

ProtVista

Open-Source Protein Feature Visualisation with
reusable Web Components

Aurélien Luciani



www.uniprot.org



UniProt

A comprehensive, high-quality and publicly accessible resource of protein sequence and functional information

It includes, among other things:

- Physiological protein function including subcellular location, pathways, reactions, interactions and involvement in disease
- Sequence annotation of domains, PTMs, mutations, disease variants and much more



European Bioinformatics Institute (EMBL-EBI),
Hinxton, Cambridge, UK

Find your protein

UniProtKB ▾ Examples: Insulin, APP, Human, P05067, organism_id:9606 Advanced | List Search

UniProt is the world's leading high-quality, comprehensive and freely accessible resource of protein sequence and functional information. [Cite UniProt](#)

⚠️ Our Proteomes and UniProtKB/TrEMBL resources are undergoing a significant transition. Please read our help page, view affected entries and [proteomes](#), or contact us with any questions.

The Protein Information Resource (PIR) logo, featuring the letters "PIR" in white on a red circular background.

Proteins
UniProt Knowledgebase
Reviewed (Swiss-Prot)
Unreviewed (TrEMBL)

Species
Proteomes
Protein sets for species with sequenced genomes from across the tree of life

Protein Clusters
UniRef
SIB
Clusters at 100% identity

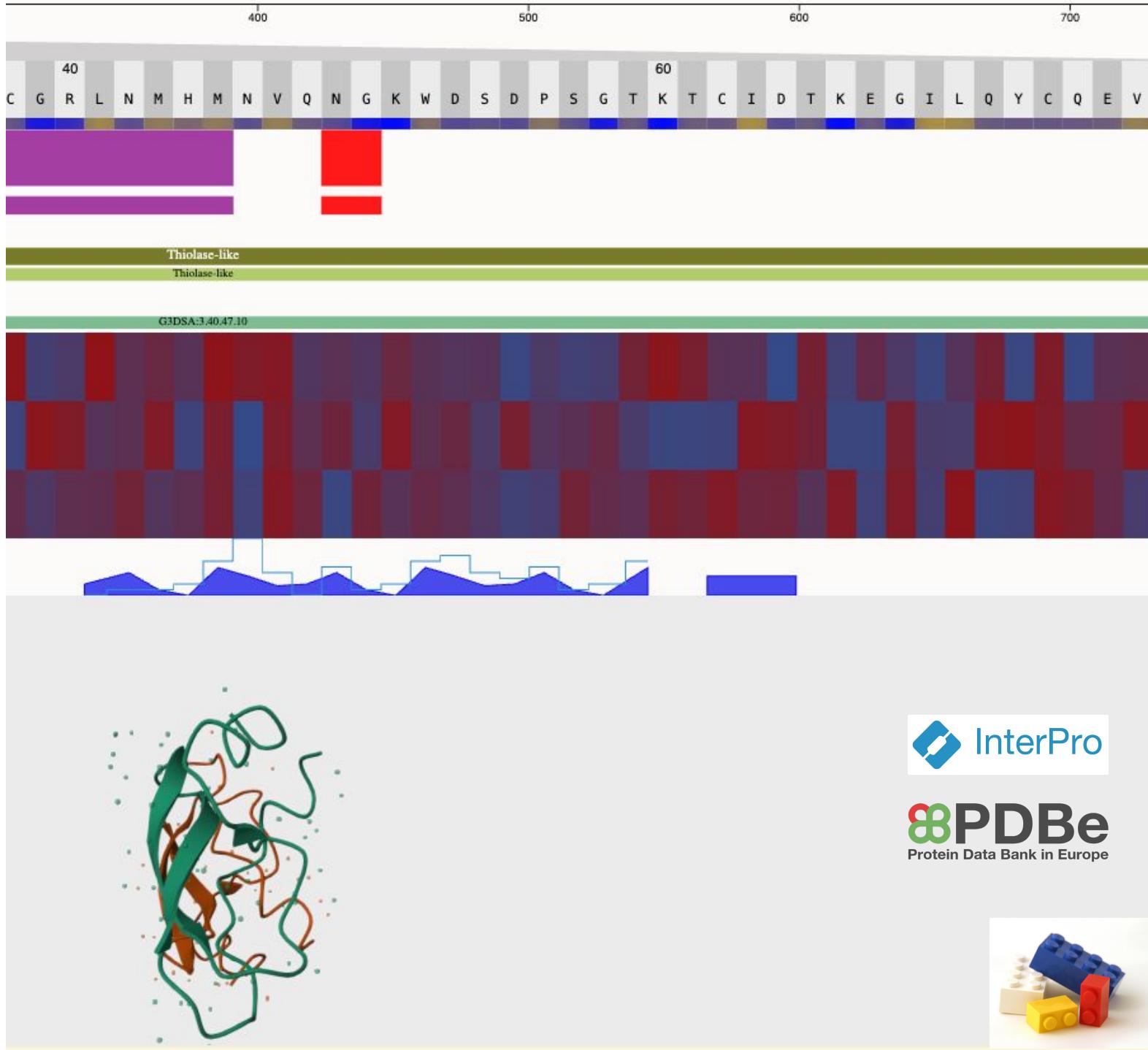
Sequence archive
UniParc
Non-redundant archive of publicly available protein sequences seen across different databases

Protein Information Resource (PIR), Washington DC and Delaware, USA

SIB Swiss Institute of Bioinformatics (SIB), Geneva, Switzerland

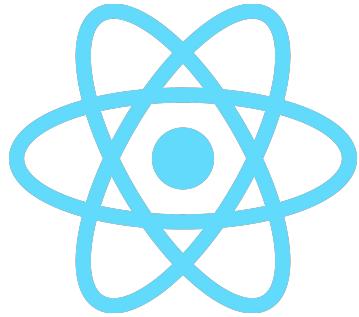
Nightingale

- An **open-source** visualization library of **standard web components**
- Designed for representing protein-related data
- **Composable** architecture facilitates vertical stacking of tracks, aiding visual comparison across datasets
- **Interoperable** with other standard components, regardless of the underlying framework
- **Compatible** with any component following the **Nightingale APIs**



Building using the standards: Web Components

- Combination of standard APIs: **Custom elements + Shadow DOM + HTML templates**
- **Not dependent** on any framework



React
and related

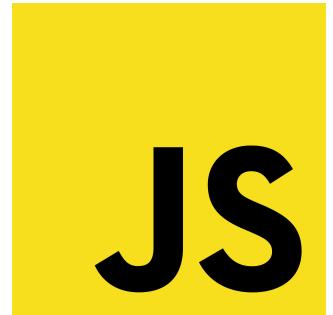


Angular



Vue.js

...



Vanilla JS

- **Limited** library dependencies
- Set of components, as **independent packages** under the `@nightingale-elements` scope on NPM

Same blocks, different ways



Visualization showing family and domains in UniProt's entry page of [P05067](#) using minimal number of tracks as needed

Family & Domainsⁱ

Features
Showing features for regionⁱ, domainⁱ, compositional biasⁱ, motifⁱ.

1 100 200 300 400 500 600 700 770

TYPE All ID POSITION(S) DESCRIPTION

+ Region 28-123 GFLD subdomain PROSITE-ProRule Annotation

+ Domain 28-189 E1 PROSITE-ProRule Annotation

Tools Add

Tools Add

Entry matches to this proteinⁱ

Feature Display Mode Summary Full

100 200 300 400 500 600 700 770

Families

Representative families

F Amyloid_glyco - Amyloidogenic domain
PANTHER: ALZHEIMER'S DISEASE BETA-PRINTS: AMYLOIDA4 - AMYLOIDA4

Domains

Representative domains

H Amyloid_glyco_heparin-bd_sf - /
SSF: A heparin-binding domain
CATHGENE3D: Amyloidogenic glycoprotein

D Amyloid_glyco_extra - Amyloidogenic domain
SMART: A4_EXTRA - amyloid A4
PROFILE: APP_E1 - Amyloid precursor protein

D Amyloid_glyco_heparin-bd - Amyloidogenic domain
PFAM: APP_N - Amyloid A4 N-terminal domain

H Amyloid_Cu-bd_sf - Amyloidogenic domain
SSF: Amyloid beta a4 protein copper binding domain
CATHGENE3D: Amyloidogenic glycoprotein

D Amyloid_glyco_Cu-bd - Amyloidogenic domain
PFAM: APP_C - Amyloid A4 C-terminal domain

A4_EXTRA
APP_E1
APP_N
APP_C

InterPro's protein viewer using Nightingale components for [P05067](#)

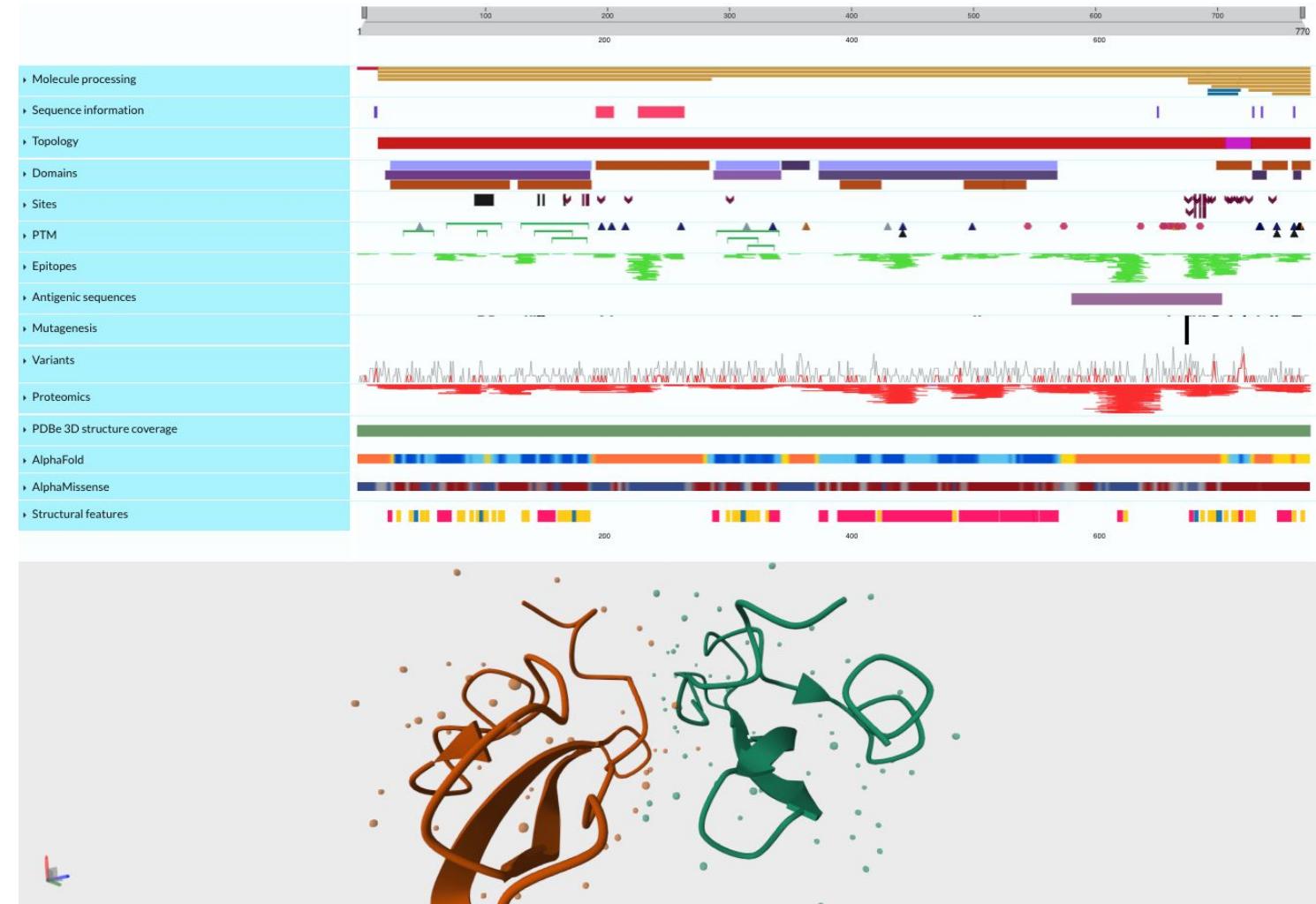


PDBe is an other user of Nightingale components

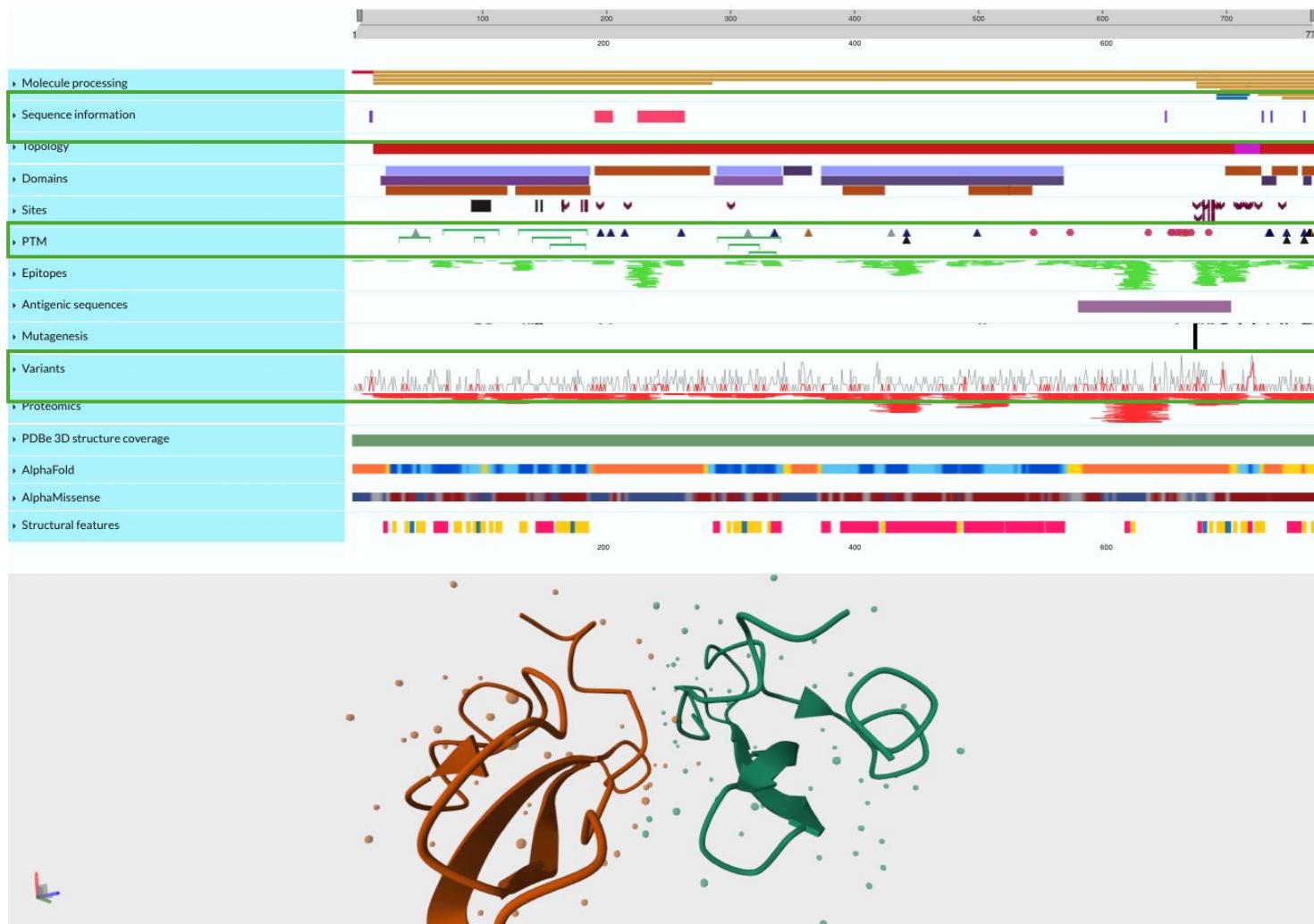


UniProt's ProtVista visualization

- **Turnkey** combination of tracks
- **Chooses** and aligns various tracks to best fit our needs
- Uses **different components** of Nightingale
- Uses UniProt's **data source**
- **Customizes** tooltips as needed
- Is itself a **Web Component**



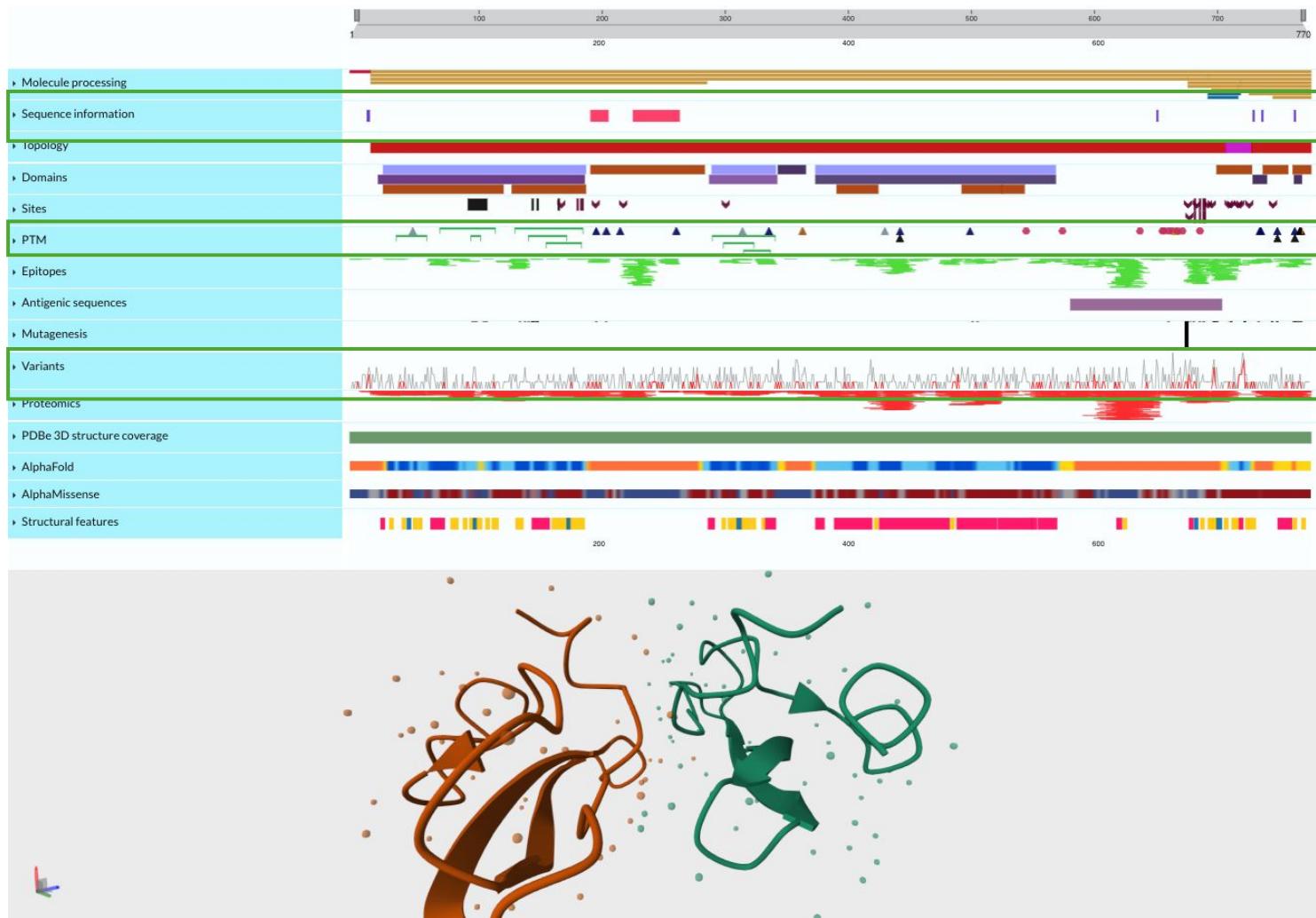
Human Amyloid-beta precursor protein in UniProt's ProtVista visualization



Human Amyloid-beta precursor protein in UniProt's ProtVista visualization

The viewer is composed of tracks – the fundamental building blocks.



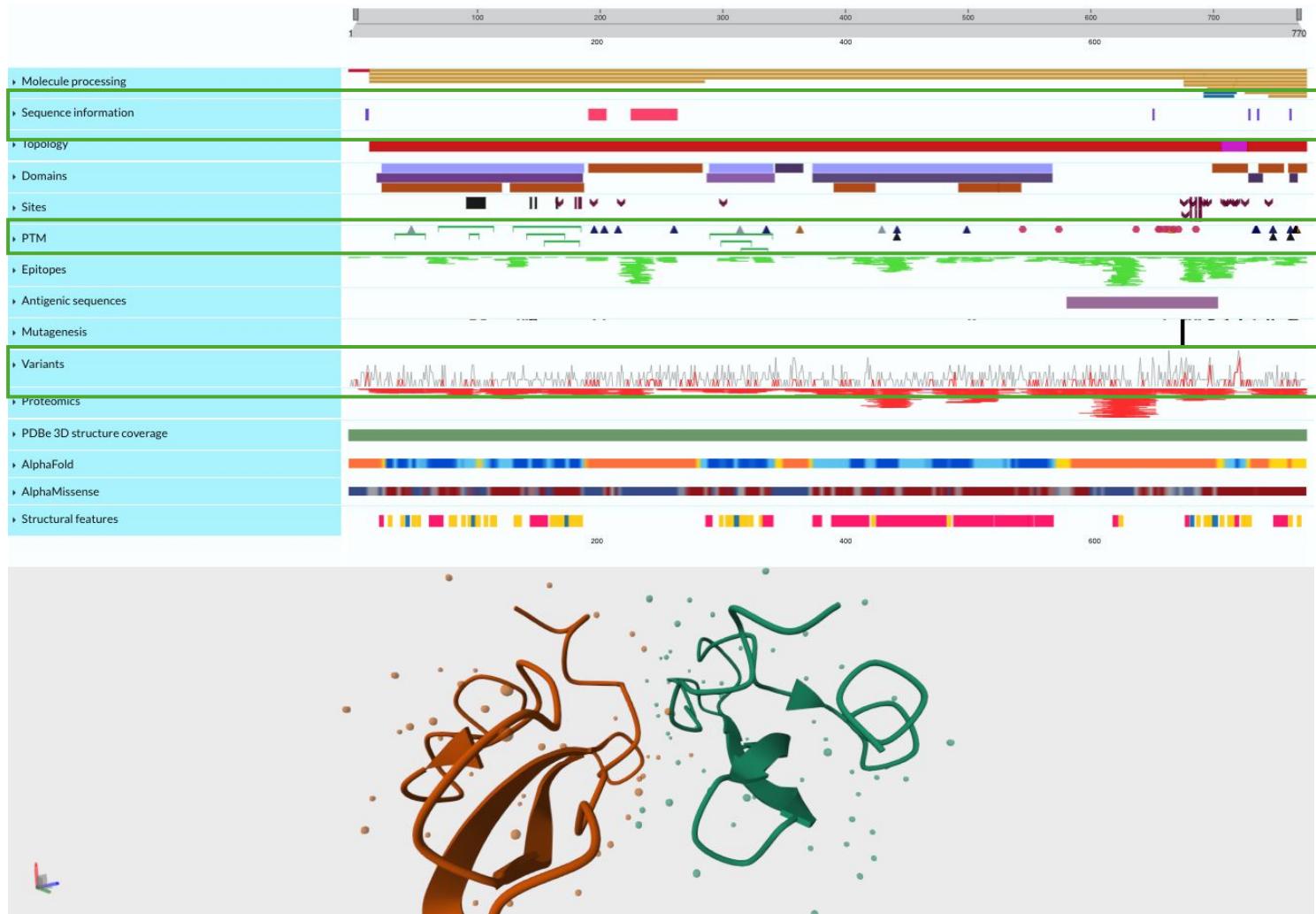


Human Amyloid-beta precursor protein in UniProt's ProtVista visualization

The viewer is composed of tracks – the fundamental building blocks.

Each track can be used individually or combined as shown here





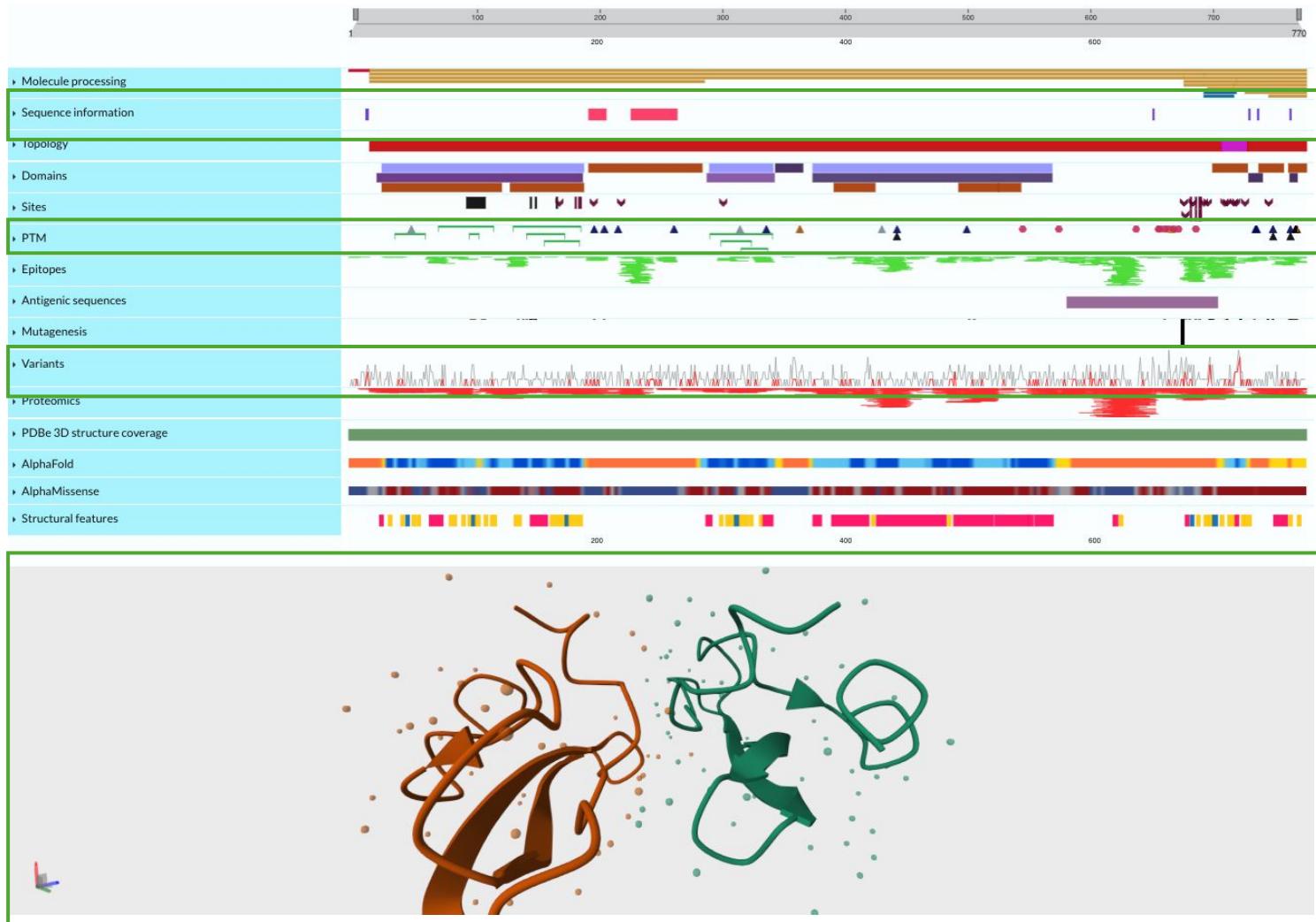
Human Amyloid-beta precursor protein in UniProt's ProtVista visualization

The viewer is composed of tracks – the fundamental building blocks.

Each track can be used individually or combined as shown here

ProtVista fetches all the data and each track is responsible for representing its own data.





Human Amyloid-beta precursor protein in UniProt's ProtVista visualization



The viewer is composed of tracks – the fundamental building blocks.

Each track can be used individually or combined as shown here

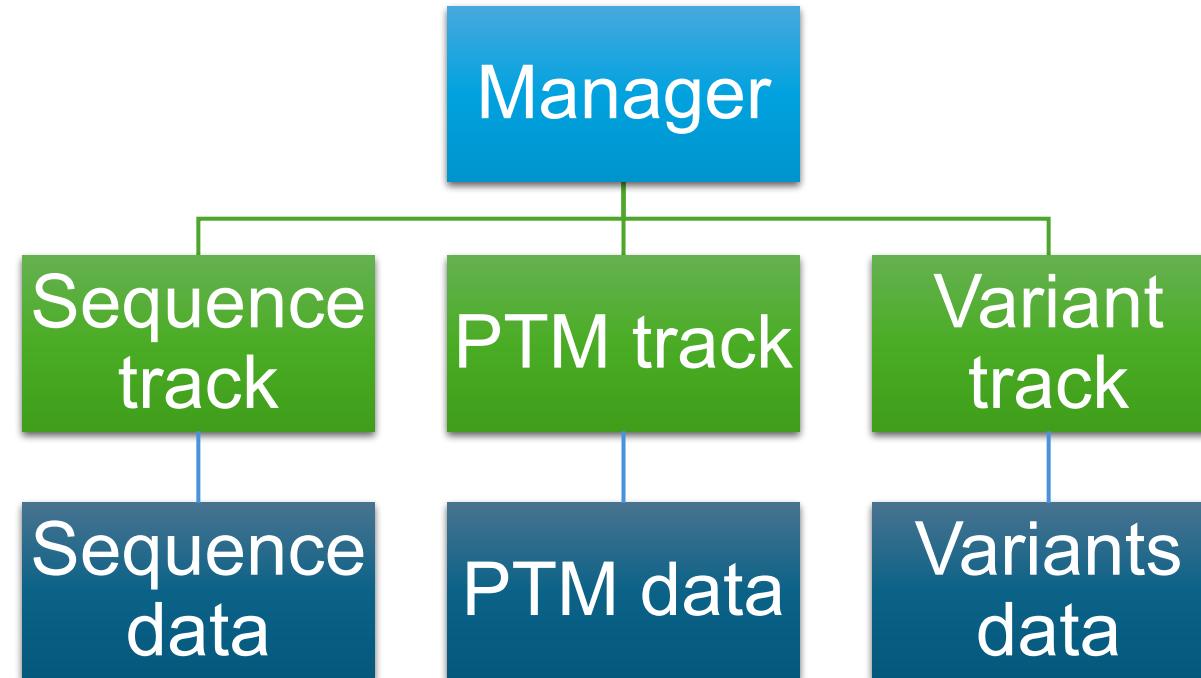
ProtVista fetches all the data and each track is responsible for representing its own data.

The structure viewer is also treated as a track that render data in 3D context



Architecture

The manager serves as a container for all the tracks, handling events from its descendants and ensuring **synchronized** behavior across tracks.



User clicks on a variant in the Variant track



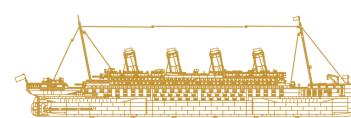
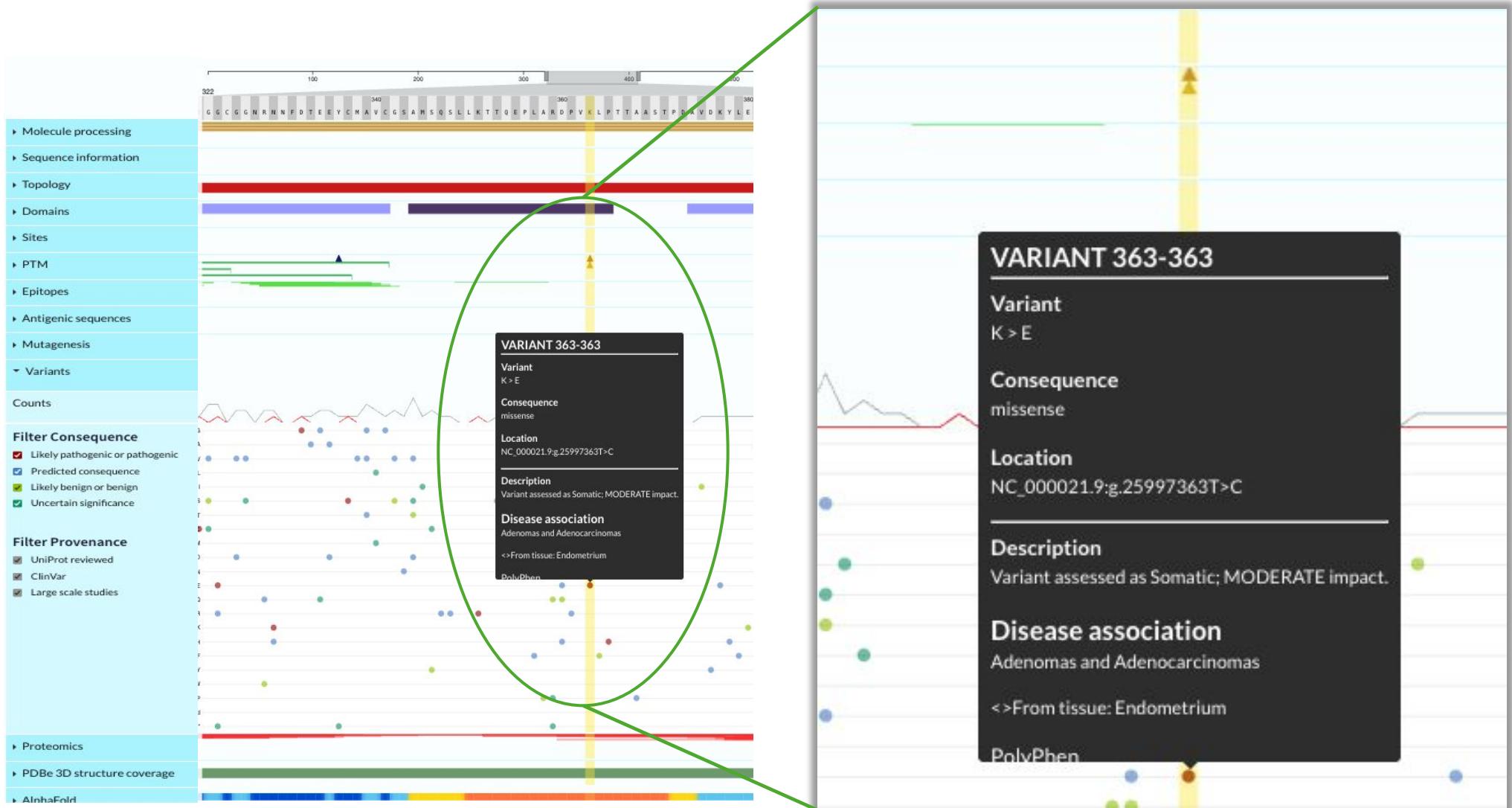
Track emits a 'highlight' event with the selected coordinates



The manager detects the event



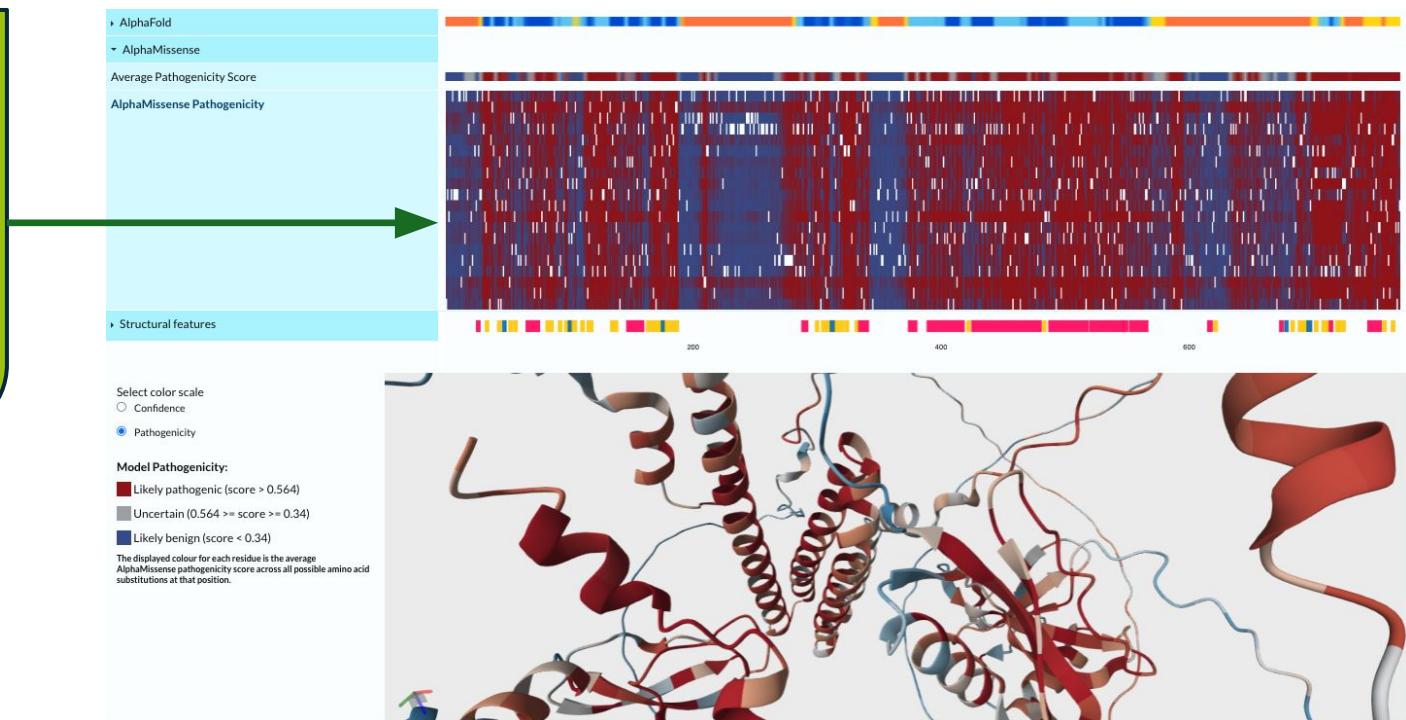
The manager sets the value to all the other tracks' highlight attribute



Under the hood

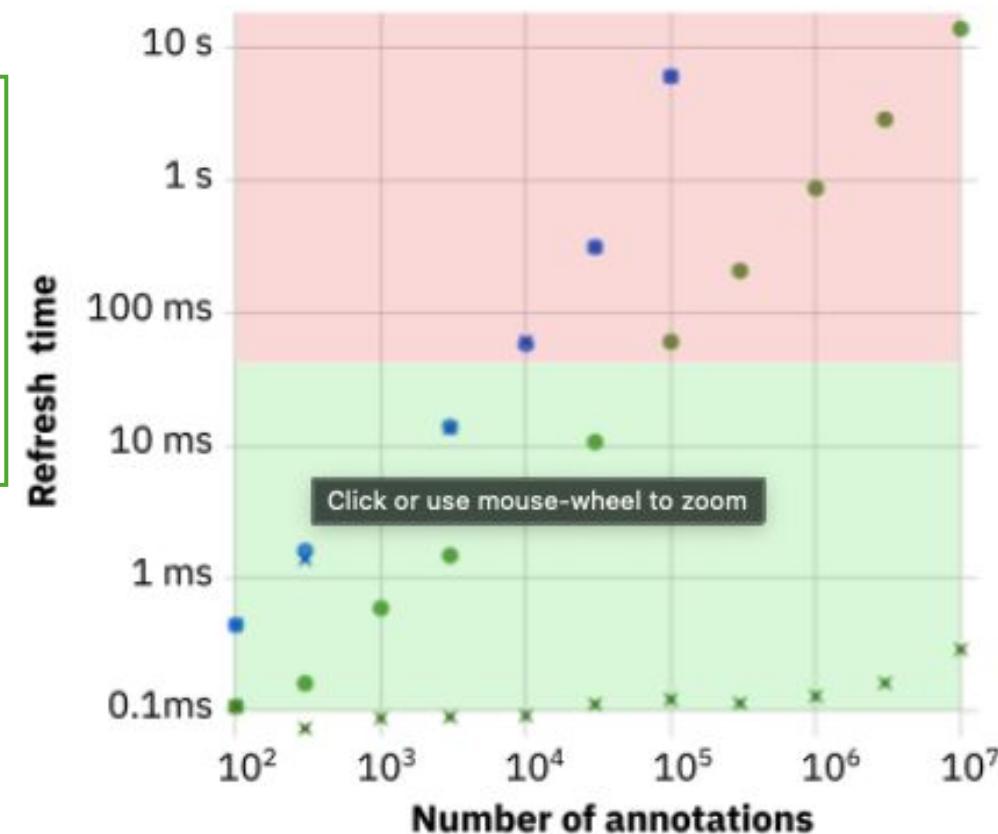
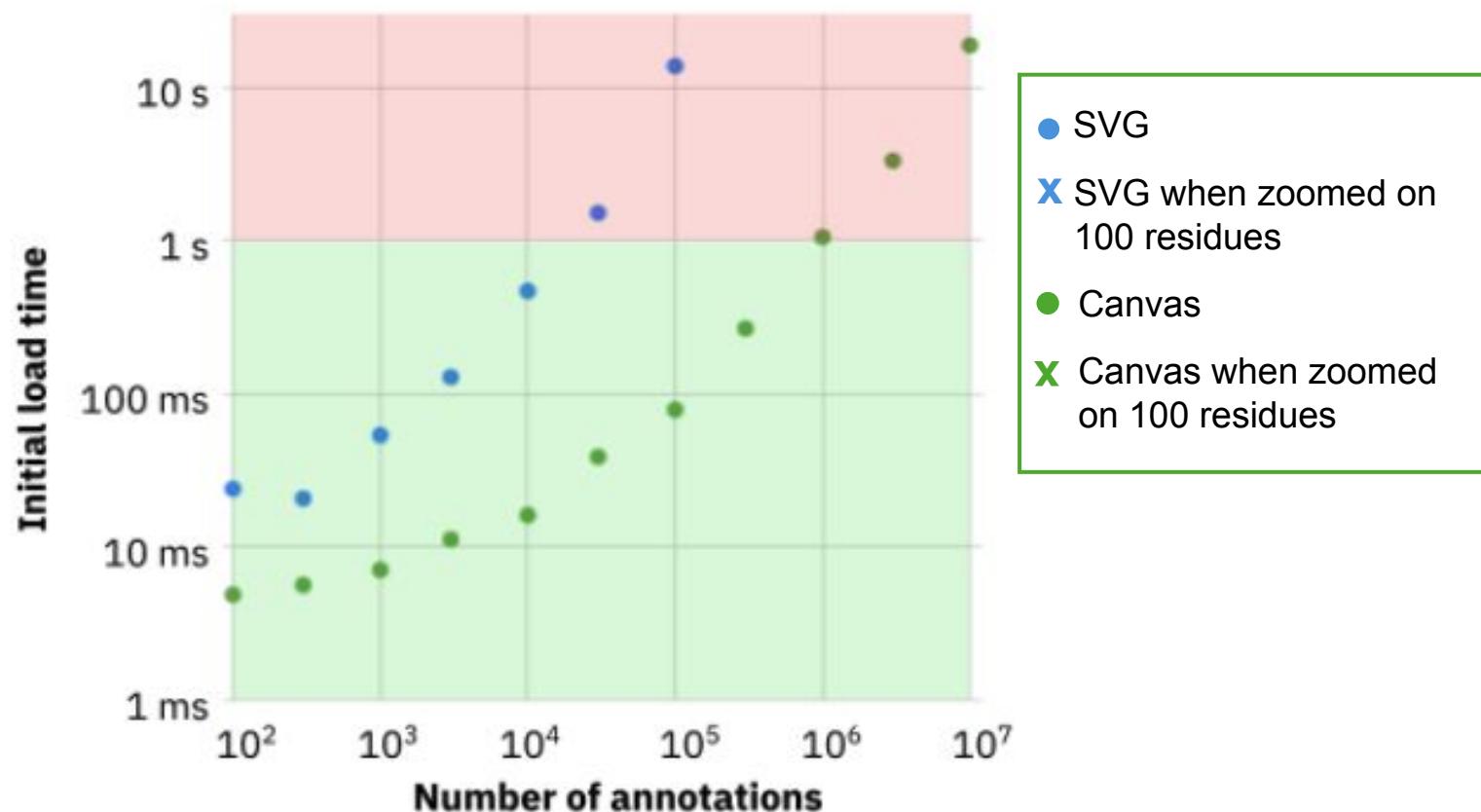
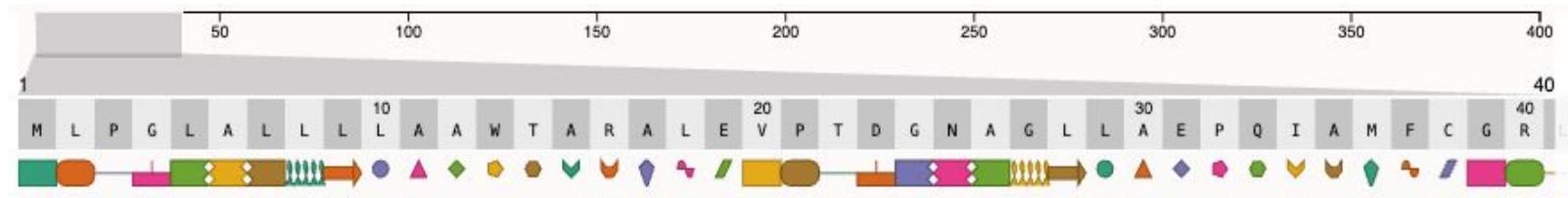
- **Lit** - for building reusable Web Components
- **D3.js** - for data-driven rendering via SVG and Canvas
- **HTML5 Canvas** – leveraged for high performance rendering in heavily used tracks
- **SVG overlay** – enables interactivity on top of Canvas for minimal performance impact

- Rendered using Canvas
- Pixels drawn directly onto the screen
- Doesn't retain all the elements in the DOM like SVG
- Reduces memory overhead



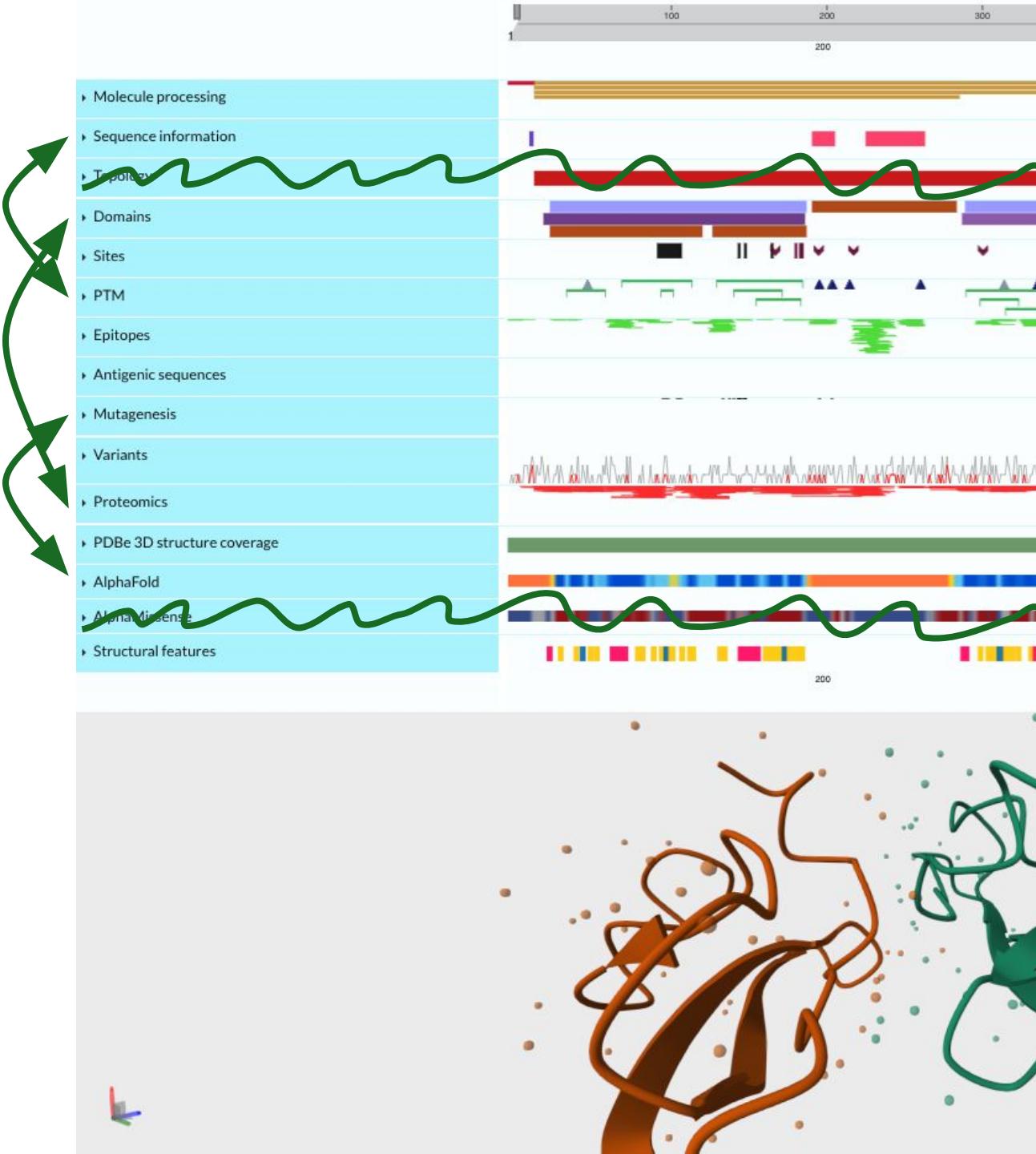
Performance optimization

Performance comparison
between SVG and Canvas
when visualizing the same
dataset



Challenges and future

- From no dependency, to minimum dependencies
 - Avoiding footguns
 - Embracing TypeScript
- Package publication best practices
 - Managing monorepos & bundles
- Performance improvements
 - Full transition to Canvas
 - Explore WebGL
- Improve developer and user experiences
- Engaging a wider community



Funding



InterPro



It is a collaboration with InterPro and PDBe and licensed under the MIT License.

UniProt is funded by joint awards by the National Institute of Health (NIH) [U24HG007822]; European Molecular Biology Laboratory (EMBL) core funds; Wellcome Trust [221320/Z/20/Z]; the UK Biotechnology and Biological Science Research Council [BB/W017970/1] and the United States National Science Foundation (NSF) [DBI-2129634]

Further ProtVista development will be supported by the Research Software Maintenance Fund, managed by the Software Sustainability Institute and funded by UKRI grant reference AH/Z000114/1



EMBL-EBI





Nightingale monorepo

<https://github.com/ebi-webcomponents/nightingale>

Nightingale Example pages

<https://ebi-webcomponents.github.io/nightingale>

ProtVista repo

<https://github.com/ebi-webcomponents/protvista-uniprot>

Any questions? 