



UniProt:

The Universal

Protein Knowledgebase

About UniProt

- The Universal Protein Resource (UniProt) is the central hub for the collection of functional information on proteins with accurate, consistent, and rich annotation.
- Captures core data mandatory for each UniProt entry: **amino acid sequence, protein name or description, taxonomic data, citation information and as much annotation information is added.**
- The database contains **over 60 million sequences**, of which over half a million sequences have been curated by experts who critically review experimental and predicted data for each protein.
- UniProt is a comprehensive resource for protein sequence and annotation data. The UniProt databases are the [UniProt Knowledgebase \(UniProtKB\)](#), the [UniProt Reference Clusters \(UniRef\)](#), and the [UniProt Archive \(UniParc\)](#).

The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

UniProtKB
UniProt Knowledgebase

Swiss-Prot (553,655)
Manually annotated and reviewed.

TrEMBL (77,483,538)
Automatically annotated and not reviewed.

UniRef
Sequence clusters

UniParc
Sequence archive

Proteomes

Supporting data

Literature citations

Cross-ref. databases

Taxonomy

Diseases

Subcellular locations

Keywords

News

Forthcoming changes
Planned changes for UniProt

UniProt release 2017_02
Freshwater fish see red | Cross-references to Araport, TAIR and IMGT/Gene-DB | Removal of sequence similarity annotations for domains

UniProt release 2017_01
Sheep in wolves' clothing | Change of the UniRef FASTA header

News archive

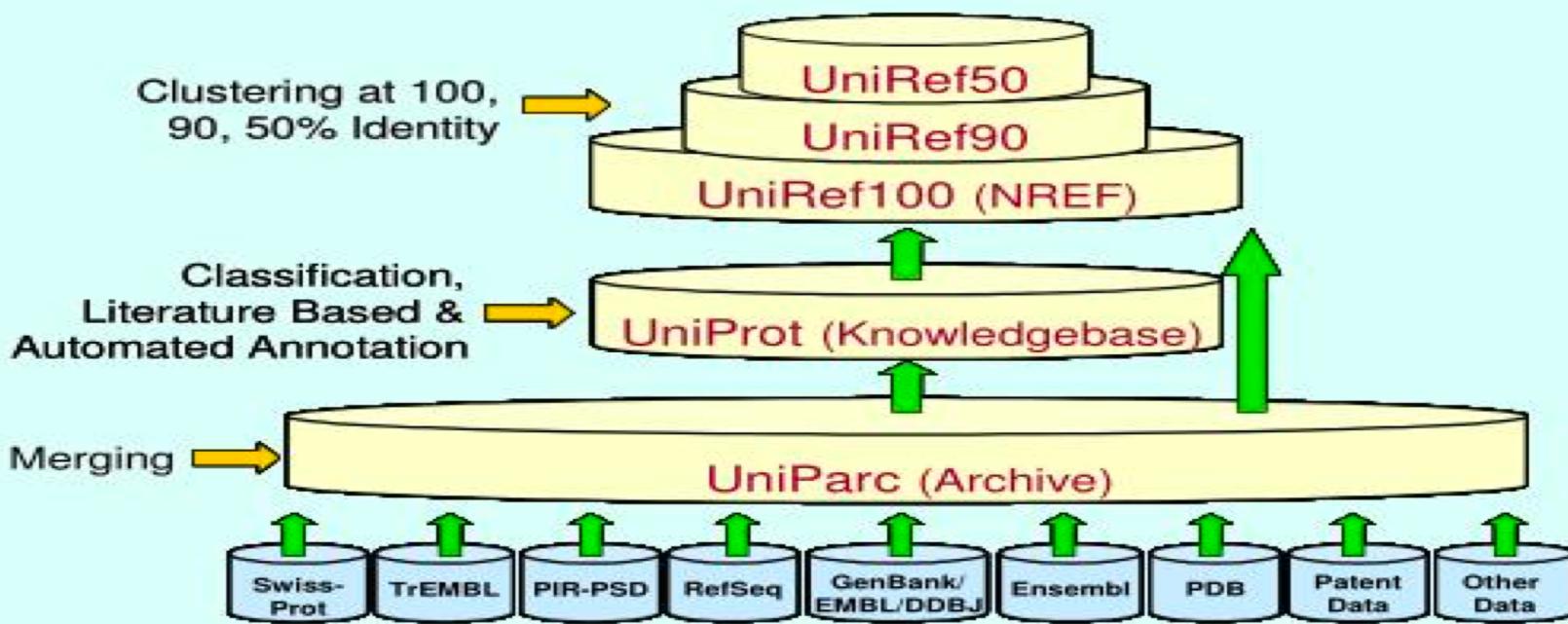
The people and funding behind UniProt

- UniProt is a collaboration between the [European Bioinformatics Institute \(EMBL-EBI\)](#), the [SIB Swiss Institute of Bioinformatics](#) and the [Protein Information Resource \(PIR\)](#).
- Across the three institutes more than [100 people](#) are involved through different tasks such as database curation, software development and support.
- The UniProt consortium is headed by [Alex Bateman](#) (PI), [Cathy Wu](#), and [Alan Bridge](#), supported by [key staff](#), and receives valuable input from an independent [Scientific Advisory Board](#).
- UniProt is supported by the [National Eye Institute \(NEI\)](#), [National Human Genome Research Institute \(NHGRI\)](#), [National Heart, Lung, and Blood Institute \(NHLBI\)](#), [National Institute of Allergy and Infectious Diseases \(NIAID\)](#), [National Institute of Diabetes and Digestive and Kidney Diseases \(NIDDK\)](#), [National Institute of General Medical Sciences \(NIGMS\)](#), and [National Institute of Mental Health \(NIMH\)](#) of the [National Institutes of Health \(NIH\)](#).



UniProt Databases

- UniParc: Comprehensive Sequence Archive with Sequence History
- UniRef: Non-redundant Reference Databases for Sequence Search
- UniProtKB: Knowledgebase with Full Classification and Functional Annotation



About dataset:

Uniprot dataset for *Staphylococcus aureus*: (total number of rows 124,980)

Entry	Protein names (biological process)	Gene names Gene ontology (cellular component)	Organism Organism ID	Ensembl transcript	EnsemblBacteria transcript	Cross-reference (GeneID) Cross-reference (ExpressionAtlas)	Gene ontology Cross-reference (Genevisible)
Q8NX66	Iron-regulated surface determinant protein B (Fur-regulated protein B)	(Staphylococcal iron-regulated protein H) (Staphylococcus aureus surface protein J)	isdB frpB sa				
sirH MW1011	Staphylococcus aureus (strain MW2)	196620	BAB94876;	pathogenesis [GO:0009405]	cell wall [GO:0005618]; extracellular region	[GO:0005576]; membrane [GO:0016020]	[GO:0005576]; membrane [GO:0016020]
				INDUCTION: Repressed by fur in the presence of iron (By similarity). Transcriptionally up-regulated by hydrogen peroxide and to a lesser extent by hypochlorous acid. Slightly down-regulated by human neutrophil azurophilic granule proteins.	{ECO:0000250, ECO:0000269 PubMed:18097052}.		
Q2FUW1	Serine-rich adhesin for platelets (Staphylococcus aureus surface protein A)	sraP sasA SAOUHSC_02990	Staphylococcus aureus (strain NCTC 8325)	93061			
ABD31977;	3921472; pathogenesis [GO:0009405]	cell wall [GO:0005618]; extracellular region [GO:0005576]; membrane [GO:0016020]	INDUCTION: Most high expressed in the transient phase between exponential and stationary growth. A further 2-fold induction occurs in secG or secG/secY2 disruption mutants.	{ECO:0000269 PubMed:20472795}			
A6QG30	Iron-regulated surface determinant protein B (Fur-regulated protein B)	(Staphylococcal iron-regulated protein H) (Staphylococcus aureus surface protein J)	isdB frpB sa				
sirH NWMN_1040	Staphylococcus aureus (strain Newman)	426430	BAF67312;	pathogenesis [GO:0009405]	cell wall [GO:0005618]; extracellular region	[GO:0005576]; membrane [GO:0016020]	[GO:0005576]; membrane [GO:0016020]
				INDUCTION: Repressed by fur in the presence of iron.			
Q7A656	Iron-regulated surface determinant protein B (Fur-regulated protein B)	(Staphylococcal iron-regulated protein H) (Staphylococcus aureus surface protein J)	isdB frpB sa				
sirH SA0976	Staphylococcus aureus (strain N315)	158879	BAB42225;	pathogenesis [GO:0009405]	cell wall [GO:0005618]; extracellular region	[GO:0005576]; membrane [GO:0016020]	[GO:0005576]; membrane [GO:0016020]
				INDUCTION: Repressed by fur in the presence of iron. {ECO:0000250}.			
Q2FZF0	Iron-regulated surface determinant protein B (Fur-regulated protein B)	(Staphylococcal iron-regulated protein H) (Staphylococcus aureus surface protein J)	isdB frpB sa				
sirH SAOUHSC_01079	Staphylococcus aureus (strain NCTC 8325)	93061	ABD30196; 3919242; pathogenesis [GO:0009405]	cell wall [GO:0005618]; extracellular region [GO:0005576]; membrane [GO:0016020]	cell wall [GO:0005618]; membrane [GO:0016020]		
				INDUCTION: Repressed by fur in the presence of iron.			
Q2FXJ2	Iron-regulated surface determinant protein H (Haptoglobin receptor A)	(Staphylococcus aureus surface protein I)	isdH harA sasI SAOUHSC_01843	Staphylococcus aureus (strain NCTC 8325)	INDUCTION:		
	93061	ABD30910; 3920522; cell wall [GO:0005618]; extracellular region [GO:0005576]; membrane [GO:0016020]					

UniProt entries where table will join BacMet

Goals of UniProt

- The accelerating growth of sequenced genomes poses great challenges for databases with their major growth being driven by sequencing of very similar and almost identical strains of the same bacterial species.
- UniProt provides a method to identify and remove highly redundant bacterial proteomes within species groups by performing pairwise alignments of sets of sequences and applies graph theory to find dominating sets that result in removal of redundant bacterial proteomes.

UniProt Stats

Database Cross-Reference Statistics

Database	Entities linked to	Entries
Sequence databases		
EMBL	162,900,792	141,591,179
PIR	162,549	130,313
RefSeq	46,368,754	45,206,991
UniGene	907,896	770,847
3D structure databases		
DisProt	96	96
PDB	40,137	19,105
PDBsum	39,143	18,686
ProteinModelPortal	7,013,797	7,013,797
SMR	1,479,352	1,479,352
Protein-protein interaction databases		
CORUM	245	245
ComplexPortal	195	140
DIP	3,191	3,190
ELM	100	100
IntAct	26,544	26,387
MINT	2,892	2,892
STRING	6,345,439	6,345,178

Citation usage		
Citation type	Citations	Entries
Submission	119,311,135	104,837,977
Journal article	53,453,325	50,811,394
Book	13,508	13,443
Thesis	15,758	15,698
Patent	1	1
Unpublished observations	0	0
Online journal article	0	0

Gene expression databases		
Bgee	529,449	529,248
CollectTF	195	195
ExpressionAtlas	587,986	587,981
Genevisible	15,812	15,805
Ontologies		
Family and domain databases		
CDD	26,053,126	22,860,000
Gene3D	63,181,545	52,475,534
HAMAP	15,895,377	15,713,254
InterPro	370,362,442	112,505,756
PANTHER	33,841,194	32,652,919
PIRSF	12,547,411	12,438,629
PRINTS	18,651,016	16,872,413
PROSITE	71,509,181	47,854,842
Pfam	141,626,841	102,840,522
ProDom	1,944,380	1,870,409
SFLD	910,492	708,177
SMART	33,973,072	25,780,073
SUPFAM	94,076,626	74,515,739
TIGRFAMs	30,203,212	27,784,701

References:

- <https://2rdnmg1qbg403gumla1v9i2h-wpengine.netdna-ssl.com/wp-content/uploads/sites/3/2017/10/mrsa-650x450.jpg>
- The UniProt Consortium (2016). UniProt: the universal protein knowledgebase. *Nucleic acids research*, 45(D1), D158-D169.
- The UniProt Consortium **UniProt: a worldwide hub of protein knowledge** [Nucleic Acids Res. 47: D506-515 \(2019\)](https://doi.org/10.1093/nar/gky493)
- <https://www.google.com/url?sa=i&source=images&cd=&cad=rja&uact=8&ved=2ahUKEwii1vzz2pHhAhWFMd8KHRiwAeoQjRx6BAgBEAU&url=https%3A%2F%2Fwww.slideserve.com%2Ftuari%2Fu-niprot-universal-protein-resource&psig=AOvVaw1Mhjk2IY4isx7MtoMft-s1&ust=1553204618984898>