

Tools to search, explore and interrogate your proteomics data with UniProt.

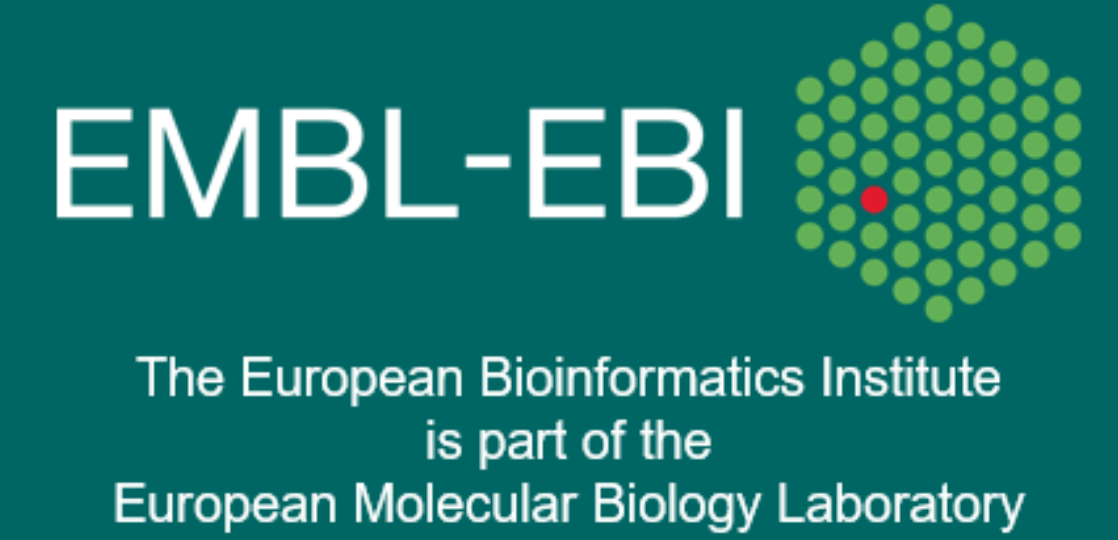


Emily Bowler-Barnett¹, Sandra Orchard¹, and the UniProt Consortium^{1, 2, 3}

¹ European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Cambridge, UK

² Swiss Institute of Bioinformatics (SIB), Geneva, Switzerland

³ Protein Information Resource (PIR), Washington DC, USA



Explore key protein features

UniProt protein feature viewer

Integrating proteomics and genomics data with the variant viewer

- Based on canonical protein sequences.
- Can view protein domains, PTM sites, structural features, proteomic peptides, experimental mutagenesis sites, and natural variants.
- Cross references to publications and mapped databases.
- Selected features are highlighted on protein 3D structure, when available.

3D protein structure viewer

PDB Entry	Method	Resolution	Chain	Positions	Links
2G57	NMR		A	19-44	PDBJ PDBSu... PDBe RCSB... PDB RCSB... PDB RCSB...
2GL7	X-ray	2.60 Å	A/D	138-686	PDBe RCSB... PDB RCSB... PDB RCSB...
2Z6H	X-ray	2.20 Å	A	138-781	PDBe RCSB... PDB RCSB... PDB RCSB...
3DIW	X-ray	2.10 Å	C/D	772-781	PDBe RCSB...

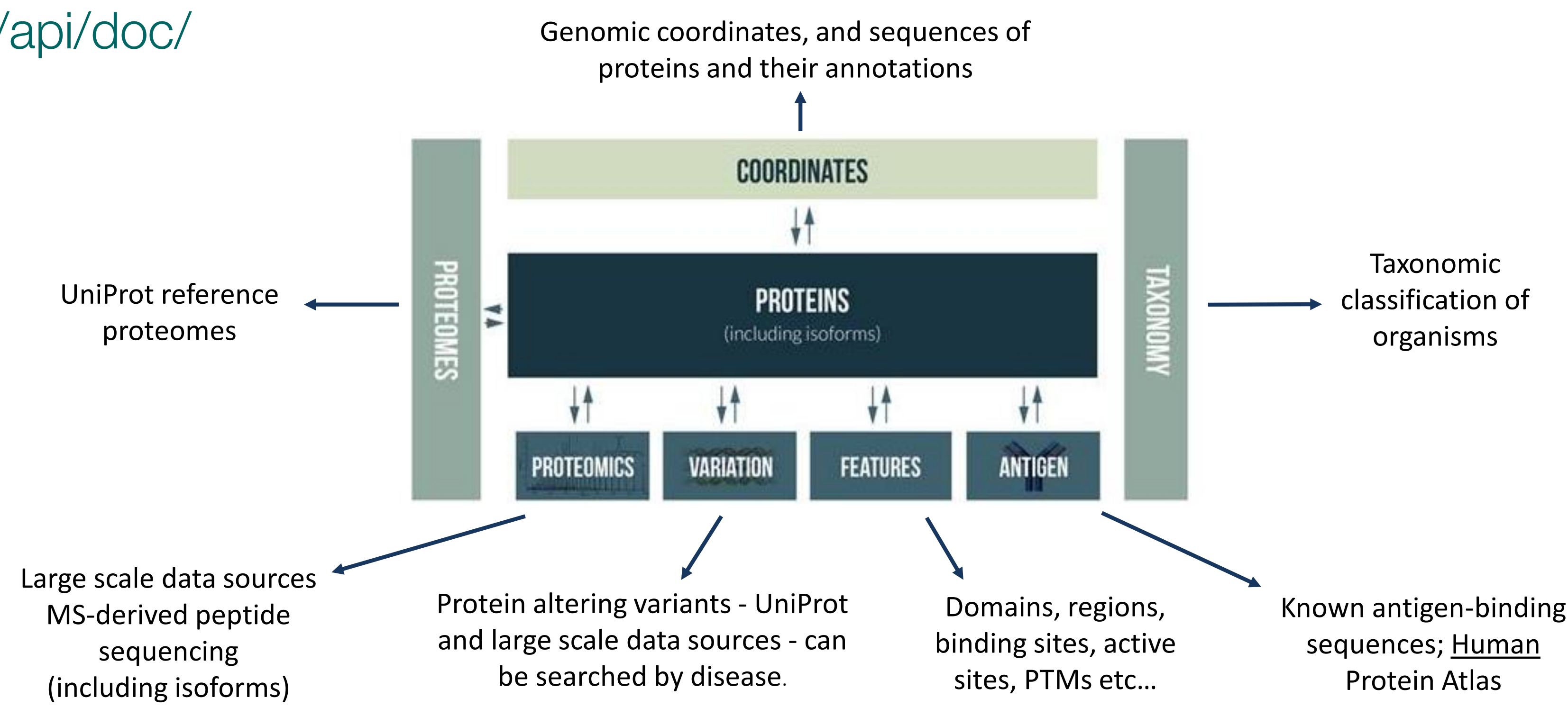
UniProt proteins API – database programmatic access.

www.ebi.ac.uk/protins/api/doc/



- Proteomics** - MS peptide sequences and positions
 - "begin": "171",
 - "end": "185"
 - ...
 - "peptide": "AAVMVHQLSKKEASR"
 - "unique": true
- Genomic coordinates** – protein genomic position
 - "chromosome": "3",
 - "start": "41224069"
 - "end": "41239339"
- Exon genomic coordinates**
 - "proteinStart": 171, (Exon start)
 - "geneStart": 41225349,
 - "proteinEnd": 185, (Exon end)
 - "geneEnd": 41225393,

- **No programming experience required!**
- Structured, multiple query search functionality.
- Example request code in Curl, Perl, Python, Ruby, Java, and R.
- Programmatic interface for accessing genomic coordinates of UniProt sequences, including isoforms – full genomic coordinates for proteins, exons (including exon sequences), and protein features.
- Downloadable formats; XML, JSON, FASTA, and GFF. Variation data also returns PEFF formatted results (HUPO Proteomics Standards Initiative).
- Variation, proteomics, and antigen services provide imported and mapped annotations from large scale data sources:
- 1000 genomes, ExAC, ClinVar, TCGA, COSMIC, TOPMed, gnomAD, PeptideAtlas, MaxQB, EPD, ProteomicsDB, and HPA.



Variation data output

```

"wildType": "P",
"polyphenPrediction": "benign",
"polyphenScore": 0.149,
"siftPrediction": "tolerated",
"siftScore": 0.12625,
"cytogeneticBand": "3p22.1",
"consequenceType": "missense",
"genomicLocation": "NC_000003.12:g.41224558C>A",
"sourceType": "large_scale_study"
    
```

Variant prediction scores and genomic location

```

"type": "VARIANT",
"description": "[LSS_COSMIC]: primary tissue(s): thyroid, large_intestine",
"alternativeSequence": "V",
"begin": "20",
"end": "20",
    
```

Variant residue change, site, and disease identifier

```

"name": "cosmic curated",
"id": "COSM5702",
...
"name": "ExAC",
"id": "rs757325337",
    
```

Variant evidence and references

