

Swiss Institute of
Bioinformatics

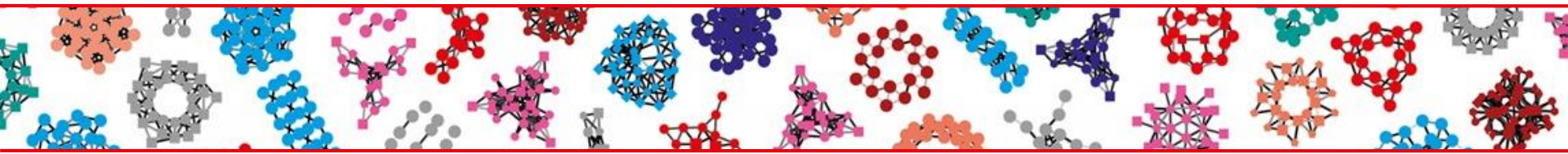
SPARQLing neXtProt data

SWAT4HCLS, Edinburgh, United Kingdom
Monique Zahn



www.sib.swiss

Overview



01

• **Introduction to neXtProt**

02

• **Data model**

03

• **SPARQLing in neXtProt**

04

• **Summary**

neXtProt – the SIB knowledgebase on human proteins

<https://www.nextprot.org/>

The screenshot displays the neXtProt website interface. At the top, there is a navigation bar with the neXtProt logo on the left and menu items: Tools, Portals, Download, Help, About, and Contact. On the right side of the navigation bar, there is a Login button and the SIB logo.

The main content area features the neXtProt logo and the tagline "Exploring the universe of human proteins". Below this, there are search options: "Simple search" (selected) and "Advanced search (SPARQL)". A search input field contains "proteins" and "Gold only". A search button with a magnifying glass icon and the text "Search" is located to the right of the input field. Below the search field, there is a hint: "e.g.: Search for MSH6 in proteins, Search for author Doolittle in publications, Search for liver in terms".

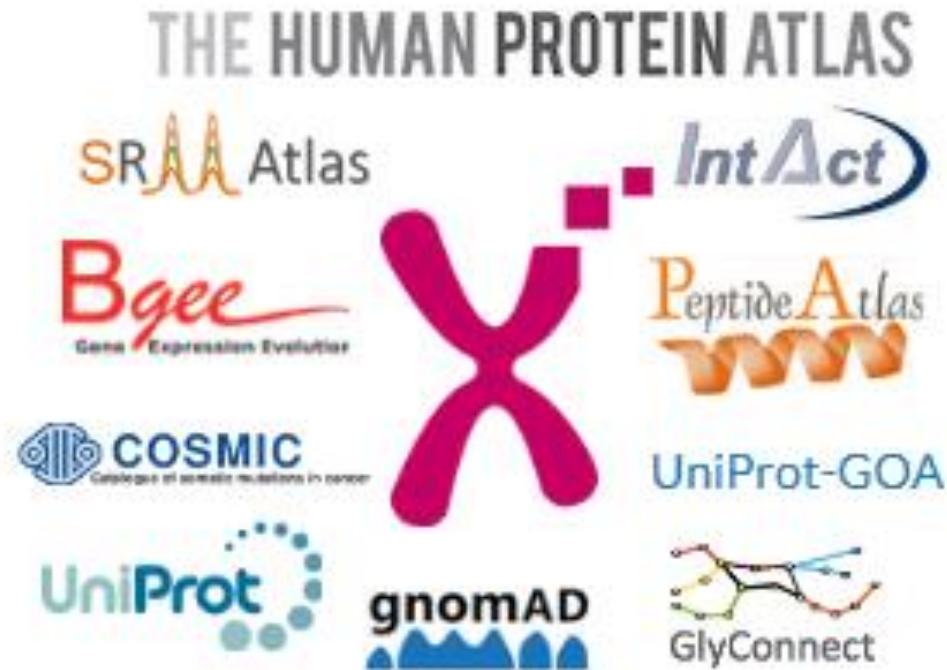
A central box highlights a "NAR 2020 PAPER" with the text: "The neXtProt knowledgebase in 2020: data, tools and usability improvements. Nucleic Acids Research, Database Issue (2020)".

Below the paper box, there are four content panels:

- Getting started:** Includes links to "The human proteome", "Simple search", "Advanced search (SPARQL)", "Analyze" (BLAST, list management, peptide uniqueness checker), "Download" (XML, FASTA, PEFF), and "Technical corner".
- Data sources:** Features logos for "THE HUMAN PROTEIN ATLAS", "SRÅ Atlas", "Bgee", "COSMIC", "UniProt", "gnomAD", "IntAct", "Peptide Atlas", "UniProt-GOA", and "GlyConnect".
- News:** Lists "Previous releases and new advanced se..." (Dec 04, 2019), "New Publication: The neXtProt knowled..." (Nov 29, 2019), and "New Publication: A Blinded Testing of..." (Oct 10, 2019).
- Release 2019-08-22:** Titled "Protein existence in neXtProt", it includes a donut chart showing the distribution of protein existence evidence: Predicted (71), Uncertain (576), Inferred from homology (510), Evidence at transcript level, and Evidence at protein level (17694).

The footer contains copyright information: "© 2011 - 2019 SIB Swiss Institute of Bioinformatics | Legal disclaimer | Data release 2019-08-22 | Application release v2.24.0 | For developers".

Data sources

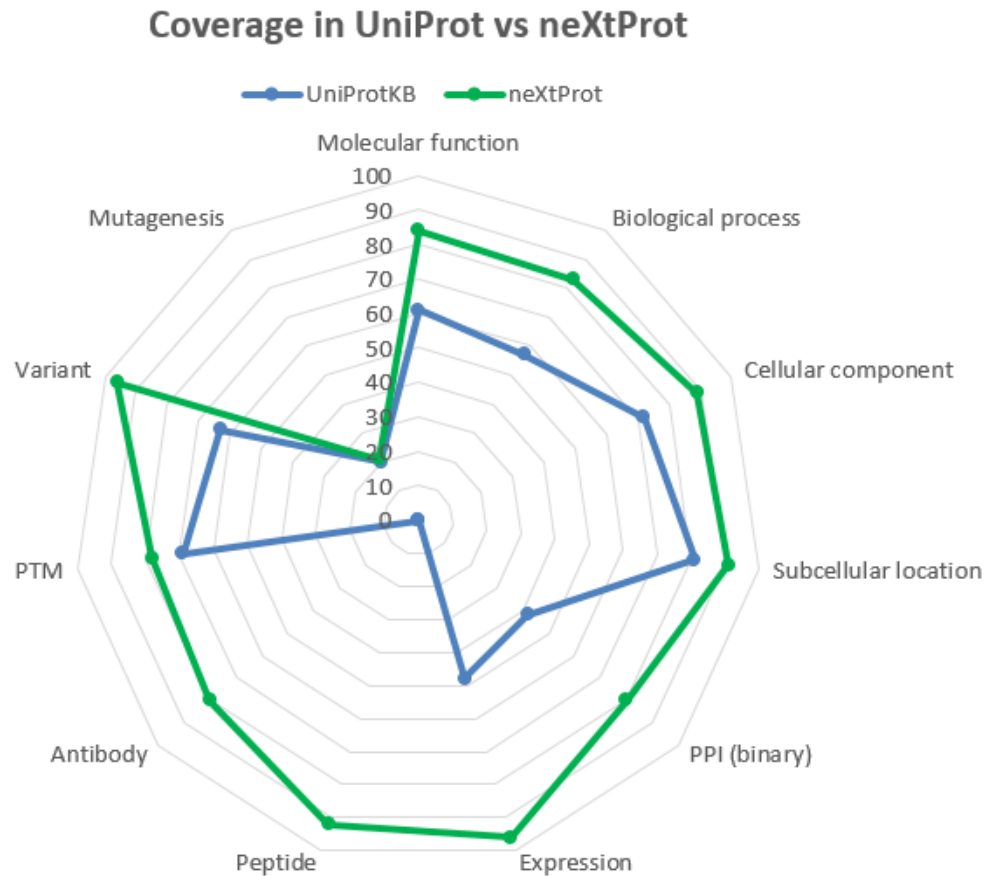


COMING SOON :



“We stand on the shoulders of giants.”

Improved coverage through data integration



Data on human proteins

https://www.nextprot.org/entry/NX_P01308/

Views

The screenshot displays the neXtProt website interface for the entry NX_P01308, titled "INS - Function". The page is organized into several sections:

- Navigation:** Includes a top menu with "Tools", "Portals", "Download", "Help", "About", and "Contact". A search bar is located in the top right, and a "Login" button is next to the SIB logo.
- Entry Header:** Shows the entry ID "NX_P01308" and the title "INS - Function". It includes statistics for "GENE" (1), "ISO" (1), and "REF" (804). A search bar and a "Gold Only" filter are also present.
- Left Sidebar:** A navigation menu with categories: "PROTEIN" (selected), "Function", "Medical", "Expression", "Interactions", "Localization", "Sequence", "Proteomics", "Structures", "Peptides", "Phenotypes", "Exons", and "Identifiers". Below this is a "REFERENCES" section with counts: "Curated publications" (98), "Additional publications" (699), "Patents" (0), and "Submissions" (4).
- Main Content Area:**
 - INS → Insulin:** A sub-header for the specific protein.
 - Metadata:** "Cleaved into: Insulin A chain; Insulin B chain", "Gene name: INS", and "Family name: Insulin".
 - Text:** "Entry whose protein(s) existence is based on evidence at protein level."
 - Annotations:** A note stating "Annotations in this section apply to all the isoforms if not specified otherwise." and a "Show evidences" button.
 - OVERVIEW:** A paragraph describing the function: "Insulin decreases blood glucose concentration. It increases cell permeability to monosaccharides, amino acids and fatty acids. It accelerates glycolysis, the pentose phosphate cycle, and glycogen synthesis in liver." with a "Gold" badge and a "1 ev" evidence count.
 - GO MOLECULAR FUNCTION:** A list of GO terms with "Definition" and "Gold" badges:
 - Hormone activity (GO:0005179)
 - Identical protein binding (GO:0042802)
 - Insulin receptor binding (GO:0005158)
 - Insulin-like growth factor receptor binding (GO:0005159)
 - Protease binding (GO:0002020)
 - Protein binding (GO:0005515)
 - Evidence Summary:** A vertical list of evidence counts and sources:
 - 1 ev UniProt
 - 3 ev BHF-UCL UniProt
 - 5 ev IntAct
 - 3 ev UniProt BHF-UCL
 - 1 ev BHF-UCL
 - 1 ev UniProt
 - 4 ev IntAct

Types of data (1)

https://www.nextprot.org/entry/NX_P01308/

The screenshot shows the NextProt website interface for the entry 'INS - Function'. The page is divided into several sections:

- Header:** Includes the NextProt logo, navigation menus (Tools, Portals, Download, Help, About, Contact), a search bar, and a 'Gold Only' filter.
- Left Sidebar:** Contains a 'PROTEIN' dropdown menu and a 'REFERENCES' section with counts for Curated publications (98), Additional publications (699), Patents (0), and Submissions (4).
- Main Content Area:**
 - INS - Function:** Shows 'INS → Insulin' with statistics (1 GENE, 1 ISO, 804 REF).
 - Basic Information:** Cleaved into: Insulin A chain; Insulin B chain; Gene name: INS; Family name: Insulin.
 - Annotations:** 'Entry whose protein(s) existence is based on evidence at protein level.'
 - OVERVIEW:** 'Insulin decreases blood glucose concentration. It increases cell permeability to monosaccharides, amino acids and fatty acids. It accelerates glycolysis, the pentose phosphate cycle, and glycogen synthesis in liver.' (Gold)
 - GO MOLECULAR FUNCTION:** Lists functions like 'Hormone activity' (GO:0005179), 'Identical protein binding' (GO:0042802), 'Insulin receptor binding' (GO:0005158), 'Insulin-like growth factor receptor binding' (GO:0005159), and 'Protease binding' (GO:0002020), all with 'Definition' and 'Gold' tags.
- Right Side:** Includes 'Extend overview' and 'Show evidences' buttons, and a list of evidence sources (e.g., UniProt, BHF-UCL, IntAct) with counts and 'ev' labels.

Two pink callouts are overlaid on the page:

- Free text:** A pink oval with an arrow pointing to the overview text: 'Insulin decreases blood glucose concentration. It increases cell permeability to monosaccharides, amino acids and fatty acids. It accelerates glycolysis, the pentose phosphate cycle, and glycogen synthesis in liver.'
- Structured:** A pink oval with an arrow pointing to the GO Molecular Function list.

At the bottom, a footer contains copyright information: © 2011 - 2019 SIB Swiss Institute of Bioinformatics | Legal disclaimer | Data release 2019-08-22 | Application release v2.23.2 | For developers

Types of data (2)

https://www.nextprot.org/entry/NX_P01308/sequence

Positional

Displayed isoform: Iso 1

Position: 107 Zoom: x 1

Sequence: MALWMRLLP L L A L L A L W G P D P A A A F V N Q H L C G S H L V E A L Y L V C G E R G F F Y T P K T R R E A E D L Q V G Q V E L G G G P G A G S L Q P L A L E G S L Q R G I V E Q C C T S I C S L Y Q L E N Y C N

Signal peptide: [bar chart]

Propeptide: [bar chart]

Mature protein: Insulin B chain [bar chart]

Disulfide bond: [bar chart]

Variant: [bar chart]

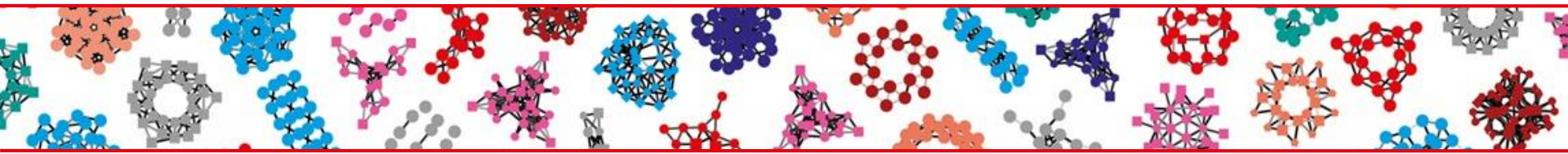
Search a feature.. SEARCH

Name	Position	Length	Description	Evidence
Processing				
Signal peptide	1-24	24	[Gold]	1 UniProtKB
Mature protein	25-54	30	Insulin B chain [Gold]	1 UniProtKB
Propeptide	57-87	31	C peptide [Gold]	1 UniProtKB
Mature protein	90-110	21	Insulin A chain [Gold]	1 UniProtKB
Modified Residue				

FASTA sequence:
1 MALWMRLLEP LALLALWGPDPAAAFVNHQLCGSHLVEALYLVCGERGFFY
51 TPKTRREAE DLQVQVELGGPGAGSLQPLALEGSLQRRGIVEQCCTSI
101 SLYQLENYC

© 2011 - 2019 SIB Swiss Institute of Bioinformatics | Legal disclaimer | Data release 2019-08-22 | Application release v2.23.2 | For developers

Overview



01 — Introduction to neXtProt

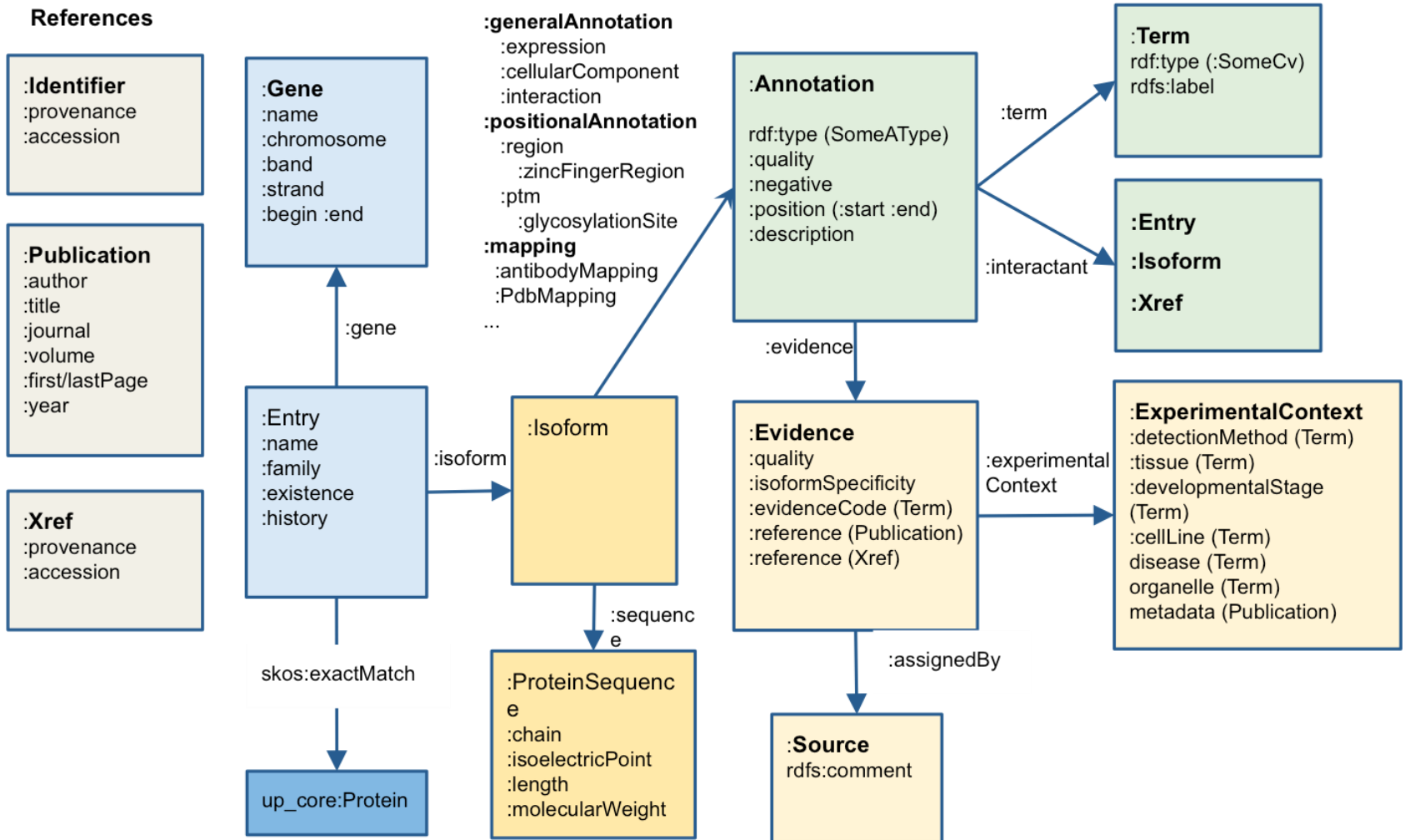
02 — **Data model**

03 — SPARQLing in neXtProt

04 — Summary

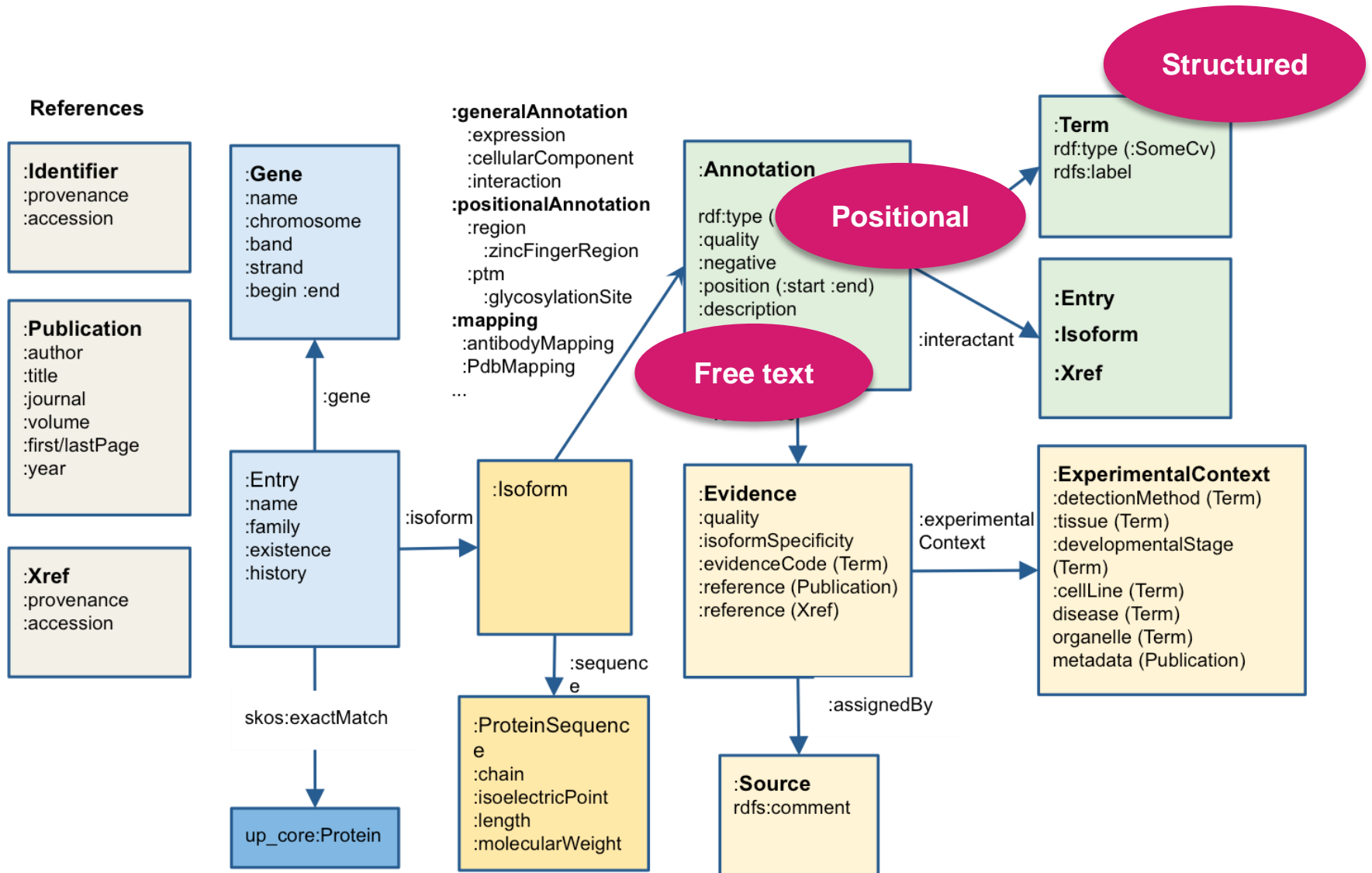
Data model

<https://www.nextprot.org/help/data-model>

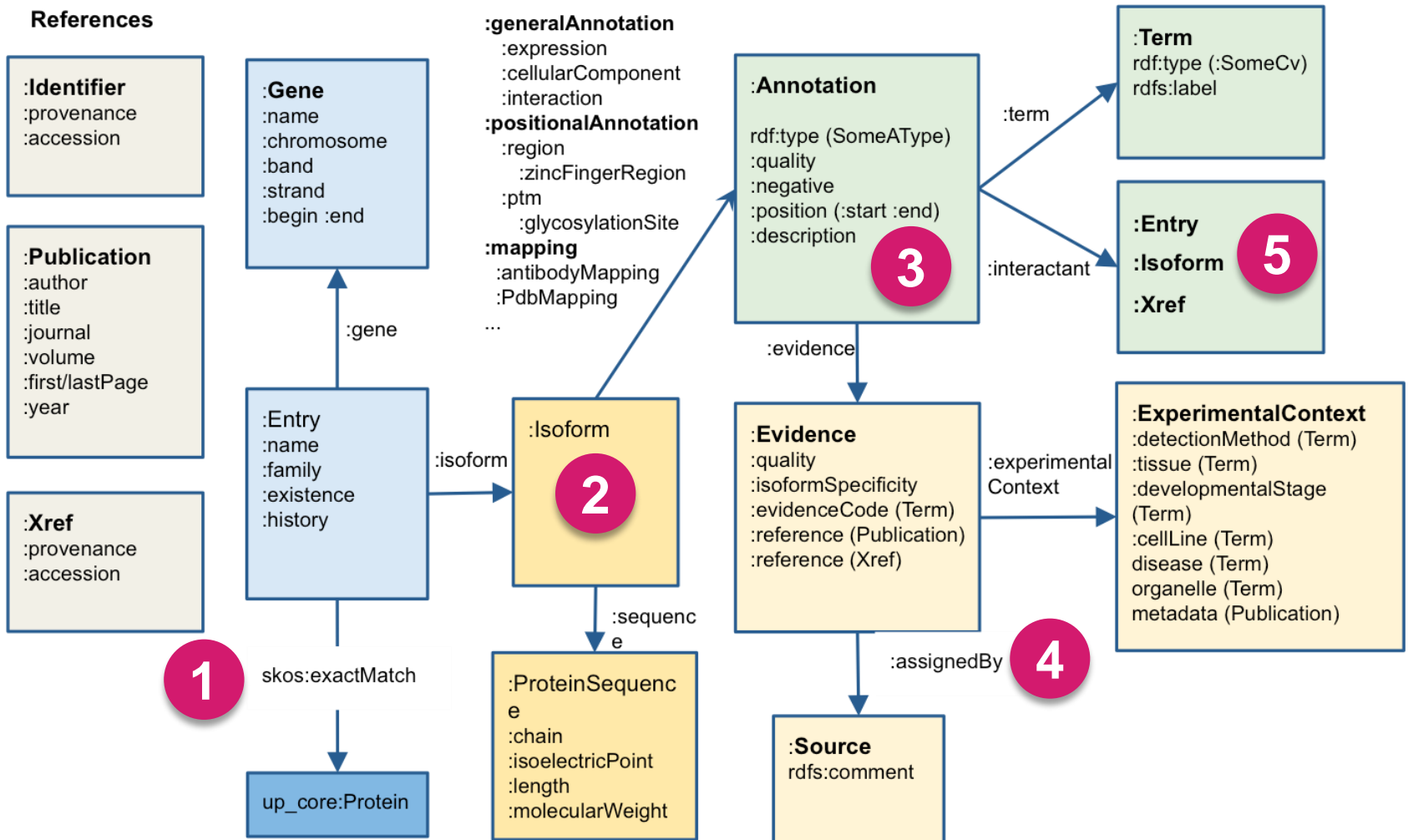


Single interoperable model (RDF)

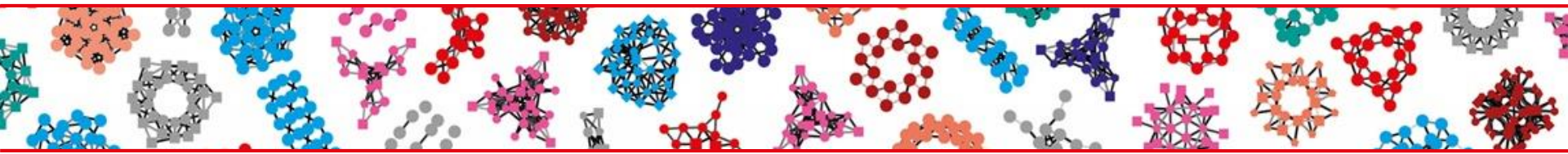
Types of data



Distinguishing features



Overview



01 — Introduction to neXtProt

02 — Data model

03 — **SPARQLing in neXtProt**

04 — Summary

SPARQL user interfaces

<https://www.nextprot.org/>

The screenshot shows the nextprot.org website. The navigation bar includes 'Tools', 'Portals', 'Download', 'Help', 'About', and 'Contact'. A 'Login' button with the SIB logo is in the top right. A search bar is centered below the navigation, with a 'Q Search' button. Below the search bar, there are four main content areas: 'Getting started', 'Data sources', 'News', and 'Release 2019-08-22'. The 'Getting started' section lists various search and analysis tools. The 'Data sources' section features logos for The Human Protein Atlas, SRAtlas, IntAct, Boee, Peptide Atlas, COSMIC, UniProt-GOA, UniProt, gnomAD, and GlyConnect. The 'News' section lists recent publications. The 'Release 2019-08-22' section includes a donut chart titled 'Protein existence in neXtProt' showing the distribution of protein evidence levels.

1 → Simple search
Advanced search (SPARQL)
My Lists
My Queries
Peptide uniqueness checker
Protein digestion
Programmatic access
API
SPARQL endpoint
SNORQL search

2 →

3 → SPARQL endpoint : <https://api.nextprot.org/sparql>
(SERVICE <https://sparql.nextprot.org>)

https://www.nextprot.org | Swiss Institute of Bioinformatics | Legal disclaimer | Data release 2019-08-22 | Application release v2.23.2 | For developers

neXtProt Advanced Search

Find answers to complex questions:

1. Proteins whose genes are on chromosome 13 **and** are associated with a disease
2. Proteins with at least one variant of the types "A->R" **or** "R->A"
3. Proteins with alternative **acetylation or Ubl conjugation** (SUMO or Ubiquitin) at the **same positions**
4. Proteins with at least two antibodies available from **Human Protein Atlas** that have associated tissue expression annotations from **immunohistochemistry** studies

These queries cannot be performed using the simple search.

neXtProt Advanced Search

<https://www.nextprot.org/proteins/search?mode=advanced>

Retrieve **entries** matching criteria

The screenshot displays the neXtProt Advanced Search interface. At the top, there is a navigation bar with the neXtProt logo, menu items (Tools, Portals, Download, Help, About, Contact), a Login button, and the SIB logo. Below the navigation bar, the search mode is set to "Advanced search (SPARQL)". A search input field contains the number "1". Below the search bar, there is a dropdown menu for "proteins" and a "Search" button. A "Tags" dropdown is set to "search in 144 queries (ex: liver)". A "New Query" button is also visible. The main content area shows a list of protein queries with their corresponding IDs and eye icons. The queries are:

- Proteins with more than 12 WD repeats (ie: with at least two beta-propellers) (NXQ_00019)
- Proteins with at least 2 HPA antibodies whose genes are located on chromosome 21 and that are highly expressed at IHC level in heart (NXQ_00020)
- Proteins with at least one HPA antibody that are located in the peroxisome (NXQ_00021)
- Proteins with no function annotated (NXQ_00022)
- Proteins that are involved in transport and located in a membrane and that are not glycosylated (experimentally or predicted) (NXQ_00023)
- Proteins with more than 10 reported interactions (of 'gold' quality) (NXQ_00024)
- Proteins with at least 50 interactors that are not associated with a disease (NXQ_00025)
- Proteins interacting with at least one protein which is located in the mitochondrion (NXQ_00026)
- Proteins with one or more glycosylation sites reported in PubMed:20570859 or PubMed:14760718 (NXQ_00027)

At the bottom of the interface, there is a footer with the copyright information: "© 2011 - 2019 SIB Swiss Institute of Bioinformatics | Le" and a link to "the model" or "contact us" for help with queries.

Advanced Search results

https://www.nextprot.org/proteins/search?mode=advanced&queryId=NXQ_00022

NXQ_00022 **Proteins** with no function annotated

The screenshot displays the nextprot website interface. At the top, there is a navigation bar with the nextprot logo, menu items (Tools, Portals, Download, Help, About, Contact), a Login button, and the SIB logo. Below the navigation bar, the search interface is shown, with 'Advanced search (SPARQL)' selected. A SPARQL query is entered in a text area, and a search button is visible. Below the query, there is a list of search results. The first result is 'Uncharacterized protein KIAA0040 (KIAA0040) [NX_Q15053]', and the second is 'Coiled-coil domain-containing protein 27 (CCDC27) [NX_Q2M243]'. Each result includes a brief description and various protein attributes. On the left side, there is a 'FILTERS' section with a dropdown menu and a '0' indicator. At the bottom, there is a footer with copyright information and a 'For developers' link.

nextprot Tools Portals Download Help About Contact Login SIB

Simple search Advanced search (SPARQL)

```
1 #Proteins with no function annotated
2 select distinct ?entry where {
3   ?entry :isoform ?iso.
4   filter not exists { ?iso :functionInfo ?_. }
5   filter not exists { ?iso :catalyticActivity ?_. }
6   filter not exists { ?iso :transportActivity ?_. }
7   filter not exists { ?iso :pathway ?_. }
8   filter not exists {
9     ?iso :function / :term ?fterm .
10    filter(?fterm != cv:GO_0005524 && ?fterm !=
11    cv:GO_0000287 && ?fterm != cv:GO_0005515 && ?fterm != cv:GO_0042802
12    && ?fterm != cv:GO_0008270 && ?fterm !=
13    cv:GO_0051260 && ?fterm != cv:GO_0005509
14    && ?fterm != cv:GO_0003676 && ?fterm !=
```

eg.: Search for MSH6 in proteins, Search for author Doolittle in publications, Search for liver in terms

Show 1 to 50 of 2222 1 of 45 page(s) Summary Details Download Sort score

FILTERS

For proteins with:

- 3D structure (32)
- Disease (40)
- Expression profile (1962)
- Mutagenesis (14)
- Proteomics (1648)

- Uncharacterized protein KIAA0040 (KIAA0040) [NX_Q15053] | [Download](#)
No functional information for this protein.
Chromosomal location: 1q25.1 Isoforms: 1 PTMs: 0 Sequence length: 99 Variants: 52
Disease: no Expression: yes Mutagenesis: no Proteomics: yes Structure: no Proteins existence: Evidence at protein level
- Coiled-coil domain-containing protein 27 (CCDC27) [NX_Q2M243] | [Download](#)
No functional information for this protein.
Chromosomal location: 1p36.32 Isoforms: 1 PTMs: 0 Sequence length: 656 Variants: 460
Disease: no Expression: yes Mutagenesis: no Proteomics: yes Structure: no Proteins existence: Evidence at protein level
- Complement component receptor 1-like protein (CR1) [NX_Q2VPA4] | [Download](#)

© 2011 - 2019 SIB Swiss Institute of Bioinformatics | Legal disclaimer | Data release 2019-01-11 | Application release v2.22.8 | For developers

SPARQL query examples

Tools ▾ Portals ▾ Download Help ▾ About ▾ Contact

Simple search Advanced search (SPARQL)

proteins ▾ 1 + ≡ ×

e.g.: Search for MSH6 in proteins, Search for author Doolittle in publications, Search for liver in terms

1 → Tags ▾ search in 144 queries (ex: liver) ← 2 New Query ×

Proteins phosphorylated and located in the cytoplasm NXQ_00001

Tags: PTM, tutorial, subcellular location, phosphorylation, cellular component

Description: This query uses the term "Phosphoprotein" (via keyword KW-0597) and the UniProt subcellular location ontology term "Cytoplasm" (SL-0886). Using the "childOf" allows to select for subcellular locations that are, in these ontologies, children of cytoplasm like for example "Cell cortex".

SPARQL: `select distinct ?entry where { values ?cytoloc {cv:SL-0086 cv:GO_0005737} # SL and GO values for cytoplasm ?entry :isoform ?iso. ?iso :keyword / :term cv:KW-0597. # Phosphorylated ?iso :cellularComponent ?loc. ?loc :term /:childOf ?cytoloc. filter not exists {?loc :negativeEvidence ?negev} # No negative localization evidence }`

Query Id: NXQ_00001

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

Proteins with 7 transmembrane regions NXQ_00003

Proteins highly expressed in brain but not expressed in testis NXQ_00004

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

neXtProt SNORQL Search

<https://snorql.nextprot.org/>

Queries to retrieve **any information** matching criteria

The screenshot displays the neXtProt SNORQL search interface. At the top, the navigation bar includes the neXtProt SNORQL logo, links for Services, Help, About, and Contact us, and a Login button. The main interface is divided into two primary sections: a query editor on the left and a results list on the right.

Query Editor: The top of the editor shows the SPARQL endpoint: `https://api.nextprot.org/sparql`. Below this, there is a text area containing a SPARQL query: `PREFIX... SELECT DISTINCT * WHERE { ?s ?p ?o } LIMIT 10`. A "Query title" field is present below the query, with a placeholder text: "Here you can write your SPARQL queries". At the bottom of the editor, there is a "Go" button, a "Reset" button, a dropdown menu set to "html", and a search input field containing the text "eg, peroxisome, liver".

Results List: The right side of the interface shows a list of search results, each with a title and a set of filter tags. The results are as follows:

- NXQ_00124** - What are the 25 most frequent families with member count. Tags: family, snorql-only, tutorial.
- NXQ_00125** - Domains that are entirely covered by 3D structures. Tags: 3Dstructure, domain, PDB, region, snorql-only, tutorial.
- NXQ_00126** - Peptides that are potential neo N-termini from undescribed isoforms. Tags: alternative splicing, peptide, snorql-only, tutorial.
- NXQ_00131** - Terms of controlled vocabularies containing some word(s). Tags: snorql-only, tutorial, vocabulary.
- NXQ_00134** - Proteins with the longest poly-AA stretches. Tags: compositional bias, sequence, snorql-only, tutorial.
- NXQ_00141** - Human proteins highly expressed in brain and observed in a PDB structure involving a virus protein. Tags: expression, federated query, interaction, PDB, snorql-only, tutorial, UniProt.
- NXQ_00148** - Variants with normal "ubiquitin-protein transferase activity" and decreased or increased binding to UBE2D1 (example: BRCA1-p.Ile89Thr). Tags: GO, interaction, phenotype, snorql-only, tutorial, variant.
- NXQ_00149** - BRCA1 variants with at least 5 different Severe phenotypes. Tags: phenotype, snorql-only, tutorial, variant.
- NXQ_00209** - Glycosylation sites and cross links positions on SwissProt canonical isoforms. Tags: glycosylation, PTM, snorql-only, tutorial.

At the bottom of the page, a footer contains the following text: "© 2011 - 2019 SIB Swiss Institute of Bioinformatics | Legal disclaimer | Data release 2019-08-22 | Application release v2.9.0 | For developers".

neXtProt SNORQL Search

NXQ_00125 **Domains** that are entirely covered by 3D structures

The screenshot displays the neXtProt SNORQL search interface. At the top, the navigation bar includes the neXtProt logo, "Services", "Help", "About", "Contact us", and "Login".

The main content area is divided into two columns. The left column contains a SPARQL endpoint URL: `https://api.nextprot.org/sparql`. Below this is a "PREFIX..." section with a SPARQL query:

```
select distinct ?domlabel where {
  ?entry :isoform ?isoform.
  ?isoform :pdbmapping ?struc.
  ?struc :start ?s1 ; :end ?s2.
  ?isoform :domain ?dom.
  ?dom :start ?d1 ; :end ?d2.
  ?dom :term/rdfs:label ?domlabel
  filter ((?s1 <= ?d1) && (?s2 >= ?d2))
}
```

The right column features a "Tags" dropdown menu and a search bar containing "Filter sparql examples". Below the search bar is a list of queries with associated tags:

- mutagenesis tutorial
- NXQ_00124 - What are the 25 most frequent families with member count (family snorql-only tutorial)
- NXQ_00125 - Domains that are entirely covered by 3D structures** (3D structure domain PDB region snorql-only tutorial)
- NXQ_00126 - Peptides that are potential neo N-termini from undescribed isoforms (alternative splicing peptide snorql-only tutorial)
- NXQ_00127 - Proteins with at least one known SUMOylation site (PTM SUMO tutorial)

An overlay window is positioned in the foreground, showing a search for "eg. peroxisome, liver". It includes a "html" dropdown, "Go" and "Reset" buttons, and a search icon. Below the search bar, it says "Found a bug? Improve this query!". The results section indicates "Query time is 0.25[s] for 538 rows" and lists domain labels:

- domlabel
- "Ig-like C2-type domain"^^xsd:string
- "protein kinase domain"^^xsd:string
- "SH3 domain"^^xsd:string
- "PDZ domain"^^xsd:string
- "Ig-like domain"^^xsd:string
- "SAM-dependent MTase PRMT-type domain"^^xsd:string
- "disintegrin domain"^^xsd:string
- "peptidase M12B domain"^^xsd:string
- "SAM-dependent MTase C5-type domain"^^xsd:string
- "F5/8 type C domain"^^xsd:string

At the bottom of the page, a footer contains the text: "© 2011 - 2019 SIB Swiss Institute of Bioinformatics | Legal disclaimer | Data release 2019-01-11 | Application release v2.6.0 | For developers".

Documentation

<https://snorql.nextprot.org/help/doc/introduction>



The screenshot shows the neXtProt SNORQL website interface. The top navigation bar includes 'neXtProt SNORQL', 'Services', 'Help', 'About', 'Contact us', and 'Login'. The left sidebar lists 'GENERALITIES' (Introduction, Exercises, Exercise1-4) and 'RDF ENTITIES' (AbsorptionMax, AbsorptionNote, ActiveSite, ActivityRegulation, Allergen, AnnotationType, AntibodyMapping). The main content area is titled 'neXtProt' and describes it as a human proteome database. It features a 'References' section with boxes for :Identifier, :Publication, and :Xref. The central part of the page displays a data model diagram with the following components and relationships:

- :Gene** (blue box): :name, :chromosome, :band, :strand, :begin, :end
- :Entry** (blue box): :name, :family, :existence, :history
- :Isoform** (yellow box): linked to :Entry via :isoform, and to :ProteinSequence via :sequence
- :Annotation** (green box): rdf:type (SomeAType), :quality, :negative, :position (:start :end), :description
- :Evidence** (yellow box): :quality, :isoformSpecificity, :evidenceCode (Term), :reference (Publication), :reference (Xref)
- :Term** (green box): rdf:type (:SomeCv), rdfs:label
- :Entry** (green box): linked to :Annotation via :interactant
- :Isoform** (green box): linked to :Annotation via :evidence
- :Evidence** (green box): linked to :Term via :term, and to :ExperimentalContext via :experimentalContext
- :ProteinSequence** (yellow box): linked to :Entry via skos:exactMatch
- :ExperimentalContext** (yellow box): :detectionMethod (Term), :tissue (Term), :developmentalStage (Term), :cellLine (Term), :disease (Term), :organelle (Term), :metadata (Publication)

Relationships shown in the diagram include: :gene (from :Entry to :Gene), :isoform (from :Entry to :Isoform), :sequence (from :Isoform to :ProteinSequence), :evidence (from :Annotation to :Evidence), :term (from :Evidence to :Term), :interactant (from :Entry to :Annotation), and :experimentalContext (from :Evidence to :ExperimentalContext).

© 2011 - 2019 SIB Swiss Institute of Bioinformatics | Legal disclaimer | Data release 2019-08-22 | Application release v2.9.0 | For developers

Documentation for :disease

<https://snorql.nextprot.org/help/entity/Disease>

Disease Values▼ 37028

Description

Involvement in disease: disease(s) associated with protein defect(s)

`:Isoform` `:disease`
`:Isoform` `:generalAnnotation`
`:Isoform` `:medical`

Subjects
for
:disease

:Disease

`:entryAnnotationId` `xsd:string` 37028 Example: AN_A0A0B4J1V
`:evidence` `:Evidence` 70836
`:quality` `:QualityQualifier` 37028
`:term` `:UniprotDiseaseCv` 14856
`rdfs:comment` `xsd:string` 37028 Example: 3MC syndrome

Objects
for
:disease

Example

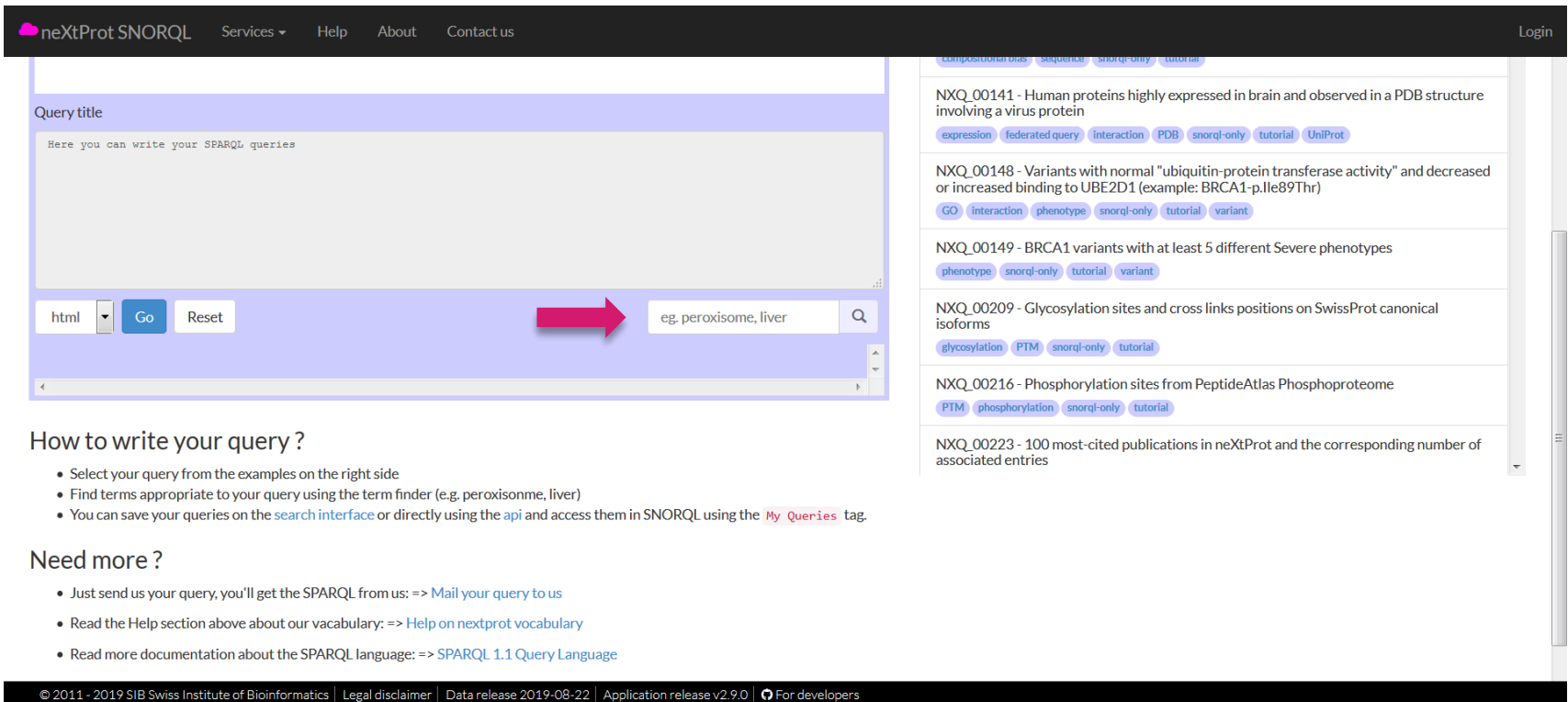
```
annotation:NX_P29597-1-AN_P29597_001466 :entryAnnotationId "AN_P29597_001466" .
annotation:NX_P29992-1-AN_P29992_XR_2718683 :evidence evidence:300002718683 .
annotation:NX_P29597-1-AN_P29597_001466 :quality :GOLD .
annotation:NX_P29597-1-AN_P29597_001466 :term cv:DI-02224 .
annotation:NX_P29597-1-AN_P29597_001466 rdfs:comment "Immunodeficiency 35 (IMD35) [MIM:611521]: A primary immunodeficiency characterized by recurrent skin abscesses, pneumonia, and highly elevated serum IgE. The disease is caused by mutations affecting the gene represented in this entry." .
```

Values

Searching for terms

<https://snorql.nextprot.org/>

Find controlled vocabulary terms quickly



neXtProt SNORQL Services ▾ Help About Contact us Login

Query title

Here you can write your SPARQL queries

html Go Reset

eg, peroxisome, liver

Compositional bias sequence snorql-only tutorial

NXQ_00141 - Human proteins highly expressed in brain and observed in a PDB structure involving a virus protein

expression federated query interaction PDB snorql-only tutorial UniProt

NXQ_00148 - Variants with normal "ubiquitin-protein transferase activity" and decreased or increased binding to UBE2D1 (example: BRCA1-p.Ile89Thr)

GO interaction phenotype snorql-only tutorial variant

NXQ_00149 - BRCA1 variants with at least 5 different Severe phenotypes

phenotype snorql-only tutorial variant

NXQ_00209 - Glycosylation sites and cross links positions on SwissProt canonical isoforms

glycosylation PTM snorql-only tutorial

NXQ_00216 - Phosphorylation sites from PeptideAtlas Phosphoproteome

PTM phosphorylation snorql-only tutorial

NXQ_00223 - 100 most-cited publications in neXtProt and the corresponding number of associated entries

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](#)

© 2011 - 2019 SIB Swiss Institute of Bioinformatics | [Legal disclaimer](#) | [Data release 2019-08-22](#) | [Application release v2.9.0](#) | [For developers](#)

Tips

NXQ_00001 - Proteins phosphorylated and located in the cytoplasm

cellular component phosphorylation PTM subcellular location tutorial

PREFIX..

```
select distinct ?entry where {  
  values ?cytoloc {cv:SL-0086 cv:GO_0005737} # SL and GO values for cytoplasm  
  ?entry :isoform ?iso.  
  ?iso :keyword / :term cv:KW-0597. # Phosphorylated  
  ?iso :cellularComponent ?loc .  
  ?loc :term /:childOf ?cytoloc.  
  filter not exists {?loc :negativeEvidence ?negev} # No negative localization evidence  
}
```

1

3

2

1

For cellular localization, use both controlled vocabularies (GO-CC and UniProtKB subcellular locations)

2

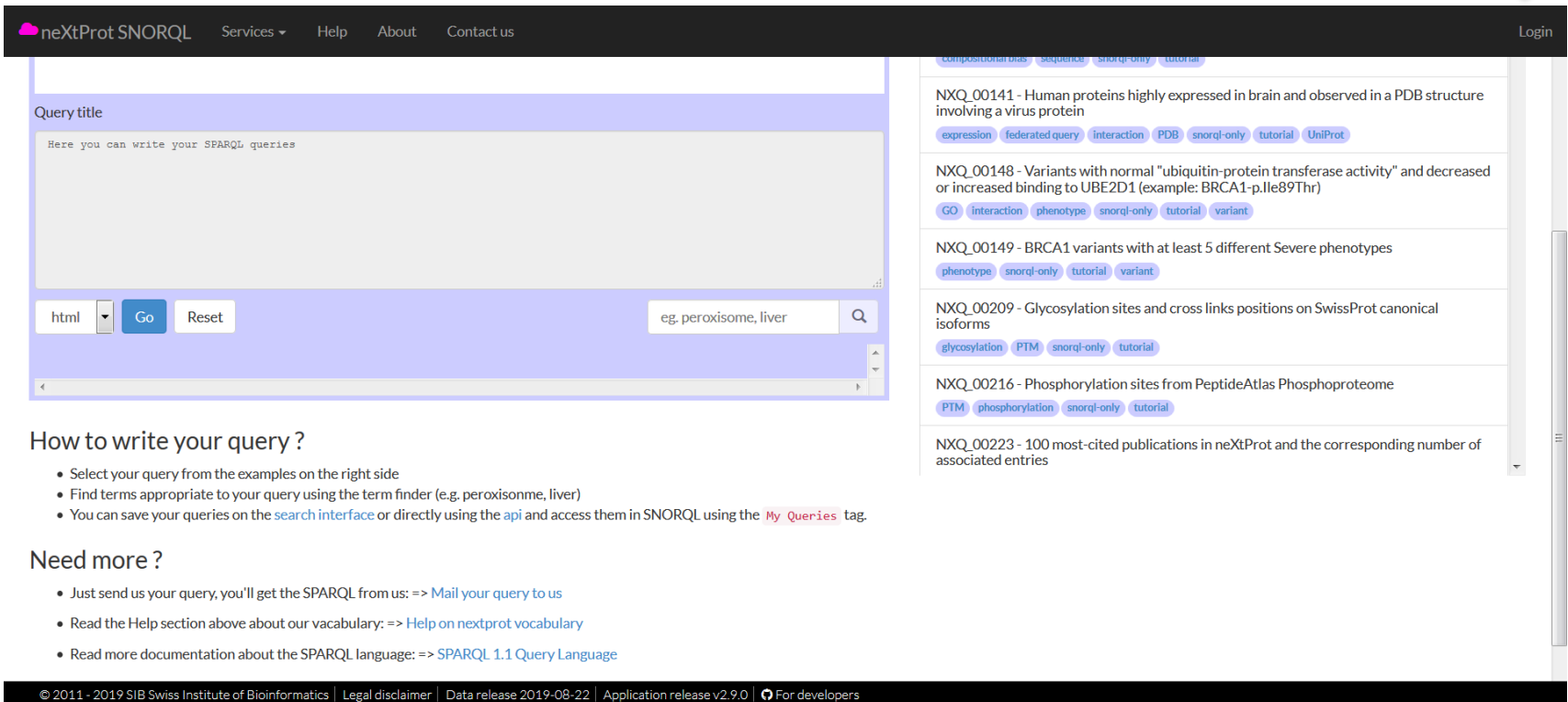
For hierarchical ontologies (i.e. GO, UniProtKB subcellular locations, neXtProt anatomy), use `:childOf` to include annotations to child terms

3

Remember to exclude negative annotations

Saving queries

<https://snorql.nextprot.org/>



The screenshot shows the neXtProt SNORQL web interface. At the top, there is a navigation bar with the neXtProt logo, 'Services', 'Help', 'About', and 'Contact us' links, and a 'Login' button in the top right corner. A red arrow points to the 'Login' button. Below the navigation bar is a large text area for writing SPARQL queries, with a placeholder text 'Here you can write your SPARQL queries'. Below this area are buttons for 'html', 'Go', and 'Reset', and a search input field containing 'eg, peroxisome, liver'. To the right of the query editor is a list of example queries, each with a title and a set of tags. The examples include: 'NXQ_00141 - Human proteins highly expressed in brain and observed in a PDB structure involving a virus protein' (tags: expression, federated query, interaction, PDB, snorql-only, tutorial, UniProt); 'NXQ_00148 - Variants with normal "ubiquitin-protein transferase activity" and decreased or increased binding to UBE2D1 (example: BRCA1-p.Ile89Thr)' (tags: GO, interaction, phenotype, snorql-only, tutorial, variant); 'NXQ_00149 - BRCA1 variants with at least 5 different Severe phenotypes' (tags: phenotype, snorql-only, tutorial, variant); 'NXQ_00209 - Glycosylation sites and cross links positions on SwissProt canonical isoforms' (tags: glycosylation, PTM, snorql-only, tutorial); 'NXQ_00216 - Phosphorylation sites from PeptideAtlas Phosphoproteome' (tags: PTM, phosphorylation, snorql-only, tutorial); and 'NXQ_00223 - 100 most-cited publications in neXtProt and the corresponding number of associated entries'.

Query title

Here you can write your SPARQL queries

html Go Reset

eg, peroxisome, liver

Compositional bias sequence snorql-only tutorial

NXQ_00141 - Human proteins highly expressed in brain and observed in a PDB structure involving a virus protein

expression federated query interaction PDB snorql-only tutorial UniProt

NXQ_00148 - Variants with normal "ubiquitin-protein transferase activity" and decreased or increased binding to UBE2D1 (example: BRCA1-p.Ile89Thr)

GO interaction phenotype snorql-only tutorial variant

NXQ_00149 - BRCA1 variants with at least 5 different Severe phenotypes

phenotype snorql-only tutorial variant

NXQ_00209 - Glycosylation sites and cross links positions on SwissProt canonical isoforms

glycosylation PTM snorql-only tutorial

NXQ_00216 - Phosphorylation sites from PeptideAtlas Phosphoproteome

PTM phosphorylation snorql-only tutorial

NXQ_00223 - 100 most-cited publications in neXtProt and the corresponding number of associated entries

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](#)

© 2011 - 2019 SIB Swiss Institute of Bioinformatics | [Legal disclaimer](#) | [Data release 2019-08-22](#) | [Application release v2.9.0](#) | [For developers](#)

Federated SPARQL queries

Used to explore the data in neXtProt and other resources

federated query▼	Filter sparql examples
NXQ_00094 - Proteins which are targets of antipsychotic drugs and highly expressed in brain drug expression federated query tutorial	
NXQ_00096 - Proteins which are targets of drugs for cardiac therapy drug federated query tutorial	
NXQ_00139 - Protein kinases which are drug targets according to ChEMBL ChEMBL drug federated query tutorial	
NXQ_00140 - Proteins that interact with viral proteins federated query interaction PPI tutorial UniProt	
NXQ_00141 - Human proteins highly expressed in brain and observed in a PDB structure involving a virus protein expression federated query interaction PDB snorql-only tutorial UniProt	
NXQ_00246 - Proteins which are enzymes catalyzing a reaction involving lipids enzyme federated query tutorial	
NXQ_00253 - Human pathways in which at least one protein is mitochondrial GOLD federated query pathway snorql-only subcellular location tutorial	
NXQ_00254 - Proteins with associated pathways in WikiPathways federated query pathway snorql-only tutorial	

Data visualization

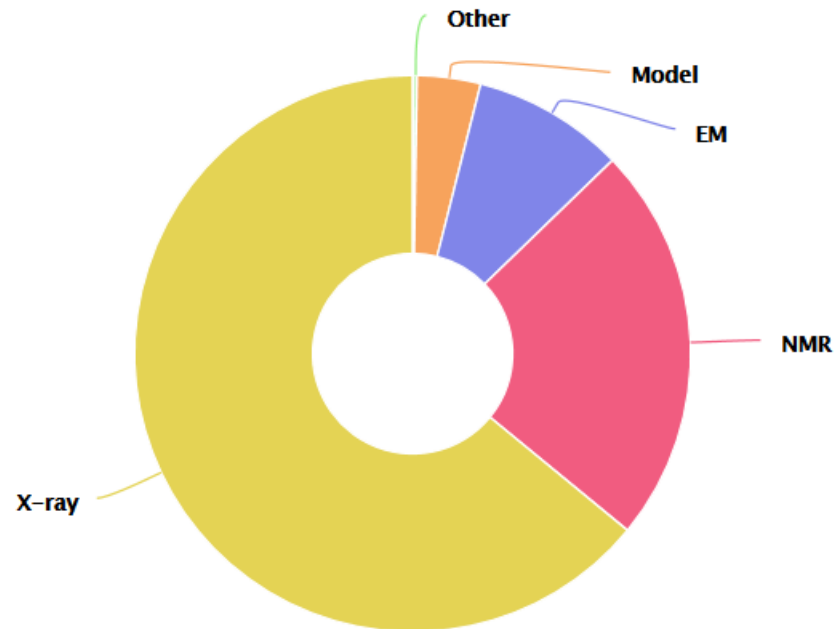
Use javascript to see SPARQL query results using Highcharts.com

```
1 <html>
2 <head>
3   <script src="http://ajax.googleapis.com/ajax/libs/jquery/1.11.2/jquery.min.js"></script>
4   <script src="http://code.highcharts.com/highcharts.js"></script>
5   <script src="http://code.highcharts.com/modules/exporting.js"></script>
6   <script src="https://cdn.rawgit.com/calipho-sib/nextprot-js/v0.0.23/dist/nextprot.min.js"></script>
7 </head>
8 <body>
9   <div id="plot" style="min-width: 310px; height: 400px; margin: 0 auto"></div>
10 </body>
11 <script type="text/javascript">
12   // Create an instance of nextprot API
13   var applicationName = "demo app for using SPARQL to view breakdown of experimental methods used in PDB structures";
14   var clientInformation = "calipho group at SIB";
15   var nx = new Nextprot.Client(applicationName, clientInformation);
16   //Define your sparql
17   var proteinsBy3DMethod = 'select distinct ?method (count (distinct ?entry) as ?entrycount) ' + SPARQL query
18   'where {?entry :isoform / :pdbMapping / :method ?method.} order by ?entrycount ';
19   //Execute the sparql and retrieve result
20   nx.executeSparql (proteinsBy3DMethod).then(function (result){ Retrieve results
21     var seriesData = [];
22     result.results.bindings.map(function (data) {
23       seriesData.push([data.method.value, parseInt (data.entrycount.value)]); //gets number of entries
24     });
25     //Draw the plot
26     $('#plot').highcharts({chart: {type: 'pie', options3d: { enabled: true, alpha: 45 }}, Plot results
27       title: { text: 'Experimentally determined 3D structures in neXtProt' },
28       plotOptions: {pie: { innerSize: 100, depth: 45 } },
29       series: [{name: 'Count',data: seriesData }]});
30   });
31 });
32 </script>
33 </html>
```

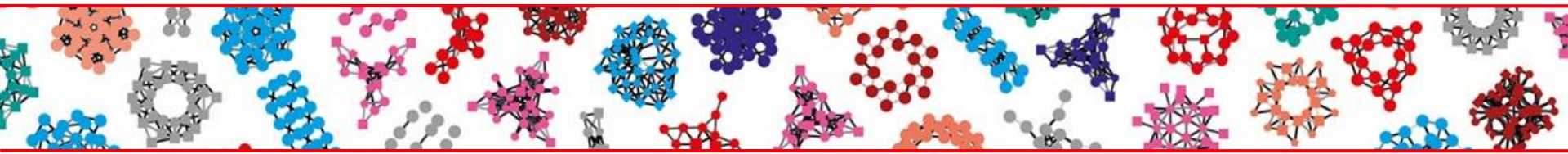
Visualizing SPARQL query results

Automatically updated at each data release

Experimentally determined 3D structures in neXtProt



Overview



01 —• Introduction to neXtProt

02 —• Data model

03 —• Uses of SPARQL

04 —• **Summary**

Querying neXtProt using SPARQL

Access

SPARQL endpoint

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

Advanced search

<https://www.nextprot.org/proteins/search?mode=advanced>

SNORQL

<https://snorql.nextprot.org/>



Data model

<https://www.nextprot.org/help/data-model>



Help > RDF entities

<https://snorql.nextprot.org/help/doc/introduction>



Sample queries

Found in the Advanced search (SPARQL) and SNORQL interfaces



User guide

<https://doi.org/10.7490/f1000research.1116829.1>

Uses of SPARQL

- **Answer complex biological questions**

Execute complex queries, across resources

- **Retrieve data**

For benchmarking, display in web sites, etc.


- **Biocuration**

- **Quality control**

For spot checks and global checks of the data

Federated query with OrthoDB

Mouse ortholog(s) found at **Tetrapoda level** for human myosin light chain 6B (MYL6B) (neXtProt entry **NX_P14649**)

 neXtProt SNORQL Services ▾ Help About Contact us

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

PREFIX...

```
prefix odb: <http://purl.orthodb.org/>
select * where {
?entry :isoform ?isoform. filter (?entry=<http://nextprot.org/rdf/entry/NX_P14649>).
service <http://sparql.orthodb.org/sparql>
{?gene rdfs:seeAlso ?entry; odb:memberOf ?og.
?og odb:ogBuiltAt [up:scientificName ?clade]
; odb:hasMember ?ortholog.
?ortholog odb:name ?ortholog_name; up:organism/a/up:scientificName 'Mus musculus'
filter (?clade='Tetrapoda')}
}}
```

Query time is 0.45[s] for 2 rows

	og	clade	ortholog	ortholog_name
3498>	< http://purl.orthodb.org/odbgroup/273280at32523 >	"Tetrapoda"	< http://purl.orthodb.org/odbgene/10090_0_0033cd >	"My16b"
3498>	< http://purl.orthodb.org/odbgroup/273280at32523 >	"Tetrapoda"	< http://purl.orthodb.org/odbgene/10090_0_0031b7 >	"My16"

Group [273280at32523](#) at Tetrapoda level
myosin light chain 6B

281 genes in 135 species >>>

Acknowledgements



neXtProt

Directors

Amos Bairoch, Lydie Lane

Biocurators

Paula Duek

Developers

Pierre-André Michel, Alain Gateau, Valentine Rech de Laval

Researcher

Mathieu Schaeffer, Kasun Samarasinghe, Vimel Rattina

Quality assurance

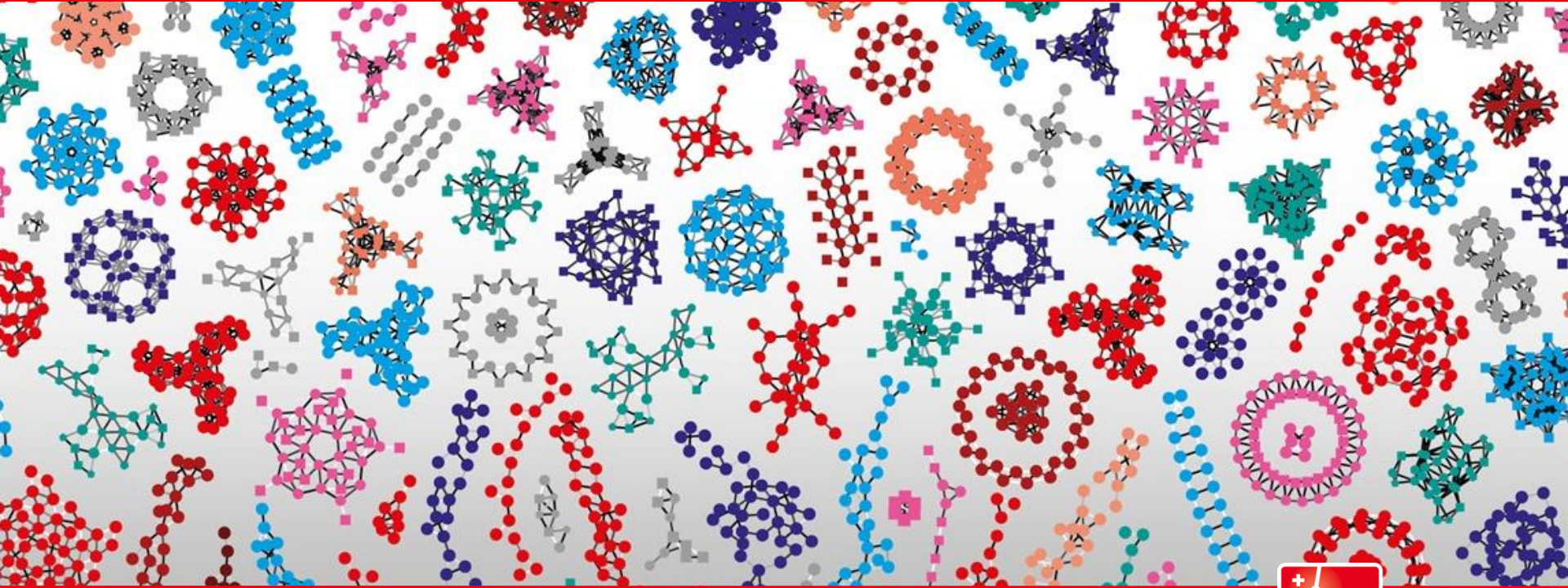
Monique Zahn

Web: <https://www.nextprot.org/>

Twitter: @neXtProt_news

ResearchGate: neXtProt project

Photograph taken by Chris James



Swiss Institute of
Bioinformatics

Thank you!

E-mail: monique.zahn@sib.swiss



www.sib.swiss