

FUNCTION FINDERS BLAST

Teacher's notes

BACKGROUND TO ACTIVITY

Function Finders BLAST provides a hands-on exercise that introduces the concept of genes coding for proteins. The activity involves translating DNA sequences into amino acid chains and using this information to find a matching protein with the correct corresponding sequence.

Estimated duration: 30 – 45 minutes

MATERIALS TO RUN THE ACTIVITY

- Student worksheets
- Codon Wheel sheets
- Function Finders presentation files
- Computer (one between two students)
- Internet connection for each computer

Optional animations (recommended for AS Level)

DNA to protein animation: www.yourgenome.org/video/from-dna-to-protein

Optional activity: View the proteins in 3D

Rasmol (available from www.rasmol.org) is a molecular modelling software that can be used to view proteins in 3D. It can be used by the teacher or students to view the proteins they have identified and initiate discussions about the tertiary structures of proteins. It is not an essential part of the activity.

ACTIVITY PREPARATION

The following activity components need to be prepared before the activity starts.

Working in pairs, students require:

- One Function Finders worksheet
- One Codon Wheel sheet
- One computer

We recommend one computer between two students, but the activity can be run with groups of three students.

OPTIONAL ACTIVITY PREPARATION

FUNCTION FINDERS BLAST

Teacher's notes

1. Download Rasmol program (optional)

If you wish to use the Rasmol programme to view the protein structure, download the programme from www.rasmol.org and select "Latest windows installer" at the top of the page.

2. Save protein files

If you plan to use the Rasmol programme ensure you have the protein files saved in an accessible folder on the laptops. The protein files are downloadable from the Function Finders BLAST resource page on www.yourgenome.org/activities/function-finders-blast.

INSTRUCTIONS FOR RUNNING THE ACTIVITY

Introducing the Codon Wheel

Slide 3 of the Function Finders presentation demonstrates how to use the codon wheel.

The codon wheel is central to the activity as it will translate the DNA sequence into amino acids.

To use the wheel students must work from the inside out. For example if the first triplet of the sequence is CAT, the amino acid it codes for is H (Histidine).

Completing the worksheet

The worksheet has eight different DNA sequences that the students have to translate into a sequence of 12 amino acid letters using the codon wheel. Once the students have completed their translation, they must find the protein by entering the amino acid sequence into the UniProt database. They can then enter the protein name, organism name and protein function on to the worksheet.

SEARCHING THE UNIPROT DATABASE

In this activity the students will be carrying out a BLAST search using the protein database UniProt.

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 do I carry out a BLAST search?

Follow the steps below to carry out a search on UniProt. The student instructions provide a step by step visual guide.

1. Go to the website www.uniprot.org/blast/
2. Enter the amino acid sequence into the search box
3. Press "Run BLAST"
4. Filter the results by "Reviewed" entries only (icon with a gold star)
5. Scroll down to the "Alignments" section
6. Click on the entry with the lowest E value and the highest % identity.

What does the E value tell me?

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 does the % identity tell me?

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

What information can I get from UniProt?

The UniProt database is a live database used by scientists around the world. It provides a lot of information on individual proteins including:

- Protein name
- Name of the gene that encodes for the protein
- Organism the protein is from
- Function of the protein
- Structure of the protein
- Protein sequence

As the main users of the UniProt database are scientists the language can be quite technical, however you can encourage students to look up and explore some of the unfamiliar terms in the text.

DISCUSSION POINTS

The section below provides some additional information on the proteins to encourage a group discussion on the role and relevance of the different proteins.

1. Antifreeze Protein type III

The Antifreeze protein (AFP) is essential to prevent organisms like the wolf fish from freezing in extremely cold conditions.

Q. Is this just exclusive to this one species / what other species may use AFPs?

AFP's are also found in some species of plants, bacteria and invertebrates, such as the Snow flea.

Q. How can this protein be relevant to us?

Understanding the structure and function of AFPs can help improve the preservation of donor organs. Currently it is not possible to freeze organs harvested for transplant operations as the freezing process irreversibly damages the tissue. However scientists are studying AFPs to see if they can be used to successfully freeze, thaw and transplant organs such as hearts and kidneys.

www.the-scientist.com/?articles.view/articleNo/34190/title/Icing-Organs/

Antifreeze proteins also have applications in food science, for example preventing ice crystallisation in ice cream, creating smoother creamier ice cream.

www.phys.org/news/2008-01-edible-antifreeze-unwanted-ice-crystals.html

2. Green Fluorescent Protein (GFP)

Green fluorescent proteins are used to communicate by organisms such as jellyfish (*Aequorea victoria*). It can also be used as a reporter gene to identify if genetic modification has been successful or to see where in an organism a particular protein is being produced. Other fluorescent proteins exist and are used as reporter genes.

www.newscientist.com/article/dn17003-fluorescent-puppy-is-worlds-first-transgenic-dog/

Q. Why is this protein relevant to us?

Reporter genes such as GFP are important for development studies. GFPs can be used to track proteins in developing embryos to get a better understanding of the location and function of the protein products of specific genes.

The image in the PowerPoint shows one day old zebrafish embryos that express green fluorescent protein in the muscle precursor cells (myotomes).

3. Luciferase

Luciferase is a protein used by fireflies to attract a mate. Luciferase catalyses a chemical reaction where chemical energy is converted to light energy. This process of light production by a living organism is called bioluminescence.

Q. What organisms use bioluminescence

A range of organisms use bioluminescence including plankton, fungi, bacteria and insects, such as fireflies and glow worms.

Q. Why is this protein relevant to us?

Like GFPs, luciferase systems are widely used as reporter genes in molecular biology techniques used in the lab.

4. Odorant receptor protein 1 (OR1)

Odorant receptor protein 1 enables *Anopheles* mosquitoes to identify human sweat and find a source of a blood meal.

Odorant receptor proteins allow organisms to detect chemical compounds in their local environment, which can help them locate and identify food or a potential mate. There are a wide range of different odorant receptors, with as many as 1,000 odorant receptor genes in the human genome. Not all of these genes are expressed or produce a functional protein.

Q. Why is this protein relevant?

Malaria is a major global disease killing around one million people every year, mostly children under the age of 5 years old. Understanding this protein and the odour compound it recognises could be used to develop solutions to help with the battle against the disease.

For example, creating a mosquito repellent that blocks the odorant receptors, preventing the mosquito from smelling and biting humans.

www.cell.com/current-biology/abstract/S0960-9822%2807%2901780-0

5. Alpha-bungarotoxin

Alpha-bungarotoxin is a component in the venom of a type of snake called the Taiwanese banded krait (*Bungarus multicinctus*). Alpha-bungarotoxin binds irreversibly to the receptor of the neurotransmitter acetylcholine in the victim, causing paralysis, respiratory failure and death.

Q. Why is this protein relevant?

Understanding how the alpha-bungarotoxin protein works helps in the development of antivenoms that can save many lives. It can also have other uses in the medical field. The blocking effect of alpha-bungarotoxin on acetylcholine receptors in the brain is being investigated as a way of altering nervous activity in schizophrenia and epilepsy.

6. Histone H2B

Histone proteins play an important role in the packaging of DNA in chromosomes within cells. DNA wraps around histone proteins to form coils, which significantly reduces the space required in the nucleus to store DNA. If you imagine the DNA-histone complex as a reel of cotton, the histone protein is the reel around which the DNA 'cotton' is wrapped.

Q. Why is this protein relevant to us?

Histone proteins are essential to life. When histone proteins like H2B are removed from yeast cells they die. Not only do histone proteins give structure to our chromosomes but they have an important role in controlling the expression of genes. Modifications to histone proteins affect how tightly or loosely wrapped the DNA is and consequently whether genes are expressed or not. Heterochromatin is a tightly packed form of DNA which means the transcription machinery cannot get to the DNA and the genes cannot be expressed. They are silenced.

Modifications to histones are associated with the development of some cancers, including leukaemias, breast cancers and ovarian cancers.

www.ncbi.nlm.nih.gov/pmc/articles/PMC1949974/pdf/11055583.pdf

7. Caspase 1

Caspase 1 is an enzyme that destroys proteins in cells leading to programmed cell death (apoptosis).

Q. Why is this protein relevant?

For us to develop correctly some of our cells must die. One example is during the formation of our fingers and toes where the cells that make up the tissues between the digits die. The image on slide 23 of the PowerPoint presentation shows a light microscope image of the developing foot of a mouse embryo showing areas of dying cells between the digits. The limb starts off as a paddle

shape, but programmed cell death (apoptosis) occurs in the areas between where the digits will form. The dead cells are removed by scavenging cells called macrophages (stained green).

8. Leghaemoglobin

Leghaemoglobin is found in leguminous plants, e.g. peas and beans whose roots have been infected by a bacterium called *Rhizobia*. The symbiotic relationship between the plant and the bacteria enables the plant to fix atmospheric nitrogen into the soil.

The protein leghaemoglobin binds to oxygen, playing an important role in controlling the levels of oxygen in the root nodules of leguminous plants. Too much oxygen can prevent nitrogen fixation taking place as it interferes with another protein called nitrogenase, which is responsible for fixing nitrogen into the soil. It can also kill the *Rhizobia* bacteria.

Q. Why is this relevant to us?

Nitrogen fixation is one of the most important biochemical processes on earth. It is the principal source of the nitrogen in the soil and without it plants would not be able to grow.

FUNCTION FINDERS BLAST

Teacher's notes

OPTIONAL EXTRA: MODEL THE PROTEIN USING RASMOL

Use of the Rasmol software provides another way of exploring the structure of proteins.

The files names for the 8 proteins featured in the activity are listed below:

Num	Rasmol file	Protein name	Sequence
1	ANP1_ANALU	Antifreeze protein type III	M K S A I L T G L L F V
2	GFP_AEQVI	Green Fluorescent Protein	M S K G E E L F T G V V
3	LUCI_LUCLA	Luciferase / Luciferin 4-monooxygenase	E N M E N D E N I V Y G
4	OR1_ANOGA	Odorant receptor protein OR1	G W A L R I M F L H L Y
5	NXL1A_BUNMU	Alpha-bungarotoxin	P G E N L C Y R K M W C
6	H2B_PEA	Histone H2B	P R E I Q T A V R L L
7	CASP1_HUMAN	Caspase 1	E K R K L F I R S M G E
8	LGB2_VICFA	Leghaemoglobin	M E F T L R Q E A L V N

Viewing the protein in Rasmol

If you are using Rasmol as a demonstration tool only, you can complete all of these steps prior to the session so you can click on your protein of choice:

1. Open Rasmol
2. Click 'File', and then 'Open...' and select the file you want (they have been labelled as the protein name for ease of identification)

Rasmol controls

- Zoom in and out:
Hold Shift & Alt & the left mouse button. Slide the mouse forwards and back.
- To centre the structure in the window:
Hold down the right mouse button and move the mouse.
- To rotate the structure:
Hold down the left mouse button and move the mouse.