Bioinformatics

CASEIN - A MAMMALIAN MILK PROTEIN

A bioinformatics study to complement our new casein practical for AH Biology

Curriculum links

- **Higher Human Biology:** Human Cells, Key Area 5a – Computer programs can be used to identify base sequences by looking for sequences similar to known genes. To compare sequence data, computer and statistical analyses (bioinformatics) are required.

- **Higher Biology:** DNA and the genome, Key Area 8a – statement as above for Higher Human Biology.

- **Advanced Higher Biology:** Cells and Proteins, Key Area 2c – Protein structure, ligand binding and conformation change.
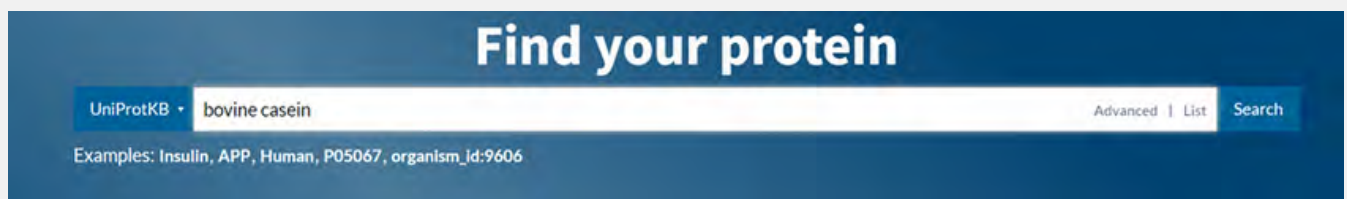
Milk is a rich source of protein (approximately 36g protein per litre of milk). The major protein constituents are whey and casein; in bovine milk, casein proteins (α, β, γ, and κ) account for 80% of milk proteins. The casein proteins are found as self-assembled particles called "micelles". SSERC have developed a practical activity to investigate the mass of casein protein in mammalian milk. Prior to this practical work, a bioinformatics approach to learning more about this protein group is beneficial.

METHOD

FINDING A PROTEIN SEQUENCE USING UNIPROT

STEP 1
Go to Uniprot – type "bovine casein" into the search box. Press "Search".

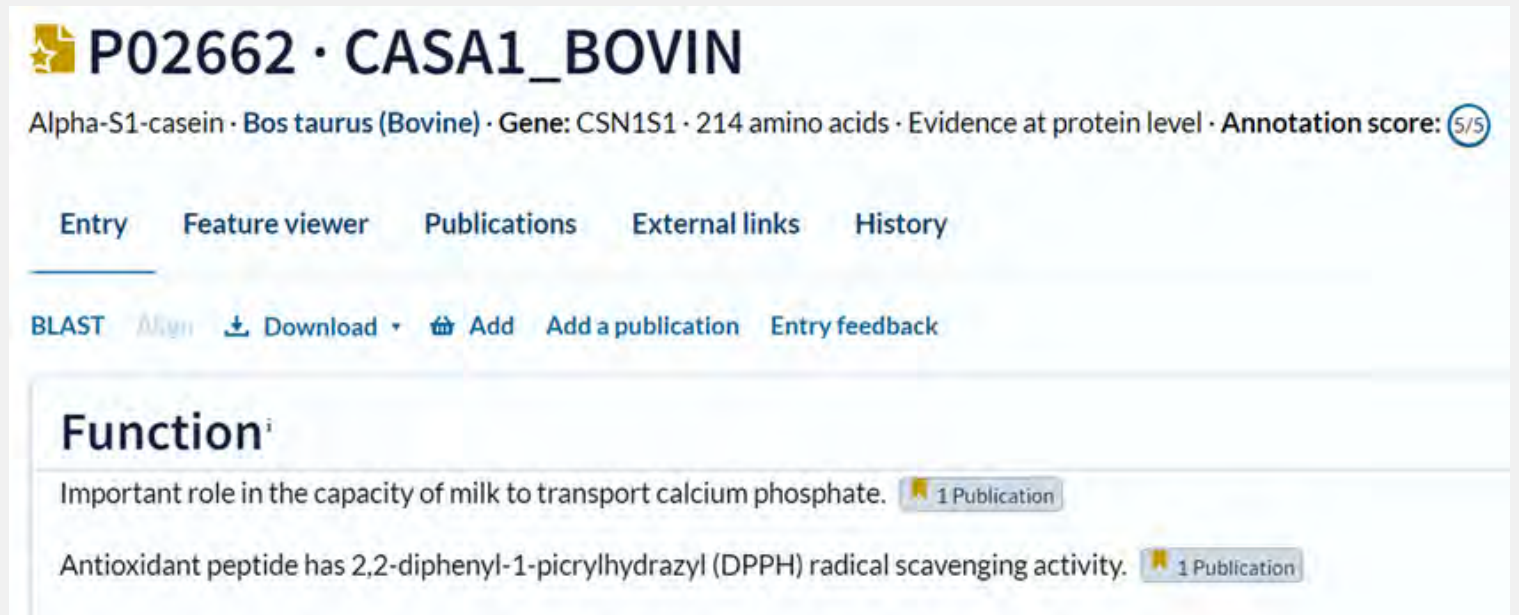
STEP 2
Uniprot will return a series of entries. Click on "P02662" to access data on α-S1-casein. You can already see from the Uniprot results page that this is a small protein of 214 amino acids.

STEP 3
The following screen will be displayed.

⭐ P02662 · CASA1_BOVIN
Alpha-S1-casein · Bos taurus (Bovine) · Gene: CSN1S1 · 214 amino acids · Evidence at protein level · Annotation score: (5/5)

Entry   Feature viewer   Publications   External links   History

BLAST   Align   ⬇ Download ▾   🎁 Add   Add a publication   Entry feedback

Function ⁱ
Important role in the capacity of milk to transport calcium phosphate. 🔖 1 Publication
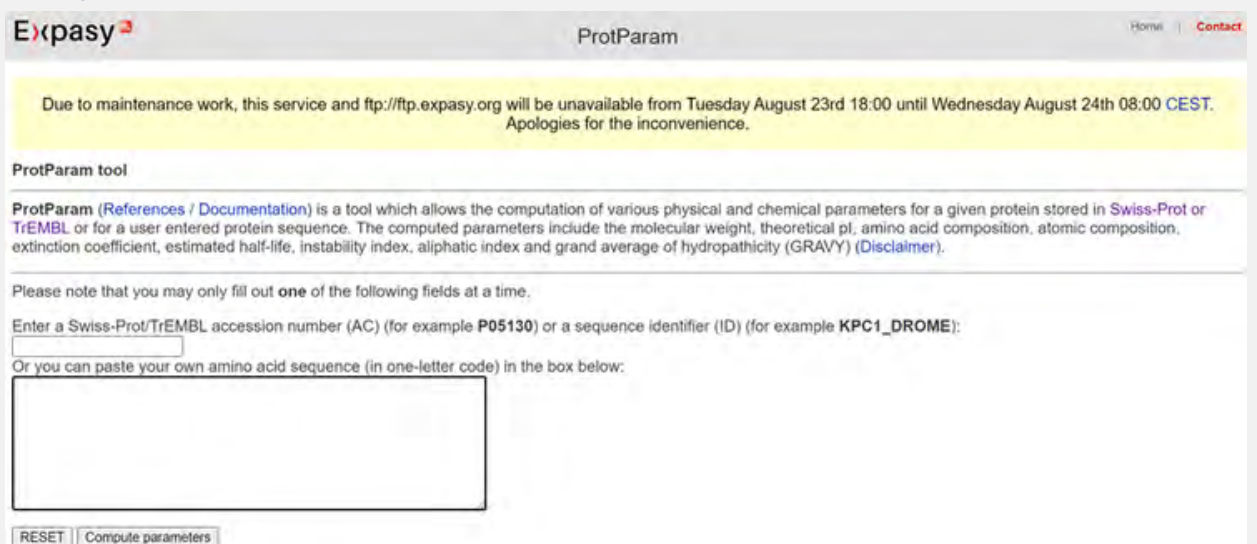
Antioxidant peptide has 2,2-diphenyl-1-picrylhydrazyl (DPPH) radical scavenging activity. 🔖 1 Publication

Use the information on this page to complete the first five entries in the "Results Table" at the bottom of the page. Do not close this tab – we will use it again later.

FINDING A PROTEIN'S ISOELECTRIC POINT USING PROTPARAM

STEP 4
Go to the ProtParam website. Paste the amino acid sequence for α-S1-Casein into the box. Click "Compute parameters". Record the molecular weight (in kilodaltons, kDa) and the theoretical pI (this is the isoelectric point) of the protein. Remember this value when you move onto the practical activity.

Expasy   ProtParam   Home | Contact

Due to maintenance work, this service and ftp://ftp.expasy.org will be unavailable from Tuesday August 23rd 18:00 until Wednesday August 24th 08:00 CEST. Apologies for the inconvenience.

ProtParam tool

ProtParam (References / Documentation) is a tool which allows the computation of various physical and chemical parameters for a given protein stored in Swiss-Prot or TrEMBL or for a user entered protein sequence. The computed parameters include the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY) (Disclaimer).

Please note that you may only fill out **one** of the following fields at a time.

Enter a Swiss-Prot/TrEMBL accession number (AC) (for example **P05130**) or a sequence identifier (ID) (for example **KPC1_DROME**):

Or you can paste your own amino acid sequence (in one-letter code) in the box below:

RESET | Compute parameters

FINDING HOMOLOGOUS PROTEINS IN OTHER SPECIES USING BLAST

STEP 5
Return to the Uniprot tab for P02662 CASA1_BOVIN. At the top of the page, click "BLAST".

⭐ **P02662 · CASA1_BOVIN**

Alpha-S1-casein · Bos taurus (Bovine) · Gene: CSN1S1 · 214 amino acids · Evidence at protein level · **Annotation score:** (5/5)

Entry   Feature viewer   Publications   External links   History

BLAST   Align   ⬇ Download ▾   🛒 Add   Add a publication   Entry feedback

STEP 6
The page that opens will present the amino acid sequence in the second box. This tool allows us to compare the similarity of the Bovine Casein protein to any other. In the practical work, you will look at the mass of protein in cow, sheep and goat milk. Let's compare the bovine sequence to the sheep casein sequence.

STEP 7
In the "Restrict by taxonomy" box, type "sheep" and select the first option that appears ("Ovis aries"). Change the "Hits" to 50. Click "Run BLAST". This will take a few minutes. When done, click "Completed".

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	sp|P02662|CASA1_BOVIN	2022-08-23 11:36	Completed ●	
ncbiblast-R20220823-113634-0330-32475650-p1m				

STEP 8

Select the tick box beside "P04653" – this is a reviewed entry as indicated by the gold icon to the left of the accession number (unlike the top entry shown in the screenshot).
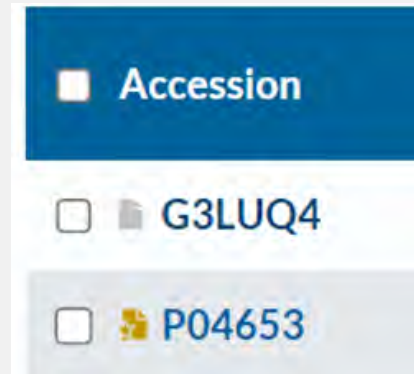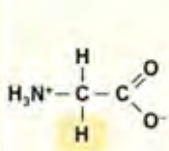

STEP 9

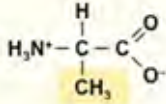Then click "Align" and then "Align selected result with query".
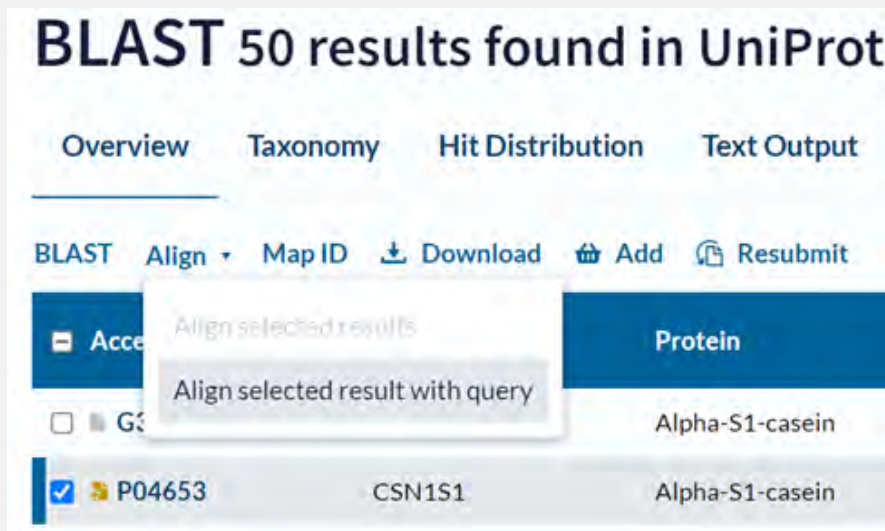
STEP 10

In the window that appears, click "Run Align". This will take a few minutes. When finished, click "Completed". Using the information available on this page, complete the remaining row in the "Results Table" at the end of this page. Can you make any observations about the non-identical regions of the alignment? You could refer to your class notes for Cells & Proteins (Key Area 2c) or the image on the next page.
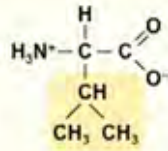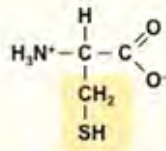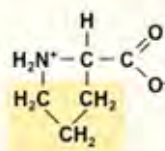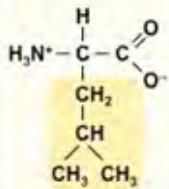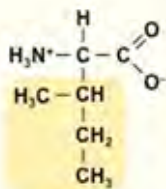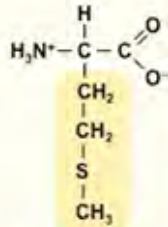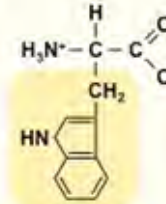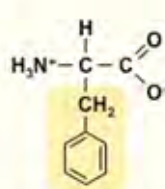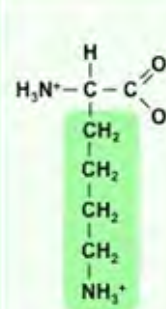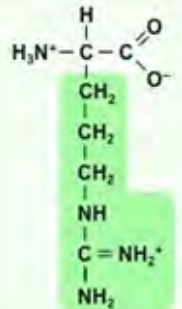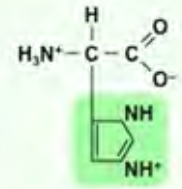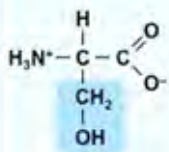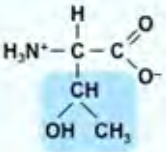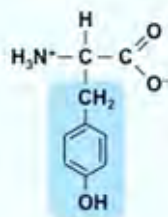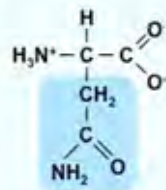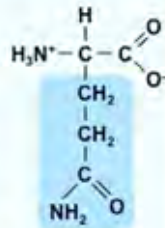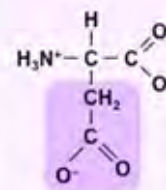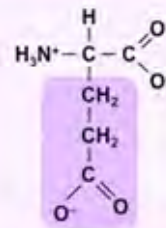
FINDING A PROTEIN STRUCTURE USING DEEPMIND ALPHAFOLD PROTEIN STRUCTURE DATABASE

Casein proteins are difficult to crystallise and so the protein data bank cannot be used to find crystal structures of α–S1–casein. However, the new AlphaFold Protein Structure database uses an algorithm to predict a protein's structure from its amino acid sequence. It is important to note that the results displayed have not been experimentally verified; they are predictions based on protein folding rules.

AlphaFold
Protein Structure Database

Developed by DeepMind and EMBL-EBI

Search for protein, gene, UniProt accession or organism | BETA | **Search**

Examples: Free fatty acid receptor 2 | At1g58602 | Q5VSL9 | E. coli | Help: AlphaFold DB search help

Feedback on structure: Contact DeepMind

STEP 11

Type in the UniProt accession number for the bovine casein protein we have been studying (P02662). Click on "Alpha-S1-Casein" when the results are displayed.
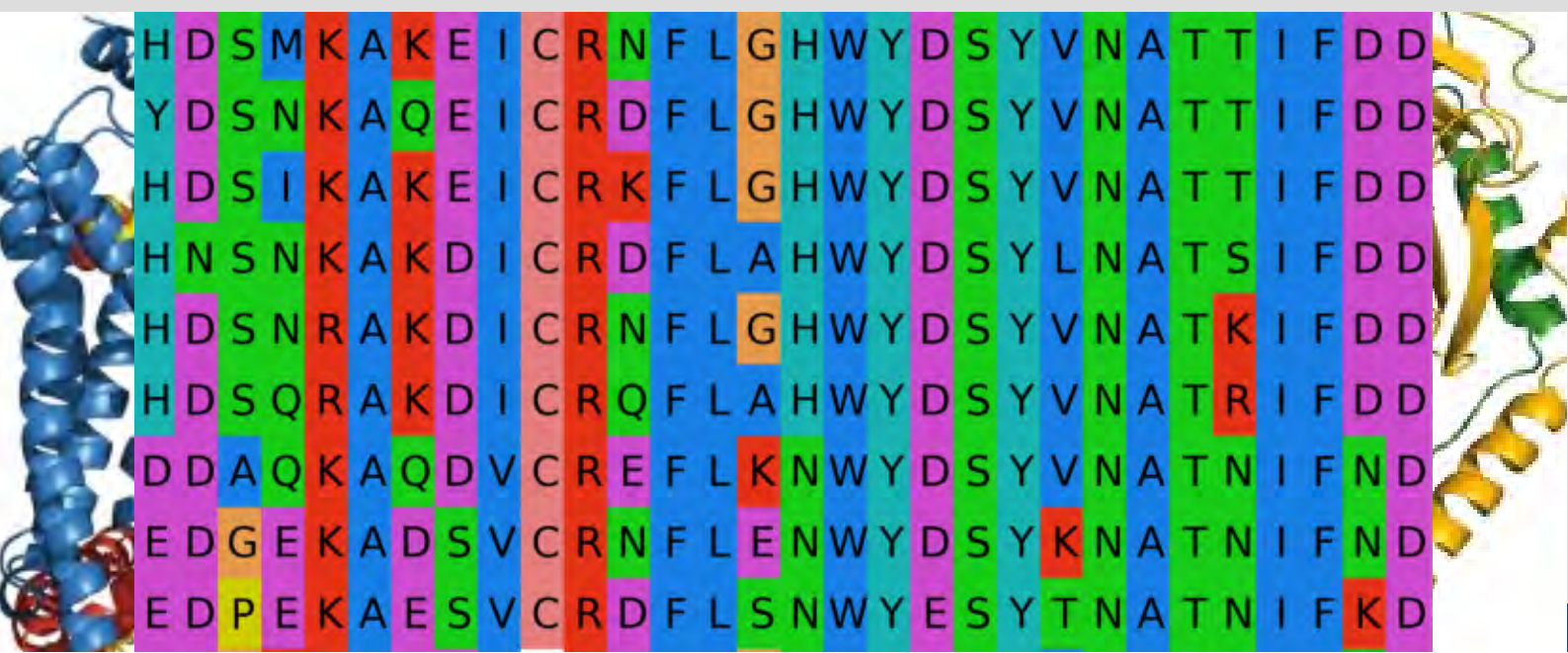

Alpha-S1-casein

P02662 (CASA1_BOVIN)

Protein	Alpha-S1-casein
Gene	CSN1S1
Source Organism	Bos taurus search this organism
UniProt	P02662 go to UniProt

STEP 12

Scroll down to observe the predicted structure and complete the final row in the Results table.

RESULTS

Results Table – α-S1-casein (Bovine)

Number of amino acids	
Function	
Organism	
Subcellular location	
Amino acid sequence (copy and paste in the right-hand cell)	
Molecular weight (kDa)	
Isoelectric point	
Sequence identity to sheep Casein	
Observed secondary structure	