

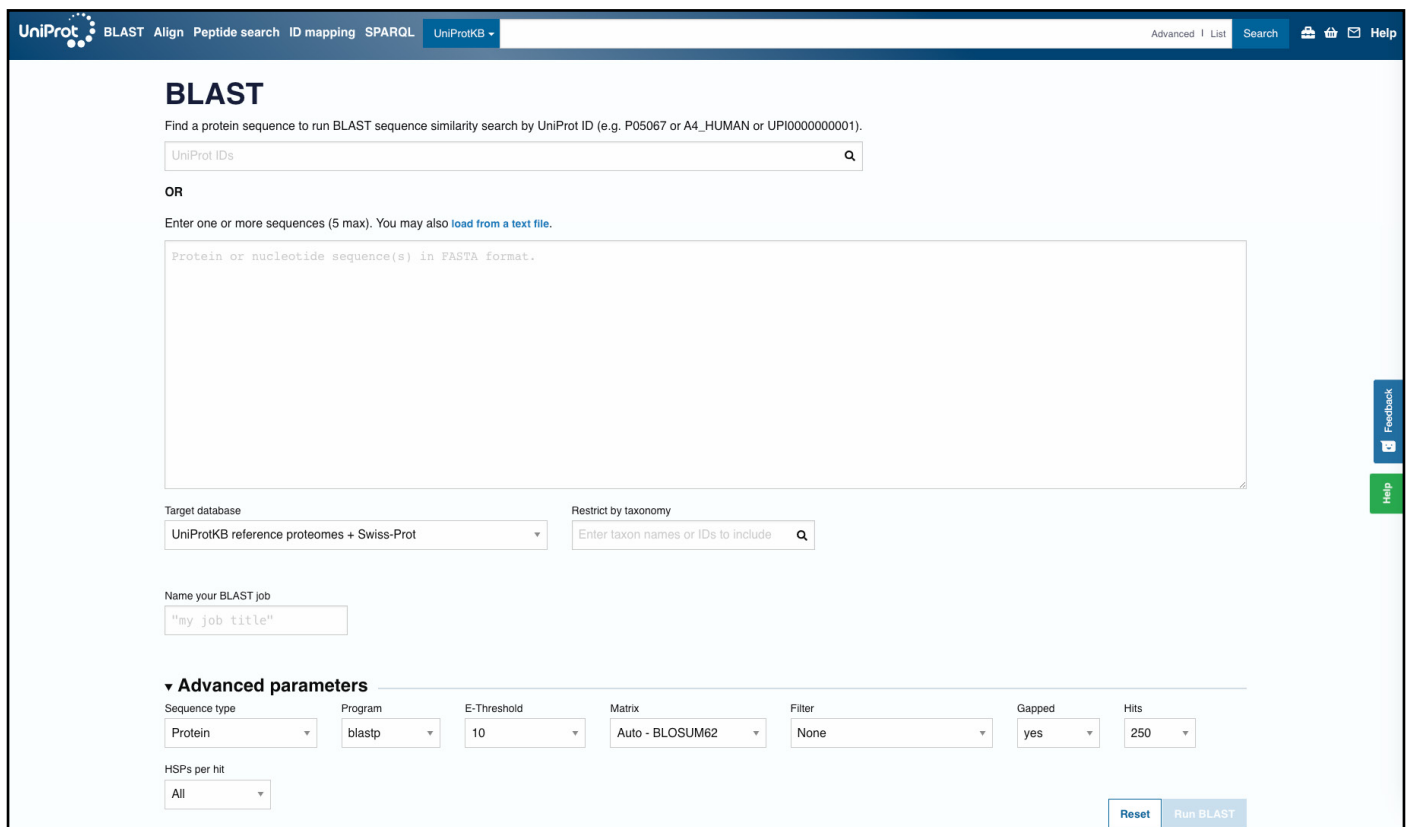
Using Uniprot.org you will complete a BLAST search of the database to discover which protein and species your amino acid sequence comes from.

What is a BLAST search?

BLAST stands for Basic Local Alignment Search Tool. A BLAST search allows researchers to compare an amino acid sequence with records held in a database. This enables them to find the closest matching protein to help them identify what the protein is and how it functions.

How to use Uniprot.org

1. Complete the Function Finders Worksheet to get your amino acid sequence
2. Go to uniprot.org/blast



The screenshot shows the UniProt BLAST search page. At the top, there is a navigation bar with links for BLAST, Align, Peptide search, ID mapping, SPARQL, and UniProtKB. The main heading is "BLAST" with a sub-heading: "Find a protein sequence to run BLAST sequence similarity search by UniProt ID (e.g. P05067 or A4_HUMAN or UPI0000000001)." Below this, there is a search input field for UniProt IDs. An "OR" section follows, with instructions to "Enter one or more sequences (5 max). You may also load from a text file." A large text area is provided for entering the protein or nucleotide sequence in FASTA format. Below the text area, there are dropdown menus for "Target database" (set to UniProtKB reference proteomes + Swiss-Prot) and "Restrict by taxonomy" (with a search field). A field for naming the BLAST job is also present. The "Advanced parameters" section includes dropdowns for Sequence type (Protein), Program (blastp), E-Threshold (10), Matrix (Auto - BLOSUM62), Filter (None), Gapped (yes), Hits (250), and HSPs per hit (All). "Reset" and "Run BLAST" buttons are at the bottom right.

Function Finders: BLAST!

Uniprot instructions

3. Type in your amino acid sequence. You might also want to name your search, such as *Sequence 1*. Click **Run BLAST**.

BLAST
Find a protein sequence to run BLAST sequence similarity search by UniProt ID (e.g. P05067 or A4_HUMAN or UPI0000000001).

UniProt IDs

OR
Enter one or more sequences (5 max). You may also [load from a text file](#).

MKSAILTGLLPV

Your input contains 1 sequence

Target database: UniProtKB reference proteomes + Swiss-Prot
Restrict by taxonomy: Enter taxon names or IDs to include

Name your BLAST job: Sequence 1

Advanced parameters

Sequence type: Protein
Program: blastp
E-Threshold: 10
Matrix: Auto - PAM30
Filter: None
Gapped: yes
Hits: 250

HSPs per hit:

4. You might see a loading page, sometimes the search can take a few minutes, so you can go back to the first page and start searching your other sequences whilst the first one searches.

Tool results
Your tool analysis results from the last 7 days are listed below. If you have tools jobs running, you can navigate away to other pages and you will be notified once the job is completed.

Job type	Name	Created	Status
BLAST	Sequence 1	2023-11-03 09:03	Running

We will notify you when your results are ready

ncbiblast-R20231103-090357-0206-95495379-p1m Target database: UniProtKB reference proteomes + Swiss-Prot

Function Finders: BLAST!

Uniprot instructions

5. Once the database has finished searching, click **Completed** to access the results. If it returns a no results outcome, double check your amino acid sequence.

The screenshot shows the UniProt BLAST interface. At the top, there are navigation links: BLAST, Align, Peptide search, ID mapping, SPARQL, and UniProtKB. On the right, there are options for 'Advanced | List', a search bar, and icons for home, mail, and help. The main heading is 'Tool results'. Below this, a message states: 'Your tool analysis results from the last 7 days are listed below. If you have tools jobs running, you can navigate away to other pages and you will be notified once the job is completed.'

Job type	Name	Created	Status
BLAST	Sequence 1	2023-11-03 09:39	Completed (20 hits)

Below the table, it says 'Target database: UniProtKB reference proteomes + Swiss-Prot'. There are also icons for star, share, and delete. A 'Feedback' button is visible on the right side.

6. You'll be presented with a list of results, and first you need to filter by **Reviewed Status** - this limits the results to ones that have been manually reviewed.

The screenshot shows the UniProt BLAST results page. The main heading is 'BLAST 20 results found in UniProtKB'. There are tabs for 'Overview', 'Taxonomy', 'Hit Distribution', 'Text Output', 'Input Parameters', and 'API Request'. The 'Overview' tab is selected. On the left, there are 'Blast parameters' including Identity (64.7), Score (65), and E-Value (0.0034). There are also 'Taxonomy' and 'Status' filters. The 'Status' filter is highlighted with a red box, showing 'Reviewed (Swiss-Prot) (7)' and 'Unreviewed (TrEMBL) (13)'. Below the filters, there is a table of results with columns: Entry, Entry Name, Protein Names, Gene Names, Organism, Length, and a progress bar. The table shows 20 results, with the first few being Type-3 ice-structuring proteins from Anarhichas lupus and Zoarces americanus.

Entry	Entry Name	Protein Names	Gene Names	Organism	Length
P12417	ANP2_ANALU	Type-3 ice-structuring protein 1.5[...]		Anarhichas lupus (Atlantic wolffish)	88 AA
P12416	ANP1_ANALU	Type-3 ice-structuring protein 1.9[...]		Anarhichas lupus (Atlantic wolffish)	88 AA
P19606	ANP3_ZOAM	Ice-structuring protein lambda OP-3[...]		Zoarces americanus (Ocean pout) (Macrozoarces americanus)	91 AA
P19605	ANPE_ZOAM	Ice-structuring protein C10[...]		Zoarces americanus (Ocean pout) (Macrozoarces americanus)	87 AA
P19604	ANPD_ZOAM	Ice-structuring protein C7[...]		Zoarces americanus (Ocean pout) (Macrozoarces americanus)	87 AA
P19607	ANP5_ZOAM	Ice-structuring protein lambda OP-5[...]		Zoarces americanus (Ocean pout) (Macrozoarces americanus)	87 AA
P07457	ANP1C_ZOAM	Ice-structuring protein SP1-C [...]		Zoarces americanus (Ocean pout) (Macrozoarces americanus)	87 AA

7. From the reduced list of entries, look for the list of E values on the right hand side of the page. Find the entry with the lowest E value and highest % identify and click on the **Entry number**.

BLAST 7 results found in UniProtKB

Entry	Entry Name	Protein Names	Gene Names	Organism	Length	Identity	E-Value
P12417	ANP2_ANALU	Type-3 Ice-structuring protein 1.5[...]		Anarhichas lupus (Atlantic wolffish)	88 AA	100%	0.0034
P12416	ANP1_ANALU	Type-3 Ice-structuring protein 1.9[...]		Anarhichas lupus (Atlantic wolffish)	88 AA	100%	0.0034
P19606	ANP3_ZOAAAM	Ice-structuring protein lambda OP-3 [...]		Zoarces americanus (Ocean pout) (Macrozoarces americanus)	91 AA	91.7%	0.054
P19605	ANPE_ZOAAAM	Ice-structuring protein C10[...]		Zoarces americanus (Ocean pout) (Macrozoarces americanus)	87 AA	91.7%	0.054
P19604	ANPD_ZOAAAM	Ice-structuring protein C7[...]		Zoarces americanus (Ocean pout) (Macrozoarces americanus)	87 AA	91.7%	0.054
P19607	ANP5_ZOAAAM	Ice-structuring protein lambda OP-5 [...]		Zoarces americanus (Ocean pout) (Macrozoarces americanus)	87 AA	91.7%	0.054
P07457	ANP1C_ZOAAAM	Ice-structuring protein SP1-C[...]		Zoarces americanus (Ocean pout) (Macrozoarces americanus)	87 AA	91.7%	0.054

What is an E value?

The E value, or Expect value, is a measure of how statistically significant your result is. E values indicate the number of hits you would expect to achieve by chance within a database of a particular size.

For example, an E value of 1.0 means you would expect to get one match in the database purely by chance. The lower the E value or the closer it is to zero, the less likely that match is to have occurred by chance. An E value should ideally be in the range of 0.01 to 0.1.

What is the % identity?

The % identity is a measure of how closely your protein sequence matches with an entry in the database. 100% match means the sequence is exactly the same.

Using these two feature gives you an idea of how reliable your results are.

Function Finders: BLAST!

Uniprot instructions

8. On your results page, find the **Protein name**, **Organism** and **Function**, and add those to your worksheet.

UniProt BLAST Align Peptide search ID mapping SPARQL UniProtKB

P12417 · ANP2_ANALU

Protein Type-3 ice-structuring protein 1.5

Status UniProtKB reviewed (Swiss-Prot)

Organism Anarhichas lupus (Atlantic wolffish)

Amino acids 88 (go to sequence)

Protein existence Inferred from homology

Annotation score 2/5

Function
Contributes to protect fish blood from freezing at subzero sea water temperatures. Lowers the blood freezing point. Binds to nascent ice crystals and prevents further growth (By similarity).

Features
Showing features for site¹.

TYPE	ID	POSITION(S)	DESCRIPTION
Site		30	Important for ice-binding
Site		35	Important for ice-binding
Site		39	Important for ice-binding
Site		65	Important for ice-binding

There's lots more information on the result page that you can explore, for example the 3D structure of the protein under **Structure**.

UniProt BLAST Align Peptide search ID mapping SPARQL UniProtKB

Structure

Model Confidence:

- Very high (pLDDT > 90)
- Confident (90 > pLDDT > 70)
- Low (70 > pLDDT > 50)
- Very low (pLDDT < 50)

AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions with low pLDDT may be unstructured in isolation.

SOURCE	IDENTIFIER	METHOD	RESOLUTION	CHAIN	POSITIONS	LINKS
AlphaFold	AF-P12417-F1	Predicted			1-88	AlphaFold Foldseek

3D structure databases

AlphaFoldDB | [P12417](#)

ModBase | [Search...](#)