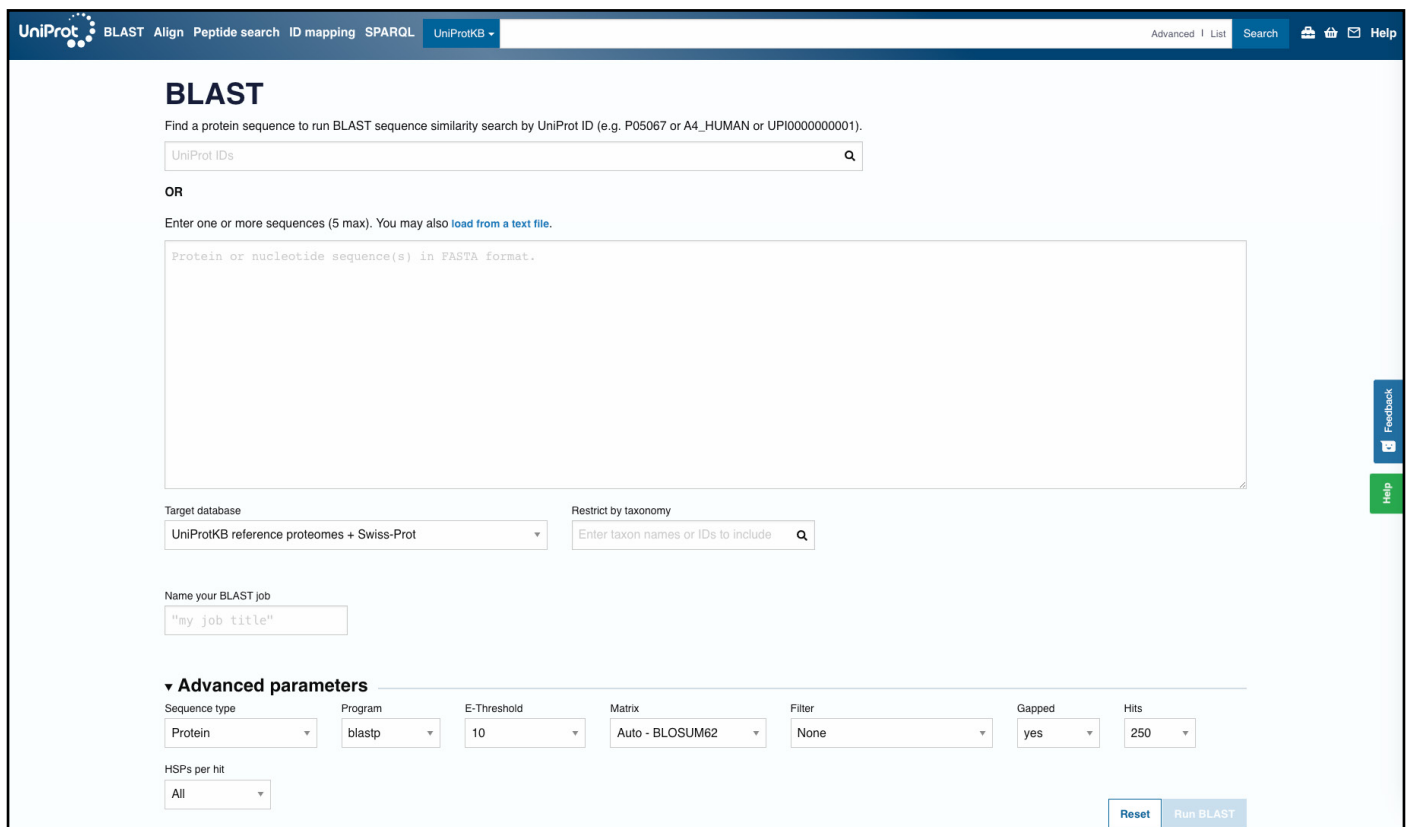# Function Finders: BLAST!
Uniprot instructions

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

yourgenome.org

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

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.

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.

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.

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**.

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.

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

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