cv:SL-0086.  
}
```

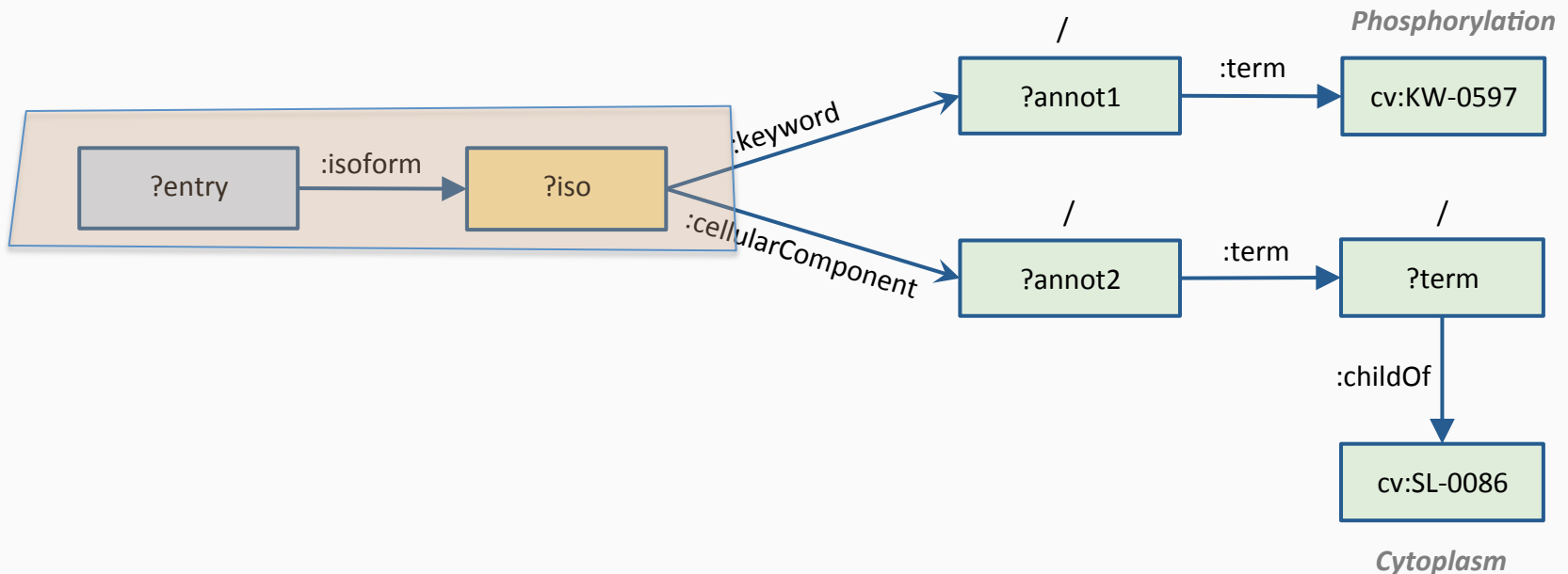


neXtProt – understanding the query

SPARQL endpoint: <https://api.nextprot.org/sparql>

PREFIX ...

```
select distinct ?entry where {  
  ?entry :isoform ?iso.  
  ?iso :keyword / :term cv:KW-0597.  
  ?iso :cellularComponent /:term /:childOf cv:SL-0086.  
}
```

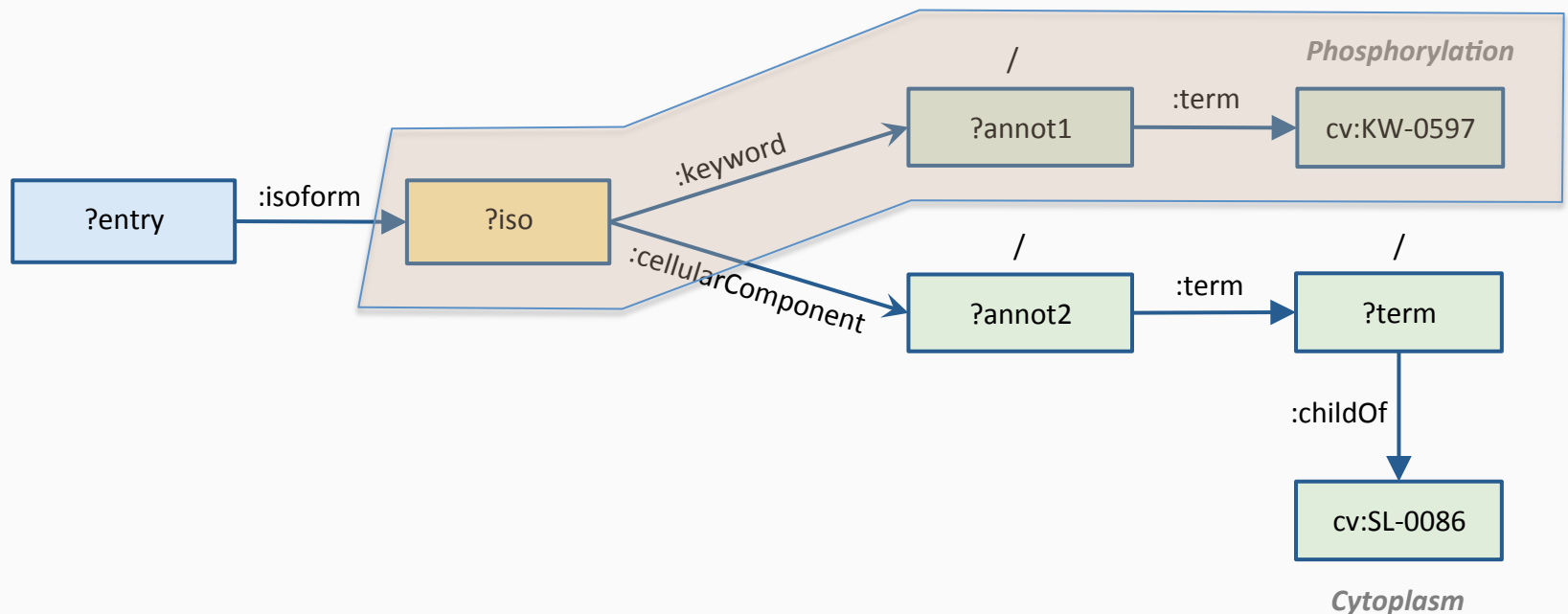


neXtProt – understanding the query

SPARQL endpoint: <https://api.nextprot.org/sparql>

PREFIX ...

```
select distinct ?entry where {  
  ?entry :isoform ?iso.  
  ?iso :keyword / :term cv:KW-0597.  
  ?iso :cellularComponent /:term /:childOf cv:SL-0086.  
}
```

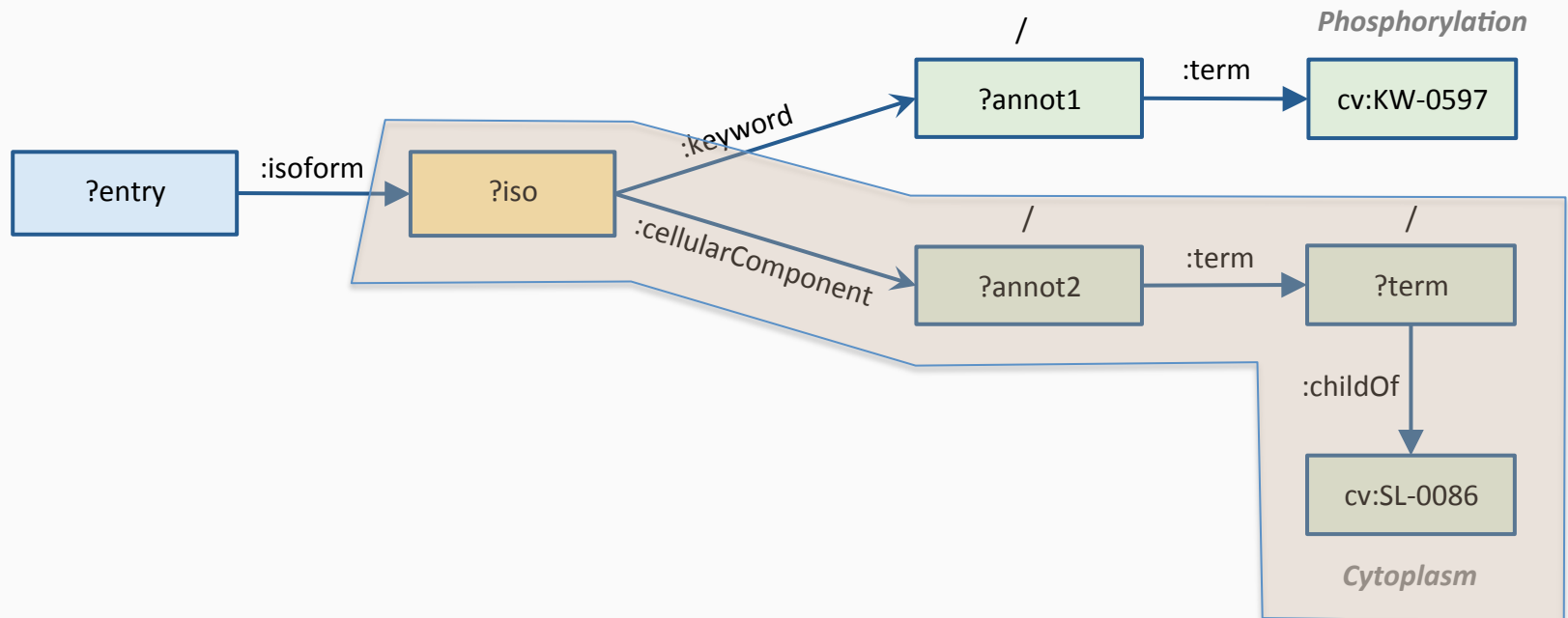


neXtProt – understanding the query

SPARQL endpoint: <https://api.nextprot.org/sparql>

PREFIX ...

```
select distinct ?entry where {  
  ?entry :isoform ?iso.  
  ?iso :keyword / :term cv:KW-0597.  
  ?iso :cellularComponent /:term /:childOf cv:SL-0086.  
}
```



neXtProt – I am lost, help !

neXtProt SnorQL Resources Help About Contact us Login

SPARQL endpoint: <https://api.nextprot.org/sparql>

PREFIX ...

```
select distinct ?entry where {
  ?entry isoform ?iso.
  ?iso keyword /:term cv:KW-0597.
  ?iso :cellularComponent /:term /:childOf cv:SL-0086.
}
```

subcellular location Filter sparql examples

NXQ_00001 - Proteins phosphorylated and located in the cytoplasm
cellular component phosphorylation PTM subcellular location

NXQ_00002 - Proteins that are located in both the cytoplasm and in the nucleus
cellular component subcellular location

NXQ_00005 - Proteins located in mitochondrion and that lack a transit peptide
cellular component subcellular location

Proteins phospho

In this query w
term "Cytoplasm
Using the "chil
cytoplasm like

html Go

Found a bug? Im

How to write your query ?

- Select your query from the examples on the right side
- Find terms appropriate to your query using the term finder (e.g. peroxisome, liver)
- You can save your queries on the [search interface](#) or directly using the [api](#) and access them in Snorql using the **My Queries** tag.

Need more ?

- Just send us your query, you'll get the SPARQL from us: => [Mail your query to us](#)
- Read the Help section above about our vocabulary: => [Help on nextprot vocabulary](#)
- Read more documentation about the SPARQL language: => [SPARQL 1.1 Query Language](#)

How to write your query

- Select your query from the examples on the right side
- Find terms appropriate to your query using the term finder (e.g. peroxisome, liver)
- You can save your queries on the [search interface](#) or directly using the [api](#) and access them in Snorql using the **My Queries** tag.

Need more ?

- Just send us your query, you'll get the SPARQL from us: => [Mail your query to us](#)
- Read the Help section above about our vocabulary: => [Help on nextprot vocabulary](#)
- Read more documentation about the SPARQL language: => [SPARQL 1.1 Query Language](#)

located in the mitochondrion


cellular component interaction PPI subcellular location

NXQ_00029 - Proteins anchored to the membrane via a GPI-anchor
PTM subcellular location

NXQ_00035 - Proteins located in the mitochondrion and which are enzymes
cellular component enzyme function subcellular location

NXQ_00042 - Proteins that bind a metal and are secreted
cellular component metal subcellular location

neXtProt – searching terms

 neXtProt SnorQL

Resources ▾ Help About Contact us

Login

SPARQL endpoint: <https://api.nextprot.org/sparql>

PREFIX ...

```
select distinct ?entry where {  
  ?entry :isoform ?iso.  
  ?iso :keyword / :term cv:KW-0597.  
  ?iso :cellularComponent / :term /:childOf cv:SL-0086.  
}
```

Proteins phosphorylated and located in the cytoplasm

In this query we use the keyword "Phosphorylation" (KW-0597) and the UniProt subcellular location ontology term "Cytoplasm" (SL-0086). Using the "childOf" allows to select for subcellular locations that are, in that ontology, children of cytoplasm like for example "Cell cortex".

html ▾ Go Reset

eg. peroxisome, liver

Found a bug? [Improve this query!](#)

Tags ▾ Filter sparql examples

NXQ_00001 - Proteins phosphorylated and located in the cytoplasm
cellular component phosphorylation PTM subcellular location

NXQ_00002 - Proteins that are located in both the cytoplasm and in the nucleus
cellular component subcellular location

NXQ_00003 - Proteins with 7 transmembrane regions
domain region topology transmembrane

NXQ_00004 - Proteins highly expressed in brain but not expressed in testis
expression




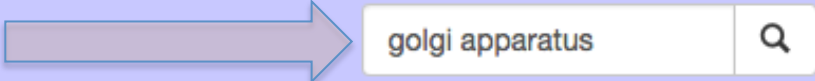
NXQ_00005 - Proteins located in mitochondrion and that lack a transit peptide
cellular component subcellular location

NXQ_00006 - Proteins whose genes are on chromosome 13 and are associated with a disease
chromosomal location disease















NXQ_00007 - Proteins associated with diseases that are associated with cardiovascular aspects
disease

NXQ_00008 - Proteins whose genes are less than 50000 bp away from the location of the gene coding for protein p53
chromosomal location gene

neXtProt – searching terms

html    

Query time is 0.026[s] for 7 rows

term	label	type
cv:GO_0005794 	"Golgi apparatus"^^xsd:string	:GoCellularComponentCv 
cv:KW-0333 	"Golgi apparatus"^^xsd:string	:UniprotKeywordCv 
cv:SL-0133 	"Golgi apparatus lumen"^^xsd:string	:UniprotSubcellularLocationCv 
cv:SL-0134 	"Golgi apparatus membrane"^^xsd:string	:UniprotSubcellularLocationCv 
cv:SL-0132 	"Golgi apparatus"^^xsd:string	:UniprotSubcellularLocationCv 
cv:D006056 	"Golgi Apparatus"^^xsd:string	:MeshAnatomyCv 
cv:GO_0044431 	"Golgi apparatus part"^^xsd:string	:GoCellularComponentCv 

neXtProt – searching terms

html

golgi apparatus

Query time is 0.026[s] for 7 rows

term	label	type
cv:GO_0005794 [http]	"Golgi apparatus"^^xsd:string	:GoCellularComponentCv [http]
cv:KW-0333 [http]	"Golgi apparatus"^^xsd:string	:UniprotKeywordCv [http]
cv:SL-0133 [http]	"Golgi apparatus lumen"^^xsd:string	:UniprotSubcellularLocationCv [http]
cv:SL-0134 [http]	"Golgi apparatus membrane"^^xsd:string	:UniprotSubcellularLocationCv [http]
cv:SL-0132 [http]	"Golgi apparatus"^^xsd:string	:UniprotSubcellularLocationCv [http]
cv:D006056 [http]	"Golgi Apparatus"^^xsd:string	:MeshAnatomyCv [http]
cv:GO_0044431 [http]	"Golgi apparatus part"^^xsd:string	:GoCellularComponentCv [http]

neXtProt – modifying the query

html

golgi apparatus

Query time is 0.026[s] for 7 rows

term	label	type
cv:GO_0005794 [http]	"Golgi apparatus"^^xsd:string	:GoCellularComponentCv [http]
cv:KW-0333 [http]	"Golgi apparatus"^^xsd:string	:UniprotKeywordCv [http]
cv:SL-0133 [http]	"Golgi apparatus lumen"^^xsd:string	:UniprotSubcellularLocationCv [http]
cv:SL-0134 [http]	"Golgi apparatus membrane"^^xsd:string	:UniprotSubcellularLocationCv [http]
cv:SL-0132 [http]	"Golgi apparatus"^^xsd:string	:UniprotSubcellularLocationCv [http]
cv:D006056 [http]	"Golgi Apparatus"^^xsd:string	:MeshAnatomyCv [http]
cv:GO_0044431 [http]	"Golgi apparatus part"^^xsd:string	:GoCellularComponentCv [http]

SPARQL endpoint: <https://api.nextprot.org/sparql>

PREFIX ...

```
select distinct ?entry where {  
  ?entry :isoform ?iso.  
  ?iso :keyword / :term cv:KW-0597.  
  ?iso :cellularComponent /:term /:childOf cv:SL-0132.  
}
```

neXtProt – running the query

SPARQL endpoint: <https://api.nextprot.org/sparql>

PREFIX ...

```
select distinct ?entry where {  
  ?entry :isoform ?iso.  
  ?iso :keyword / :term cv:KW-0597.  
  ?iso :cellularComponent /:term /:childOf cv:SL-0132.  
}
```

Proteins phosphorylated and located in the cytoplasm

In this query we use the keyword "Phosphorylation" (KW-0597) and the UniProt subcellular location ontology term "Cytoplasm" (SL-0886). Using the "childOf" allows to select for subcellular locations that are, in that ontology, children of cytoplasm like for example "Cell cortex".

html

Found a bug? [Improve this query!](#)

neXtProt – running the query

SPARQL endpoint: <https://api.nextprot.org/sparql>

PREFIX ...

```
select distinct ?entry where {
  ?entry :isoform ?iso.
  ?iso :keyword / :term cv:KW-0597.
  ?iso :cellularComponent /:term /:childOf cv:SL-0132.
}
```

Proteins phosphorylated and located in the cytoplasm

In this query we use the keyword "Phosphorylation" (KW-0597) and the UniProt subcellular location ontology term "Cytoplasm" (SL-0886).
Using the "childOf" allows to select for subcellular locations that are, in that ontology, children of cytoplasm like for example "Cell cortex".

html

Found a bug? [Improve this query!](#)

Query time is 0.332[s] for 628 rows

entry
entry:NX_P36956 --- (neXtProt link) [http]
entry:NX_O15258 --- (neXtProt link) [http]
entry:NX_P22059 --- (neXtProt link) [http]

neXtProt – saving your query

The screenshot displays the neXtProt Search web application. At the top, the navigation bar includes the neXtProt Search logo, links for Resources, Help, About, and Contact us, and a user profile for Pierre-Andre Michel. Below the navigation bar, there are search options: Simple search and Advanced search (selected). A search input field contains the number '1', and a dropdown menu shows 'proteins'. A blue arrow points from the search input field to a modal window titled 'New Query'. The modal window has a 'Tags' dropdown and a search bar containing 'search in 117 queries (ex: liver)'. It lists nine protein queries with their corresponding IDs and eye icons for visibility. The queries are: 'Proteins phosphorylated and located in the cytoplasm' (NXQ_00001), 'Proteins that are located in both the cytoplasm and in the nucleus' (NXQ_00002), 'Proteins with 7 transmembrane regions' (NXQ_00003), 'Proteins highly expressed in brain but not expressed in testis' (NXQ_00004), 'Proteins located in mitochondrion and that lack a transit peptide' (NXQ_00005), 'Proteins whose genes are on chromosome 13 and are associated with a disease' (NXQ_00006), 'Proteins associated with diseases that are associated with cardiovascular aspects' (NXQ_00007), 'Proteins whose genes are less than 50000 bp away from the location of the gene coding for protein p53' (NXQ_00008), and 'Proteins with 3 disulfide bonds and that are not annotated as hormones' (NXQ_00009). At the bottom of the modal window, there is a note: 'For help with these queries look at the model or contact us.'

neXtProt Search Resources ▾ Help ▾ About Contact us

Simple search Advanced search

proteins ▾ 1

Search

Show 1 to 0 of 0 ▾

FILTERS
For proteins with:

Legal disclaimer

Sort score ▾

New Query ✕

Tags ▾ search in 117 queries (ex: liver)

Proteins phosphorylated and located in the cytoplasm	NXQ_00001	👁
Proteins that are located in both the cytoplasm and in the nucleus	NXQ_00002	👁
Proteins with 7 transmembrane regions	NXQ_00003	👁
Proteins highly expressed in brain but not expressed in testis	NXQ_00004	👁
Proteins located in mitochondrion and that lack a transit peptide	NXQ_00005	👁
Proteins whose genes are on chromosome 13 and are associated with a disease	NXQ_00006	👁
Proteins associated with diseases that are associated with cardiovascular aspects	NXQ_00007	👁
Proteins whose genes are less than 50000 bp away from the location of the gene coding for protein p53	NXQ_00008	👁
Proteins with 3 disulfide bonds and that are not annotated as hormones	NXQ_00009	👁

Version: 0.1.4 (build 844)

For help with these queries look at the model or contact us.

neXtProt – saving your query

neXtProt Search Resources ▾ Help ▾ About Contact us 📄 ★ Pierre-Andre Michel ▾

☐ Simple search ☒ Advanced search

proteins ▾ 1 ☰ ✕

Show 1 to 0 of 0 ▾ phosphorylated proteins in golgi ✕ Sort score ▾

FILTERS
For proteins with:

```
1 select distinct ?entry where {  
2   ?entry :isoform ?iso.  
3   ?iso :keyword / :term cv:KW-0597.  
4   ?iso :cellularComponent /:term /:childOf cv:SL-0132.  
5 }
```

Write some description...

[Legal disclaimer](#) 🔗 Version: 0.1.4 (build 844)

For help with these queries look at [the model](#) or [contact us](#). ➡

neXtProt – reusing your query

The screenshot displays the neXtProt Search interface. At the top, the navigation bar includes the neXtProt logo, 'Search', 'Resources', 'Help', 'About', and 'Contact us'. The user 'Pierre-Andre Michel' is logged in. Below the navigation bar, there are radio buttons for 'Simple search' and 'Advanced search', and a search input field containing the number '1'. A dropdown menu for 'proteins' is also visible.

The main content area shows a list of search results. On the left, there is a 'FILTERS' section with the text 'For proteins with:'. A blue arrow points from this section to the 'Tags' dropdown menu. The 'Tags' dropdown menu is open, showing a list of filters: 'Clear filter', 'My queries', '3D structure', 'active site', 'annotation quality', 'antibody', 'carbohydrate', 'cellular component', 'cellular process', 'CHEMBL', 'chromosomal location', 'classification', 'coiled coil', 'compositional bias', 'cross-reference', 'dataset', and 'disease'. The 'My queries' option is highlighted.

The search results table has a header 'Tags' and a search input field 'search in 118 queries (ex: liver)'. A 'New Query' button is located next to the search input field. The table lists various queries with their IDs (NXQ_00001 to NXQ_00009) and a 'Sort score' column. The queries are:

- located in the cytoplasm (NXQ_00001)
- both the cytoplasm and in the nucleus (NXQ_00002)
- ane regions (NXQ_00003)
- brain but not expressed in testis (NXQ_00004)
- rdion and that lack a transit peptide (NXQ_00005)
- n chromosome 13 and are associated with a disease (NXQ_00006)
- ases that are associated with cardiovascular aspects (NXQ_00007)
- ess than 50000 bp away from the location of the gene (NXQ_00008)
- ds and that are not annotated as hormones (NXQ_00009)

At the bottom right, the version information 'Version: 0.1.4 (build 844)' is displayed.

neXtProt – reusing your query

The screenshot displays the neXtProt Search web application. At the top, the navigation bar includes the neXtProt Search logo, links for Resources, Help, About, and Contact us, and a user profile for Pierre-Andre Michel. Below the navigation bar, there are search options: Simple search and Advanced search (selected). A search bar contains the text 'proteins' and a dropdown menu showing '1'. A 'Search' button is visible. On the left side, there is a 'FILTERS' section with the text 'For proteins with:'. A large blue arrow points from this section to a modal window. The modal window, titled 'My queries', shows a list of saved queries: 'test query' (associated with dSMdpFoW) and 'phosphorylated proteins in golgi' (associated with MJM9EVX9). Each query has an eye icon next to it. At the bottom of the modal, there is a note: 'For help with these queries look at [the model](#) or [contact us](#).' The footer of the page includes a 'Legal disclaimer' link and the version information 'Version: 0.1.4 (build 844)'.

neXtProt Search Resources ▾ Help ▾ About Contact us

Simple search ☐ Advanced search ☒

proteins ▾ 1 Search

Show 1 to 0 of 0 ▾

FILTERS
For proteins with:

My queries ▾ search in 2 queries (ex: liver) New Query ✕

test query	dSMdpFoW	
phosphorylated proteins in golgi	MJM9EVX9	

For help with these queries look at [the model](#) or [contact us](#).

[Legal disclaimer](#) Version: 0.1.4 (build 844)

neXtProt – reusing your query

The screenshot displays the neXtProt Search web application. At the top, there is a navigation bar with the neXtProt logo, search options (Simple search, Advanced search), and a user profile for Pierre-Andre Michel. Below the navigation bar, a search bar contains the text 'proteins' and a search button. A modal window is open in the center, titled 'My queries', showing a list of saved queries. The first query is 'test query' with ID 'dSMdpFoW'. The second query is 'phosphorylated proteins in golgi' with ID 'MJM9EVX9'. A 'New Query' button is visible in the top right of the modal. At the bottom of the modal, there is a link to 'the model' and a 'contact us' link. The background of the page shows a 'Show 1 to 0 of 0' result indicator and a 'Sort score' dropdown.

neXtProt Search Resources ▾ Help ▾ About Contact us

Simple search Advanced search

proteins ▾ 1 Search

Show 1 to 0 of 0 ▾

FILTERS
For proteins with:

My queries ▾ search in 2 queries (ex: liver) New Query ✕

test query	dSMdpFoW	👁
phosphorylated proteins in golgi	MJM9EVX9	👁

For help with these queries look at [the model](#) or [contact us](#).

[Legal disclaimer](#) Version: 0.1.4 (build 844)

Sharable URL for your query

<https://search.nextprot.org/proteins/search?mode=advanced&queryId=MJM9EVX9>

Now do it yourself !

from: <http://sparql-playground.nextprot.org>

or: <http://localhost:7777>

SPARQL playground Documentation FAQs Data About SPARQL Resources SIB

DOCUMENTATION

- Introduction
- Exercises**
- Exercise1
- Exercise2
- Exercise3
- Exercise4
- About

RDF ENTITIES

- ActiveSite (6664)
- Allergen (12)
- AnnotationType (85)
- AntibodyMapping (37144)
- BetaStrand (116700)
- BgeeDevelopmentalStageCv (192)

Exercises

- 1. Glycoproteins located in the cytoplasm**
[How to build this query](#)
- 2. Proteins with 10 or more gold interactions with SH2 domain-containing proteins**
[How to build this query](#)
- 3. Proteins with a variant in a homeobox region**
[How to build this query](#)
- 4 Proteins with a lipidation site having a variant on it**
[How to build this query](#)

Exercises

[SPARQL playground](#) [Documentation](#) [FAQs](#) [Data](#) [About](#)

DOCUMENTATION

Introduction

Exercise1

Exercise2

Exercise3

Exercise4

About

RDF ENTITIES

ActiveSite (6664)

Allergen (12)

AnnotationType (85)

AntibodyMapping (37144)

BetaStrand (116700)

BgeeDevelopmentalStageCv (192)

BinaryInteraction (352658)

BindingSite (10587)

CalciumBindingRegion (1063)

Exercise 1

Build a SPARQL query to find:

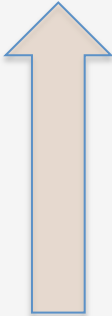
- **Glycoproteins located in the cytoplasm**

Procedure

- Select sample query 1 (NXQ_00001)
- Use the term finder to get the accession code of *glycoprotein*
- Modify the keyword term accession in the query
- Run the query
- Check the result you get by clicking one of the (neXtProt link) in the result:
 - bottom of *Localisation* TAB should contain the term *Cytoplasm* or a more specific one (a :childOf Cytoplasm)
 - *Keyword* section of Sequence TAB should contain the *Glycoprotein* term

Terms

- Cytoplasm [cv:SL-0086], Golgi apparatus [cv:SL-0132], Phosphoprotein [cv:KW-0597], Glycoprotein [cv:KW-0325]



The end

Many thanks to

Jerven Bolleman

Daniel Teixeira

Alain Gateau

Monique Zahn

Pascale Gaudet