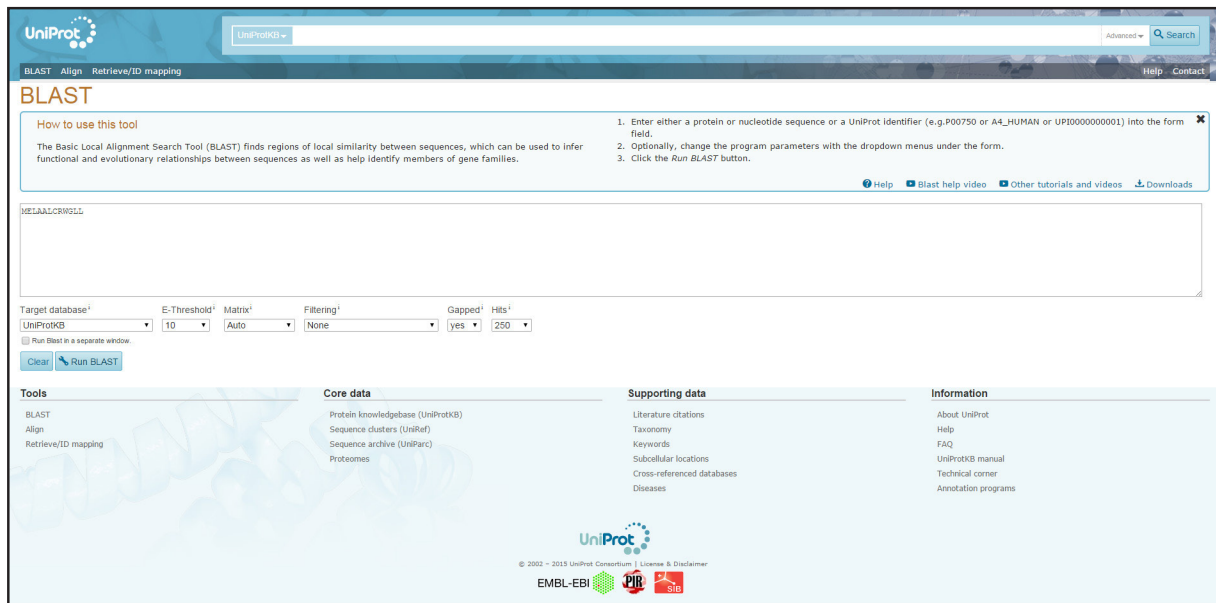


FUNCTION FINDERS BLAST

Worksheet instructions

HOW TO COMPLETE THE WORKSHEETS

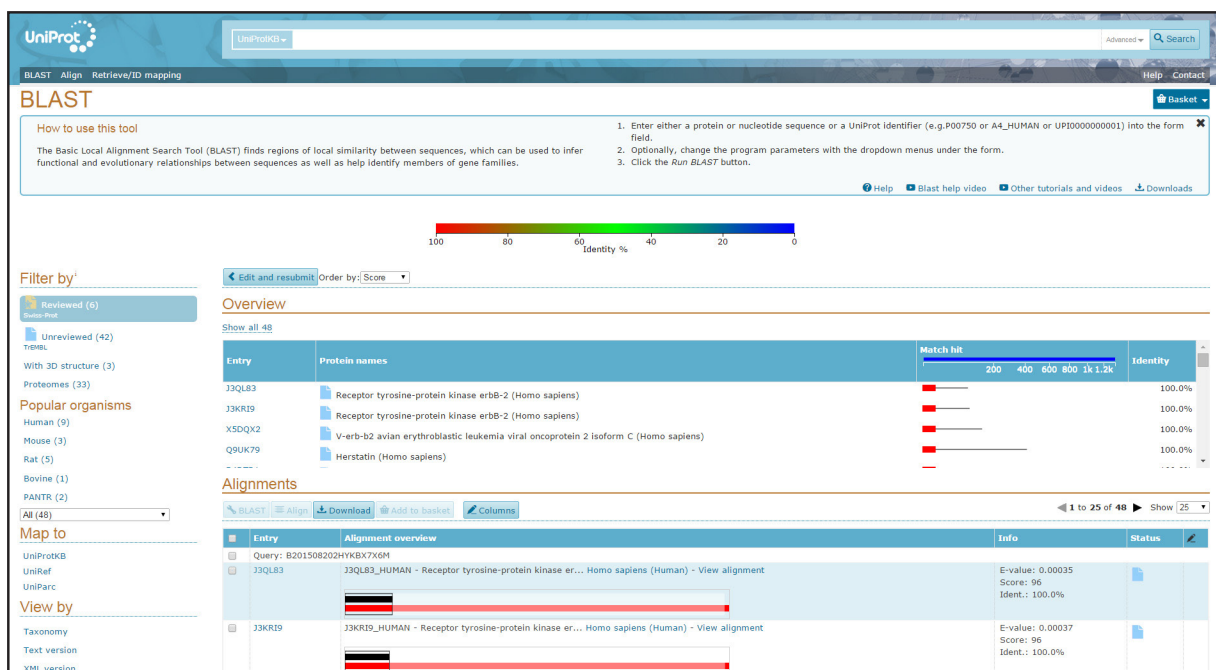
1. Use the codon wheel to **translate the DNA sequences** on the worksheet to amino acids.
2. **Type the amino acid sequence** in to the Uniprot Blast search www.uniprot.org/blast/.



The screenshot shows the UniProt BLAST search page. At the top, there's a search bar with 'UniProtKB' entered. Below it, the 'BLAST' section is active. A 'How to use this tool' box provides instructions. The search parameters are: Target database: UniProtKB, E-Threshold: 10, Matrix: Auto, Filtering: None, Gapped: yes, Hits: 250. The 'Run BLAST' button is highlighted. The page also features navigation tabs for Tools, Core data, Supporting data, and Information.

Press **Run BLAST** to get results (it may take a few seconds for results to appear).

3. When the search results appear, **filter the results** to only show reviewed entries (gold file icon with a star). Each result is known as a "hit".



The screenshot shows the UniProt BLAST search results page. The search bar contains 'UniProtKB'. The 'Filter by' section is active, showing 'Reviewed (6)' and 'Unreviewed (42)'. The 'Overview' section shows a table of results with columns for Entry, Protein names, Match hit, and Identity. The 'Alignments' section shows detailed alignment information for the top results.

Entry	Protein names	Match hit	Identity
J3QL83	Receptor tyrosine-protein kinase erbB-2 (Homo sapiens)	200 400 600 800 1k 1.2k	100.0%
J3KR19	Receptor tyrosine-protein kinase erbB-2 (Homo sapiens)		100.0%
XSDQX2	V-erb-b2 avian erythroblastic leukemia viral oncogene protein 2 isoform C (Homo sapiens)		100.0%
Q9UK79	Herstatin (Homo sapiens)		100.0%

FUNCTION FINDERS BLAST

Worksheet instructions

4. After filtering the hits should look like this. **Scroll down** to the Overview section.

The screenshot shows the BLAST results interface. The 'Overview' section lists six hits with their accession numbers and descriptions, along with their identity percentages. The 'Alignments' section shows a table of alignment details for each hit, including E-value, Score, and Identity. A red box highlights the 'Info' column in the Alignments table.

Entry	Alignment overview	Info	Status
PD4626-6	ISOFORM 6 OF RECEPTOR TYROSINE-PROTEIN KINASE ERBB-2 (HOMO SAPIENS)	E-value: 0.00060 Score: 96 Ident.: 100.0%	
PD4626	RECEPTOR TYROSINE-PROTEIN KINASE ERBB-2 (HOMO SAPIENS)	E-value: 0.00061 Score: 96 Ident.: 100.0%	
O18735	RECEPTOR TYROSINE-PROTEIN KINASE ERBB-2 (CANIS FAMILIARIS)	E-value: 0.042 Score: 83 Ident.: 91.7%	
P70424	RECEPTOR TYROSINE-PROTEIN KINASE ERBB-2 (MUS MUSCULUS)	E-value: 1.0 Score: 73 Ident.: 83.3%	
P06494	RECEPTOR TYROSINE-PROTEIN KINASE ERBB-2 (RATTUS NORVEGICUS)	E-value: 1.0 Score: 73 Ident.: 83.3%	
Q60553	RECEPTOR TYROSINE-PROTEIN KINASE ERBB-2 (MESOCRICETUS AURATUS)	E-value: 9.5 Score: 66 Ident.: 83.3%	

Look at the info column. This will give you an idea of how reliable your hits are.

The Expect value (E-value) indicates the number of random hits you would expect by chance for the given query sequence and the size of the sequence database against which the BLAST is performed.

For example, an E-value of 1.0 means that you would expect on average to get one match in the database for the submitted query simply by chance. The lower the E-value, or the closer it is to zero, the more "significant" the match is. In general, the E-values should be in the range of 0.01 to 0.1 to be statistically significant.

The identity % describes how similar your sequence is to the hit, i.e. whether the amino acids are in the same position when aligned. 100% means the sequences match exactly.

FUNCTION FINDERS BLAST

Worksheet instructions

5. **Click on the hit** that you think best matches your sequence. Find out the name of the protein which species the sequence is from (common and species name) and what the protein does. Is the protein found in other species?

The screenshot shows the UniProtKB entry for UniProtKB - P04626 (ERBB2_HUMAN). The protein is identified as Receptor tyrosine-protein kinase erbB-2. The organism is Homo sapiens (Human). The status is Reviewed with an annotation score of 5.0. The function section describes it as a protein tyrosine kinase that is part of several cell surface receptor complexes. The catalytic activity is ATP + a [protein]-L-tyrosine = ADP + a [protein]-L-tyrosine phosphate. The enzyme regulation is activated by dimerization. The sites section includes a binding site at position 753-753 and an active site at position 845-845. The regions section includes a nucleotide binding site at position 726-734. The GO - Molecular function section lists ATP binding, ErbB-3 class receptor binding, identical protein binding, protein C-terminus binding, and protein dimerization activity.

You can reduce the amount of information on the screen by unticking the blue display categories on the left hand side of the screen.

If you cannot find all the information you need, try using a Google search or Wikipedia to find out more.

6. To start a new BLAST search click **Edit and resubmit**, and enter your next set of amino acids. Repeat steps 3 to 5.