IEDB
Immune Epitope Database & Tools

IEDB, UniProt & Protein Ontology
A Collaboration

Presented by: Randi Vita, MD, Lead Ontology and Quality Manager

The IEDB Aggregates Protein Data

Many different publications describe epitopes from the “same” protein

- How to accurately and consistently capture such data?
- How to efficiently and intuitively present such data?

We employ strategies at 3 levels:

1. Curation
2. Collation
3. Collaboration

1 Curation

During curation, curators select the protein isoform that best represents the publication being curated

- May be author specified
- Must 100% contain epitope residues
- All possible accessions are accessible

Journal	Sci Transl Med	PMID:37075126
Epitope		
Epitope ID	2218297	IEDB_epitope:2218297
Chemical Type	Discontinuous peptide	
Source Name	haemagglutinin	GenPept:CAC86622.1
Source Organism	Influenza A virus (A/New Caledonia/20/1999(H1N1)) (Influenza A virus (A/New Caledonia/20/99 (H1N1)))	NCBITaxon:381512
Discontinuous Residues	W357, H368, Q370, N371, E372, Q373, G374, S375, Y377, K474, I476, G477, N478, C480, E482	

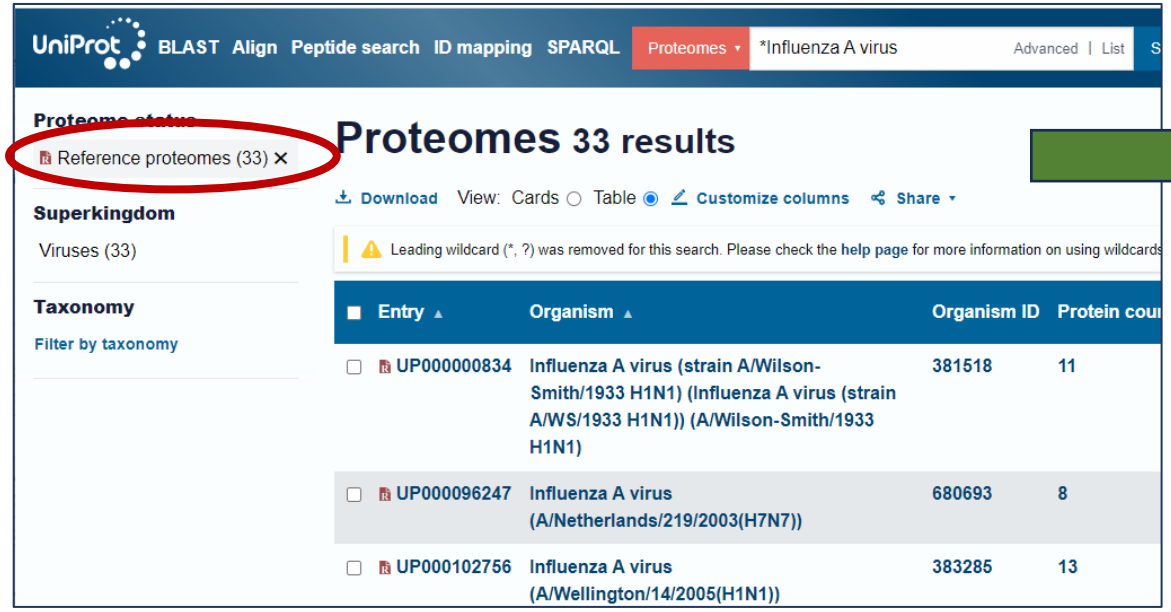
Journal	J Gen Virol	PMID:28555545
Epitope		
Epitope ID	625612	IEDB_epitope:625612
Chemical Type	Linear peptide	
Linear Sequence	SLSTASSWSY	
Source Molecule Name	hemagglutinin	GenPept:ADU02154.1
Source Organism	Influenza A virus (A/swine/Denmark/101310-1/2011(H1N1))	ONTIE:0001886
Starting Position	86	
Ending Position	95	

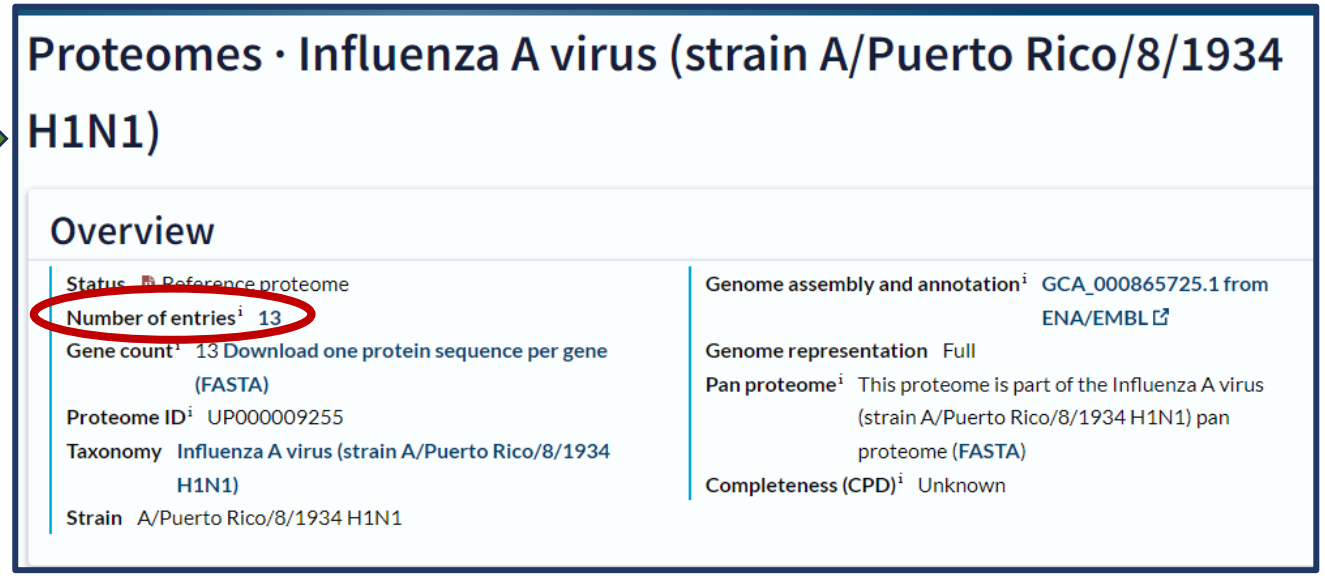
Journal	Int Immunol	PMID:10383943
Epitope		
Epitope ID	2790	IEDB_epitope:2790
Chemical Type	Linear peptide	
Linear Sequence	ALNNRFQIKGVELKS	
Source Molecule Name	hemagglutinin	GenPept:AAA87553.1
Source Organism	Influenza A virus	NCBITaxon:11320
Starting Position	511	
Ending Position	525	

2 Collation

After curation, all data belonging to a single protein (all isoforms/variants) are collated via UniProt reference proteomes:

- Sequence identity is used to identify the reference proteome that best represents the IEDB curated data
- Isoforms used in curated data are mapped selected proteome
- Updated as new data is curated and/or new proteomes become available

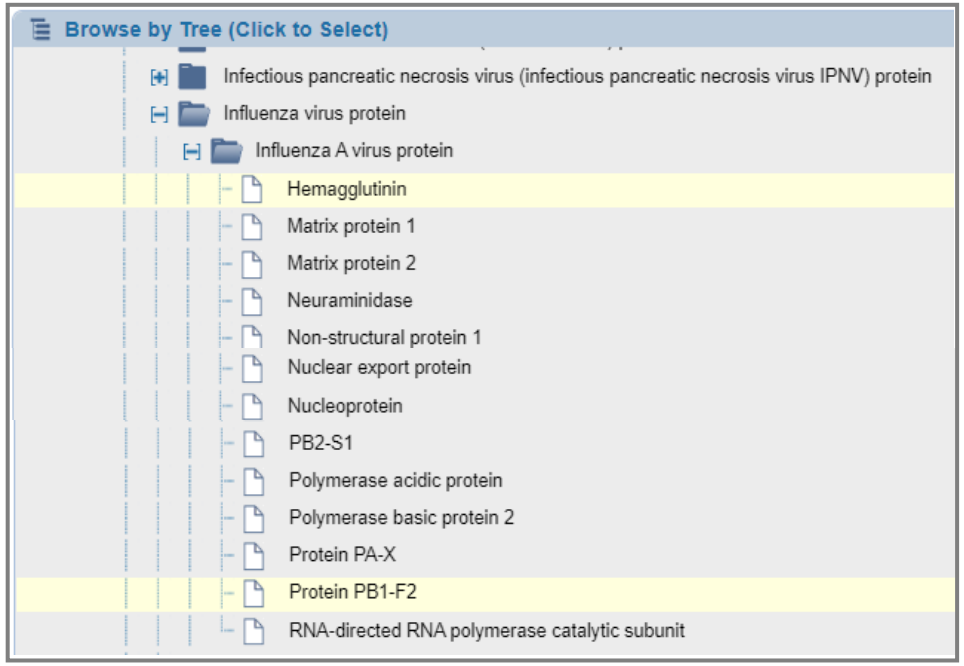


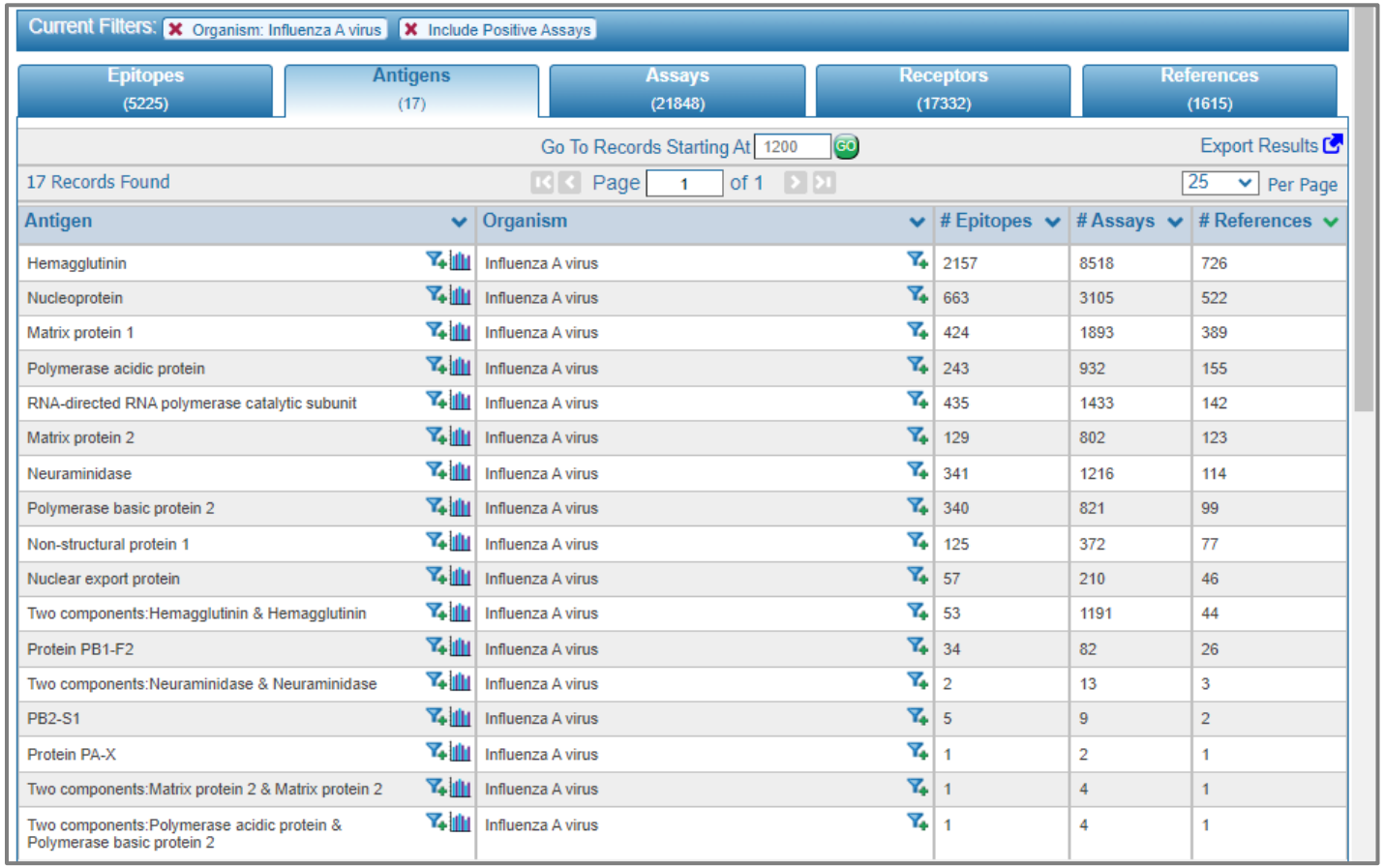


2 Collation

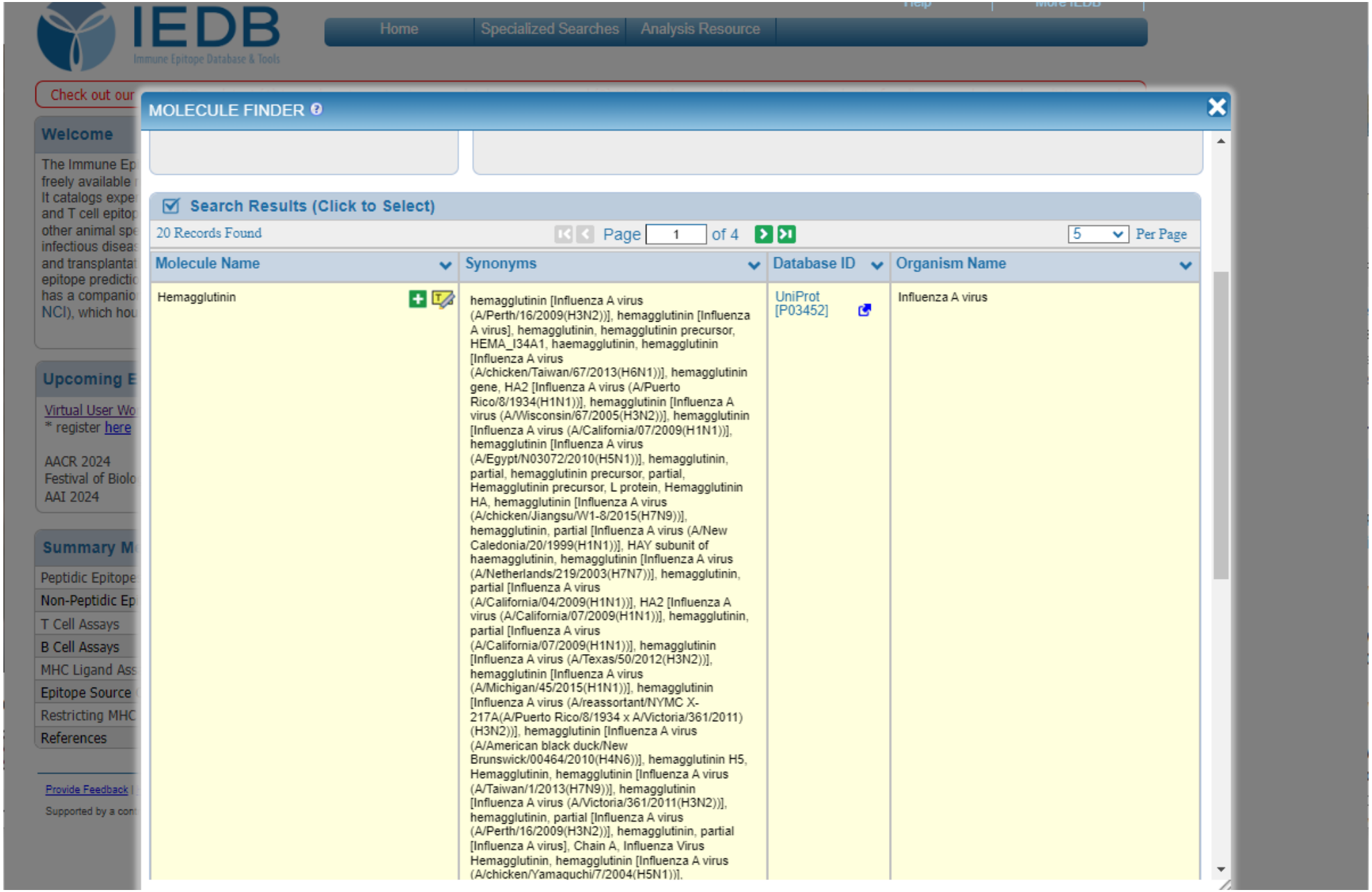
Data is searched and presented via tree view of the UniProt reference proteome

- Members of a proteome are presented as a navigable tree via Molecule Finder
- Synonyms from UniProt are searchable
- Antigens tab displays proteome members





Antigen	Organism	# Epitopes	# Assays	# References
Hemagglutinin	Influenza A virus	2157	8518	726
Nucleoprotein	Influenza A virus	663	3105	522
Matrix protein 1	Influenza A virus	424	1893	389
Polymerase acidic protein	Influenza A virus	243	932	155
RNA-directed RNA polymerase catalytic subunit	Influenza A virus	435	1433	142
Matrix protein 2	Influenza A virus	129	802	123
Neuraminidase	Influenza A virus	341	1216	114
Polymerase basic protein 2	Influenza A virus	340	821	99
Non-structural protein 1	Influenza A virus	125	372	77
Nuclear export protein	Influenza A virus	57	210	46
Two components:Hemagglutinin & Hemagglutinin	Influenza A virus	53	1191	44
Protein PB1-F2	Influenza A virus	34	82	26
Two components:Neuraminidase & Neuraminidase	Influenza A virus	2	13	3
PB2-S1	Influenza A virus	5	9	2
Protein PA-X	Influenza A virus	1	2	1
Two components:Matrix protein 2 & Matrix protein 2	Influenza A virus	1	4	1
Two components:Polymerase acidic protein & Polymerase basic protein 2	Influenza A virus	1	4	1


IEDB
Immune Epitope Database & Tools
Home
Specialized Searches
Analysis Resource
Check out our
Welcome
The Immune Ep
freely available
It catalogs expe
and T cell epitop
other animal spe
infectious disea
and transplantat
epitope predictio
has a companio
NCI), which hou
Upcoming E
Virtual User Wo
* register here
AACR 2024
Festival of Biolo
AAI 2024
Summary M
Peptidic Epitope
Non-Peptidic Ep
T Cell Assays
B Cell Assays
MHC Ligand Ass
Epitope Source
Restricting MHC
References
Provide Feedback
Supported by a con
MOLECULE FINDER
Search Results (Click to Select)
20 Records Found
Page 1 of 4
5 Per Page
Molecule Name
Synonyms
Database ID
Organism Name
Hemagglutinin
hemagglutinin [Influenza A virus (A/Perth/16/2009(H3N2))], hemagglutinin [Influenza A virus], hemagglutinin, hemagglutinin precursor, HEMA_I34A1, haemagglutinin, hemagglutinin [Influenza A virus (A/chicken/Taiwan/67/2013(H6N1))], hemagglutinin gene, HA2 [Influenza A virus (A/Puerto Rico/8/1934(H1N1))], hemagglutinin [Influenza A virus (A/Wisconsin/67/2005(H3N2))], hemagglutinin [Influenza A virus (A/California/07/2009(H1N1))], hemagglutinin [Influenza A virus (A/Egypt/N03072/2010(H5N1))], hemagglutinin, partial, hemagglutinin precursor, partial, Hemagglutinin precursor, L protein, Hemagglutinin HA, hemagglutinin [Influenza A virus (A/chicken/Jiangsu/W1-8/2015(H7N9))], hemagglutinin, partial [Influenza A virus (A/New Caledonia/20/1999(H1N1))], HAY subunit of haemagglutinin, hemagglutinin [Influenza A virus (A/Netherlands/219/2003(H7N7))], hemagglutinin, partial [Influenza A virus (A/California/04/2009(H1N1))], HA2 [Influenza A virus (A/California/07/2009(H1N1))], hemagglutinin, partial [Influenza A virus (A/California/07/2009(H1N1))], hemagglutinin [Influenza A virus (A/Texas/50/2012(H3N2))], hemagglutinin [Influenza A virus (A/Michigan/45/2015(H1N1))], hemagglutinin [Influenza A virus (A/reassortant/NYMC X-217A(A/Puerto Rico/8/1934 x A/Victoria/361/2011) (H3N2))], hemagglutinin [Influenza A virus (A/American black duck/New Brunswick/00464/2010(H4N6))], hemagglutinin H5, Hemagglutinin, hemagglutinin [Influenza A virus (A/Taiwan/1/2013(H7N9))], hemagglutinin [Influenza A virus (A/Victoria/361/2011(H3N2))], hemagglutinin, partial [Influenza A virus (A/Perth/16/2009(H3N2))], hemagglutinin, partial [Influenza A virus], Chain A, Influenza Virus Hemagglutinin, hemagglutinin [Influenza A virus (A/chicken/Yamaguchi/7/2004(H5N1))],
UniProt [P03452]
Influenza A virus

3 Collaboration

New collaboration with the **Protein Ontology (PRO)**, **UniProtKB**, and **iPTMnet** is enabling new functionality and improving IEDB data

Protein Ontology (PRO)

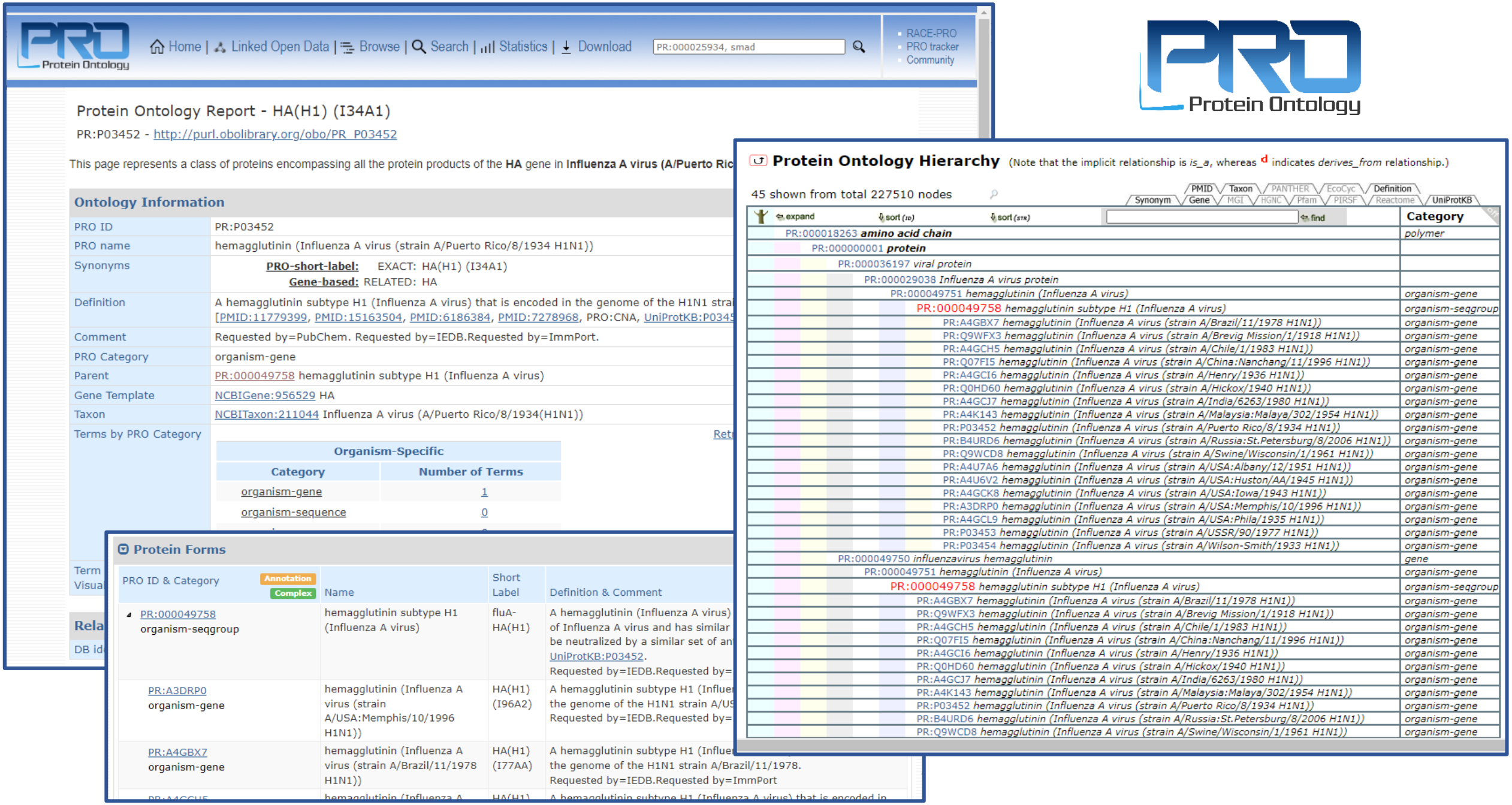
- Ontological representation of protein-related entities
- Explicit representation of proteoforms
- *proconsortium.org*

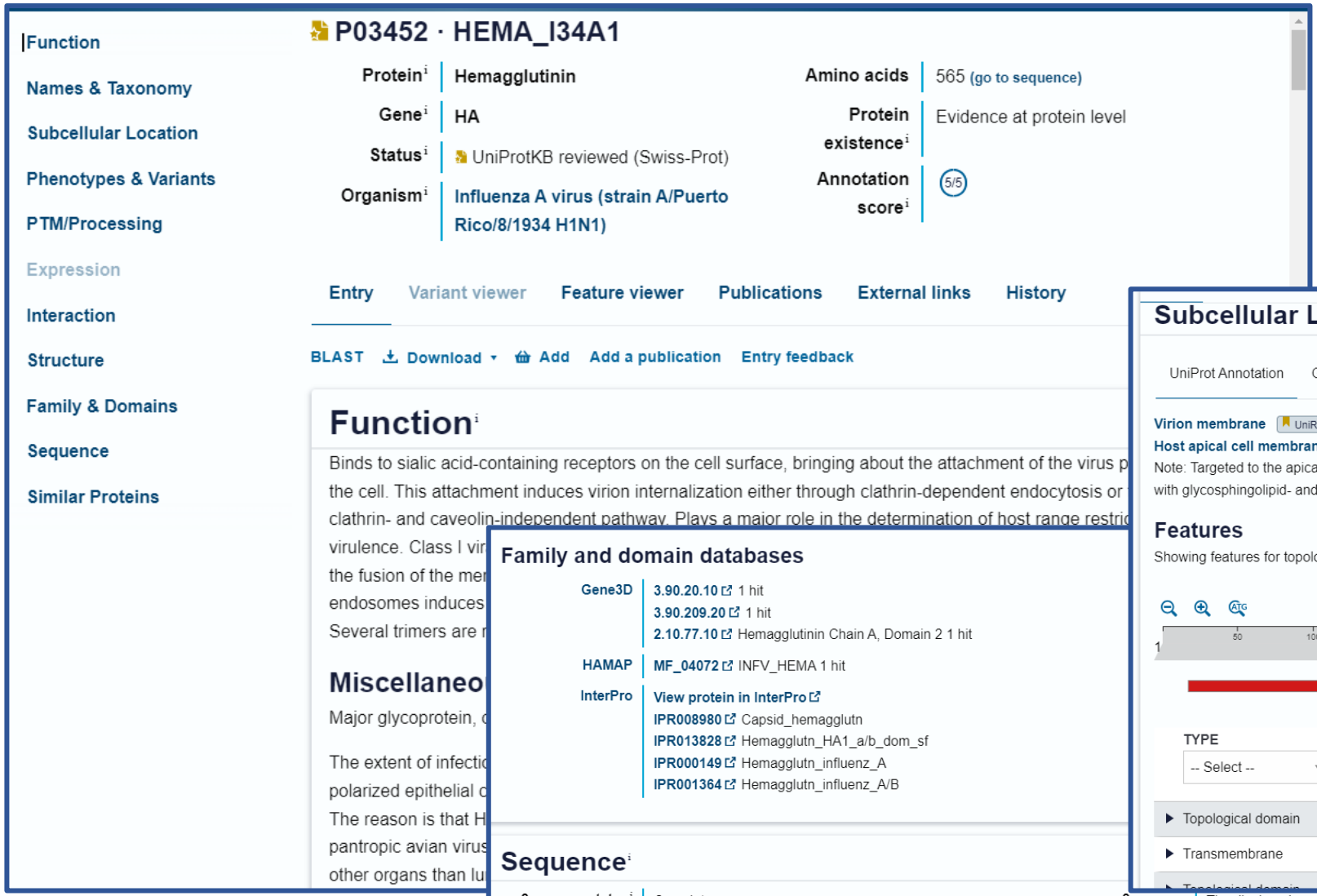
UniProtKB

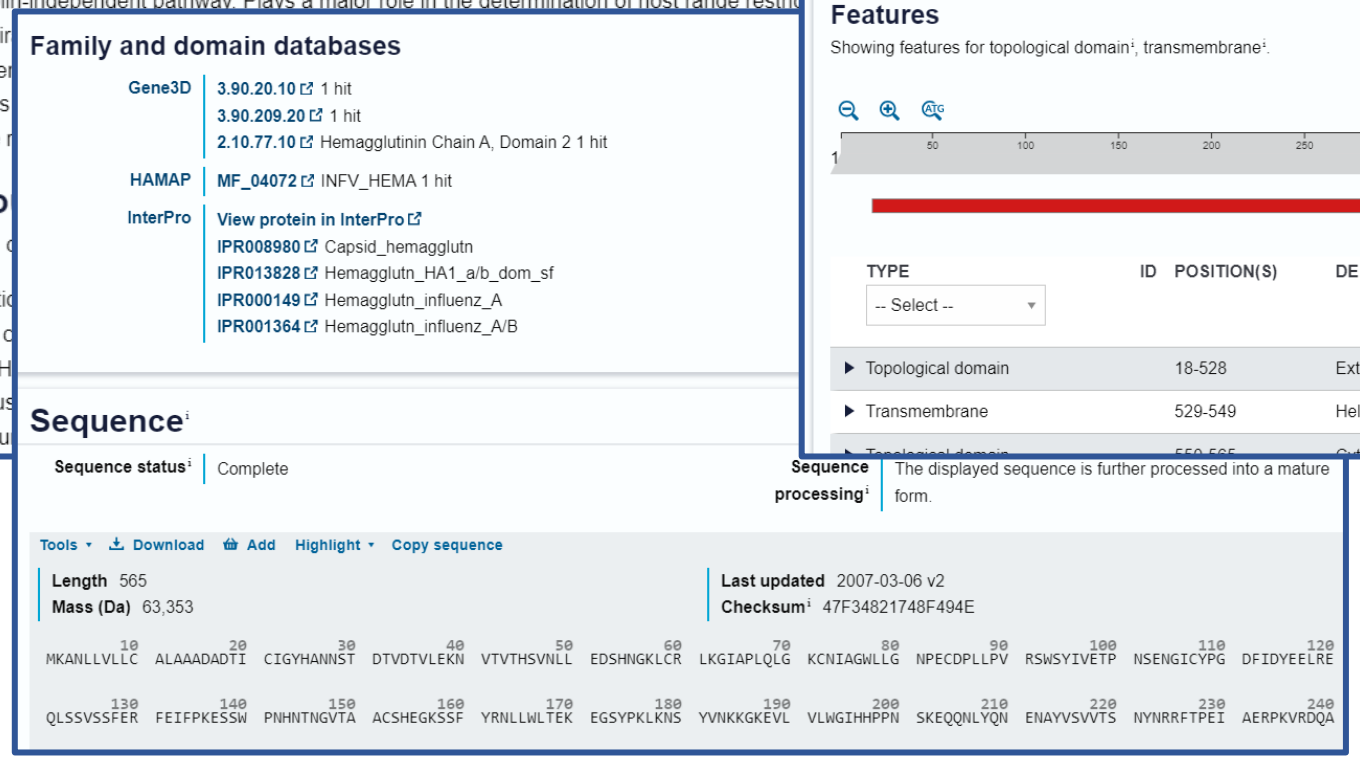
- Database of protein sequences
- Rich functional annotation of proteins
- *uniprot.org*

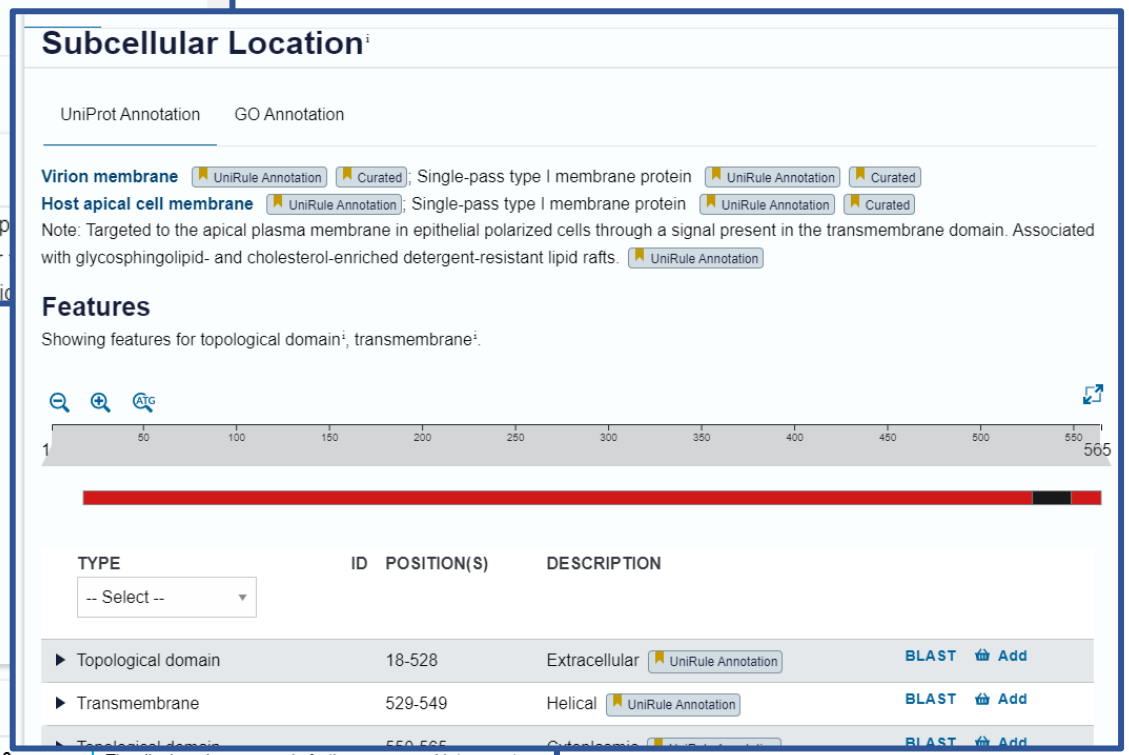
iPTMnet

- Library of post-translational modifications (PTMs)
- Integrates protein PTMs in a systems biology context
- *research.bioinformatics.udel.edu/iptmnet*


Protein Ontology
Home | Linked Open Data | Browse | Search | Statistics | Download
PR:000025934, smad
RACE-PRO
PRO tracker
Community
Protein Ontology Report - HA(H1) (I34A1)
PR:P03452 - http://purl.obolibrary.org/obo/PR_P03452
This page represents a class of proteins encompassing all the protein products of the HA gene in Influenza A virus (A/Puerto Ric
Ontology Information
PRO ID
PR:P03452
PRO name
hemagglutinin (Influenza A virus (strain A/Puerto Rico/8/1934 H1N1))
Synonyms
PRO-short-label: EXACT: HA(H1) (I34A1)
Gene-based: RELATED: HA
Definition
A hemagglutinin subtype H1 (Influenza A virus) that is encoded in the genome of the H1N1 strai
[PMID:11779399, PMID:15163504, PMID:6186384, PMID:7278968, PRO:CNA, UniProtKB:P0345
Comment
Requested by=PubChem. Requested by=IEDB.Requested by=ImmPort.
PRO Category
organism-gene
Parent
PR:000049758 hemagglutinin subtype H1 (Influenza A virus)
Gene Template
NCBIGene:956529 HA
Taxon
NCBITaxon:211044 Influenza A virus (A/Puerto Rico/8/1934(H1N1))
Terms by PRO Category
Organism-Specific
Category
Number of Terms
organism-gene
1
organism-sequence
0
Protein Forms
PRO ID & Category
Annotation
Complex
Name
Short Label
Definition & Comment
PR:000049758
organism-seqgroup
hemagglutinin subtype H1 (Influenza A virus)
fluA-HA(H1)
A hemagglutinin (Influenza A virus) of Influenza A virus and has similar be neutralized by a similar set of an UniProtKB:P03452.
Requested by=IEDB.Requested by=
PR:A3DRP0
organism-gene
hemagglutinin (Influenza A virus (strain A/USA:Memphis/10/1996 H1N1))
HA(H1) (I96A2)
A hemagglutinin subtype H1 (Influe the genome of the H1N1 strain A/US
Requested by=IEDB.Requested by=
PR:A4GBX7
organism-gene
hemagglutinin (Influenza A virus (strain A/Brazil/11/1978 H1N1))
HA(H1) (I77AA)
A hemagglutinin subtype H1 (Influe the genome of the H1N1 strain A/Brazil/11/1978.
Requested by=IEDB.Requested by=ImmPort
Protein Ontology Hierarchy (Note that the implicit relationship is is_a, whereas d indicates derives_from relationship.)
45 shown from total 227510 nodes
PMID Taxon PANTHER EcoCyc Definition
Synonym Gene MGI HGNC Pfam PIRSF Reactome UniProtKB
expand
sort (ID)
sort (STR)
find
Category
PR:000018263 amino acid chain
polymer
PR:000000001 protein
PR:000036197 viral protein
PR:000029038 Influenza A virus protein
PR:000049751 hemagglutinin (Influenza A virus)
organism-gene
PR:000049758 hemagglutinin subtype H1 (Influenza A virus)
organism-seqgroup
PR:A4GBX7 hemagglutinin (Influenza A virus (strain A/Brazil/11/1978 H1N1))
PR:Q9WFX3 hemagglutinin (Influenza A virus (strain A/Brevig Mission/1/1918 H1N1))
PR:A4GCH5 hemagglutinin (Influenza A virus (strain A/Chile/1/1983 H1N1))
PR:Q07FI5 hemagglutinin (Influenza A virus (strain A/China:Nanchang/11/1996 H1N1))
PR:A4GCI6 hemagglutinin (Influenza A virus (strain A/Henry/1936 H1N1))
PR:Q0HD60 hemagglutinin (Influenza A virus (strain A/Hickox/1940 H1N1))
PR:A4GCJ7 hemagglutinin (Influenza A virus (strain A/India/6263/1980 H1N1))
PR:A4K143 hemagglutinin (Influenza A virus (strain A/Malaysia:Malaya/302/1954 H1N1))
PR:P03452 hemagglutinin (Influenza A virus (strain A/Puerto Rico/8/1934 H1N1))
PR:B4URD6 hemagglutinin (Influenza A virus (strain A/Russia:St.Petersburg/8/2006 H1N1))
PR:Q9WCD8 hemagglutinin (Influenza A virus (strain A/Swine/Wisconsin/1/1961 H1N1))
PR:A4U7A6 hemagglutinin (Influenza A virus (strain A/USA:Albany/12/1951 H1N1))
PR:A4U6V2 hemagglutinin (Influenza A virus (strain A/USA:Huston/AA/1945 H1N1))
PR:A4GCK8 hemagglutinin (Influenza A virus (strain A/USA:Iowa/1943 H1N1))
PR:A3DRP0 hemagglutinin (Influenza A virus (strain A/USA:Memphis/10/1996 H1N1))
PR:A4GCL9 hemagglutinin (Influenza A virus (strain A/USA:Phila/1935 H1N1))
PR:P03453 hemagglutinin (Influenza A virus (strain A/USSR/90/1977 H1N1))
PR:P03454 hemagglutinin (Influenza A virus (strain A/Wilson-Smith/1933 H1N1))
PR:000049750 influenzavirus hemagglutinin
gene
PR:000049751 hemagglutinin (Influenza A virus)
organism-gene
PR:000049758 hemagglutinin subtype H1 (Influenza A virus)
organism-seqgroup
PR:A4GBX7 hemagglutinin (Influenza A virus (strain A/Brazil/11/1978 H1N1))
PR:Q9WFX3 hemagglutinin (Influenza A virus (strain A/Brevig Mission/1/1918 H1N1))
PR:A4GCH5 hemagglutinin (Influenza A virus (strain A/Chile/1/1983 H1N1))
PR:Q07FI5 hemagglutinin (Influenza A virus (strain A/China:Nanchang/11/1996 H1N1))
PR:A4GCI6 hemagglutinin (Influenza A virus (strain A/Henry/1936 H1N1))
PR:Q0HD60 hemagglutinin (Influenza A virus (strain A/Hickox/1940 H1N1))
PR:A4GCJ7 hemagglutinin (Influenza A virus (strain A/India/6263/1980 H1N1))
PR:A4K143 hemagglutinin (Influenza A virus (strain A/Malaysia:Malaya/302/1954 H1N1))
PR:P03452 hemagglutinin (Influenza A virus (strain A/Puerto Rico/8/1934 H1N1))
PR:B4URD6 hemagglutinin (Influenza A virus (strain A/Russia:St.Petersburg/8/2006 H1N1))
PR:Q9WCD8 hemagglutinin (Influenza A virus (strain A/Swine/Wisconsin/1/1961 H1N1))
organism-gene


Function
Names & Taxonomy
Subcellular Location
Phenotypes & Variants
PTM/Processing
Expression
Interaction
Structure
Family & Domains
Sequence
Similar Proteins
P03452 · HEMA_I34A1
Protein Hemagglutinin
Gene HA
Status UniProtKB reviewed (Swiss-Prot)
Organism Influenza A virus (strain A/Puerto Rico/8/1934 H1N1)
Amino acids 565 (go to sequence)
Protein existence Evidence at protein level
Annotation score 5/5
Entry
Variant viewer
Feature viewer
Publications
External links
History
BLAST
Download
Add
Add a publication
Entry feedback
Function
Binds to sialic acid-containing receptors on the cell surface, bringing about the attachment of the virus p
the cell. This attachment induces virion internalization either through clathrin-dependent endocytosis or
clathrin- and caveolin-independent pathway. Plays a major role in the determination of host range restric
virulence. Class I vir
the fusion of the mer
endosomes induces
Several trimers are r
Miscellaneo
Major glycoprotein, c
The extent of infectio
polarized epithelial c
The reason is that H
pantropic avian virus
other organs than lu


Family and domain databases
Gene3D 3.90.20.10 1 hit
3.90.209.20 1 hit
2.10.77.10 Hemagglutinin Chain A, Domain 2 1 hit
HAMAP MF_04072 INFV_HEMA 1 hit
InterPro View protein in InterPro
IPR008980 Capsid_hemagglutn
IPR013828 Hemagglutn_HA1_a/b_dom_sf
IPR000149 Hemagglutn_influenz_A
IPR001364 Hemagglutn_influenz_A/B
Sequence
Sequence status Complete
Sequence processing The displayed sequence is further processed into a mature form.
Tools
Download
Add
Highlight
Copy sequence
Length 565
Mass (Da) 63,353
Last updated 2007-03-06 v2
Checksum 47F34821748F494E
10 20 30 40 50 60 70 80 90 100 110 120
MKANLLVLLC ALAAADADTI CIGYHANNST DTVDTVLEKN VTVTHSVNLL EDSHNGKLCR LKGIAPLQLG KCNIAGWLLG NPECDPLLPV RSWSYIVETP NSENGICYPG DFIDYEELRE
130 140 150 160 170 180 190 200 210 220 230 240
QLSSVSSFER FEIFPKESSW PNHNTNGVTA ACSHEGKSSF YRNLLWLTEK EGSYPKLKNS YVNKKGKEVL VLWGIHHPPN SKEQQNLYQN ENAYVSVVTS NYNRRFTPEI AERPKVRDQA


Subcellular Location
UniProt Annotation
GO Annotation
Virion membrane UniRule Annotation Curated; Single-pass type I membrane protein UniRule Annotation Curated
Host apical cell membrane UniRule Annotation; Single-pass type I membrane protein UniRule Annotation Curated
Note: Targeted to the apical plasma membrane in epithelial polarized cells through a signal present in the transmembrane domain. Associated with glycosphingolipid- and cholesterol-enriched detergent-resistant lipid rafts. UniRule Annotation
Features
Showing features for topological domain, transmembrane.
1 50 100 150 200 250 300 350 400 450 500 550 565
TYPE ID POSITION(S) DESCRIPTION
-- Select --
Topological domain 18-528 Extracellular UniRule Annotation BLAST Add
Transmembrane 529-549 Helical UniRule Annotation BLAST Add

iPTMnet

iPTMnet Home Browse Statistics Project Info Help Result All Search

Display

Protein Information

Interactive Sequence View

O60814 (H2BC12) as Substrate

Proteoforms

PTM sites affected in variants

Back to top

Cytoscape View

Clear

Submit

iPTMnet Report for O60814 (H2BC12)

Protein Information

Cytoscape View

UniProt AC / UniProt ID	O60814 / H2B1K_HUMAN
Protein Name	Histone H2B type 1-K
Gene Name	Name: H2BC12 Synonyms:H2BFT;
Organism	Homo sapiens (Human)
PRO ID	PR:O60814
PRO Name	histone H2B type 1-K (human)
Definition	A histone H2B type 1-K that is encoded in the genome of human.
Short Label	hH2BC12
Category	organism-gene

Interactive Sequence View

Show PRO entries Modification site affected in variant

• Number of sequence: 14 • Alignment length: 126 • Scale: "–" ≈ 2 amino acids

iPTM:O60814 hH2BC12

Site	PTM Type	PTM Enzyme	Score	Source	PMID
All	All		All	All	
	Acetylation		★☆☆☆	PRO	
K6	Acetylation		★★☆☆	PRO UniProt	18552846, 16283522, 16627869
K12	Acetylation		★★☆☆	PhosphoSitePlus UniProt	16627869
K13	Acetylation		★★☆☆	PRO UniProt	18552846, 16627869, 20693536, ...
S15	Phosphorylation		★★☆☆	PRO UniProt	12757711
K16	Acetylation		★★☆☆	PRO PhosphoSitePlus UniProt	16283522, 19883641, 16627869
K17	Acetylation		★★☆☆	PRO UniProt	20693536, 16627869
K21	Acetylation		★★☆☆	PRO UniProt	18552846, 16283522, 16627869
K24	Acetylation		★☆☆☆	PhosphoSitePlus	
K35	Ubiquitination		★☆☆☆	UniProt	21726816

PTM sites affected in variants

Sear

Site	Variant	Source	PMID	Disease [Sample source]
All	All	All		All
K12	*	Biomuta		DOID:1324 / lung cancer [cosmic, icgc, tcga]
K13	R13	Biomuta		DOID:11054 / urinary bladder cancer [tcga]
S15	W15	Biomuta		DOID:11054 / urinary bladder cancer [tcga]
S15	L15	Biomuta		DOID:9256 / colorectal cancer [tcga]
K16	N16	Biomuta		DOID:1324 / lung cancer [cosmic]
K21	E21	Biomuta		DOID:9256 / colorectal cancer [tcga]

Proteoforms

Search:

PRO ID (Short Label)	Sites	PTM Enzyme	Source	PMID
All	All		All	
PR:000045079 (hHIST1H2BK/InitMet-/Ac:1)	acK13		PRO	16283522, 18552846, 20693536
PR:000045080 (hHIST1H2BK/InitMet-/Phos:1)	pS15		PRO	12757711
PR:000045081 (hHIST1H2BK/InitMet-/Ac:2)	acK21		PRO	16283522, 18552846
PR:000045083 (hHIST1H2BK/InitMet-/Ubiq:1)	ubK121		PRO	16307923, 16713563, 18400894, ...
PR:000045084 (hHIST1H2BK/InitMet-/Ac:3)	acK17		PRO	20693536
PR:000045085 (hHIST1H2BK/InitMet-/Ac:4)	acK6		PRO	18552846
PR:000045087	acK47		PRO	20693536

Collaboration-Functionality

New Antigen Summary pages added to IEDB for each protein antigen/ reference proteome member

Summary of IEDB data

- Isoform information
- Epitopes and assays
- Assay contexts such as host and disease

ANTIGEN SUMMARY

Hemagglutinin from Influenza A virus (Antigen ID UNIPROT:P03452).
Synonyms: hemagglutinin precursor, HEMA_I34A1, L protein, Hemagglutinin, Chain A, hemagglutinin HA, hemagglutinin HA2, haemagglutinin, hemagglutinin gene, Hemagglutinin HA, hemagglutinin, partial, Influenza Virus Hemagglutinin, hemagglutinin HA1, hemagglutinin subtype H5, Hemagglutinin precursor, HAY subunit of haemagglutinin, Truncated hemagglutinin, HA2, hemagglutinin H5, HA1, HA.
This antigen has 3216 epitopes and has been studied for immune reactivity in 748 publication(s), tested in 4616 T cell assays, 5541 B cell assays and 1741 MHC ligand assays and has 128 3D structure assays.
Assays in which this antigen was tested are distributed according to host: Homo sapiens (human) (6825), Mus musculus (mouse) (2941), Oryctolagus cuniculus (rabbit) (321), Vicugna pacos (alpaca) (241), Macaca mulatta (rhesus macaque) (185), Gallus gallus (chicken) (121), Sus scrofa (pig) (82), Lama glama (llama) (49), Macaca fascicularis (crab eating macaque) (39), Mustela putorius (European polecat) (24), Bos taurus (bovine) (13), Pan troglodytes (chimpanzee) (10), Anas platyrhynchos (duck) (4), Capra hircus (domestic goat) (3), Rattus norvegicus (brown rat) (2), Ovis aries (domestic sheep) (1).
Assays in which this antigen was tested are distributed according to disease state of the host: influenza (925), healthy (454), collagen-induced arthritis (CIA) (6), multiple sclerosis (6), rheumatoid arthritis (6), type 1 diabetes mellitus (3), prediabetes syndrome (1).

Collaboration-Functionality

New Antigen Summary pages added to IEDB for each protein antigen/ reference proteome member

New imported annotations

Gene: UL83
Counteracts the host antiviral immune response when activated and phosphorylated, by preventing host IRF3 from entering the nucleus. Inhibits also the type I interferon production by inactivating the enzymatic activity of DNA sensor CGAS without affecting STING1. Participates in the transactivation of viral major immediate-early genes by the recruitment of host IFI16 to the promoters of these genes.

EXTERNAL RESOURCES	
Resource	Link
PRO (multiple possible)	Protein Ontology terms
UniProt	Reference protein
IUIS	Allergy information
IPTMnet	View possible post-translational modifications
Immunome Browser	View Immunome Browser
IEDB tools	Predict...

Collaboration-Data Improvements

Protein Ontology (PRO), UniProtKB, and iPTMnet identify issues with IEDB data as they each analyze our data in their view

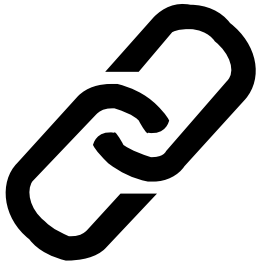
Errors fixed

- Obsoleted sources
- Organism assignment
- PTM positions and residues

New validation added

- Enforce same species
- Enforce valid PTM fields

Future Plans



New Links

- Protein Ontology (PRO) – proteoforms
- UniProtKB
- iPTMnet – modifications, functionality

Calculated Annotations

- Cancer antigen, autoantigen, allergen
- Weight based evidence

Expand Antigen Summary page coverage

- Nonpeptidics (ChEBI)