# Function Finders

Protein Profile 1

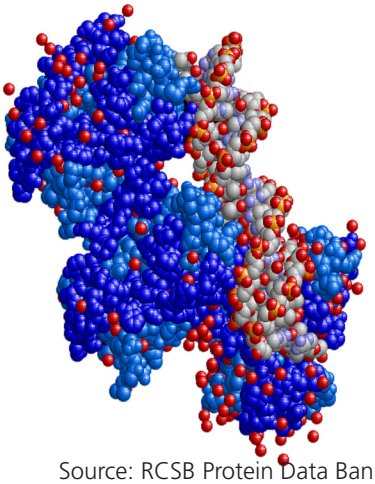
## P53: Protecting the genome

Protein name: P53

Amino acid sequence: MEEPQSDPSVEP

Organism: HUMAN

**P53 is known as the 'guardian of the genome', helping to protect us from cancer.**



Source: RCSB Protein Data Bank

## Protein Facts

- Each P53 molecule has three regions: one binds to other P53 molecules, one switches genes on and one binds to DNA.
- P53 has several anti-cancer properties: it can activate DNA repair proteins when DNA has been damaged; it can suspend cell division, allowing time for damaged DNA to be repaired; it can initiate cell suicide (apoptosis) if cells are too badly damaged to be repaired.
- Damage to the *P53* gene can be caused by UV radiation, certain chemicals, some drugs and viruses – all of which may lead to cancer.
- People who inherit only one functional copy of the gene (instead of the usual two) are more likely to develop tumours early in adulthood.
- P53 is constantly produced but in healthy cells it is prevented from initiating DNA repair or cell death by being constantly destroyed. Only when the cell is damaged does it become active.

## How does this protein relate to me?

P53 is involved in activating the "tanning machinery" in skin cells when they are exposed to sunlight. This process leads to the production of the brown pigment melanin, which absorbs UV light and helps protect the skin cells from damage.

© N. Durrell Mckenna, Wellcome images

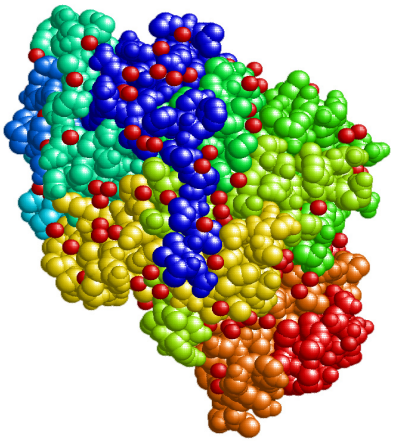
To view the protein in 3D visit: www.ebi.ac.uk/thornton-srv/databases/pdbsum/ and type **1tup** in the code box.
Find your protein sequence at www.uniprot.org and enter **P04637** in the search box

For more information on genes and proteins visit www.yourgenome.org.

# Function Finders

## Protein Profile 2

## Luciferase: Lighting the way



Source: RCSB Protein Data Bank

Protein name: Luciferase

Amino acid sequence: ENMENDENIVYG

Organism: Firefly

**Luciferase is an enzyme used by fireflies to create their characteristic light.**

### Protein Facts

- Luciferase is one of a number of proteins that are used by organisms to generate light in a process called bioluminescence.
- Luciferase uses chemical energy to speed up the reaction between luciferin and oxygen to generate light.
- The adult firefly beetle uses bioluminescence to attract a mate. Other organisms also use bioluminescence including deep sea fish such as the Anglerfish which uses it to attract prey rather than a mate.
- The word luciferase is based on the word lucifer meaning light-bearer.
- By changing just one amino acid in the protein sequence of luciferase, the light emitted changes from yellow to red.

### How does this protein relate to me?

Luciferase can be used to follow the production of other proteins in a cell or the body of an organism. This has been used to follow the development of cancerous cells and visualize how anti-cancer drugs slow down cancerous cell division.

© antfarmer, Wikipedia commons

To view the protein in 3D visit: www.ebi.ac.uk/thornton-srv/databases/pdbsum/ and type **2d1q** in the code box.
Find your protein sequence at www.uniprot.org and enter **Q01158** in the search box.

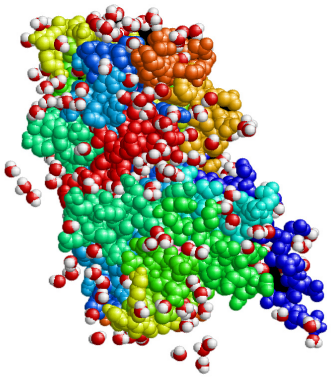
For more information on genes and proteins visit www.yourgenome.org.

# Function Finders

Protein Profile 3

## Odour receptors: Picking up the scent

| | |
|---|---|
| Protein name: | Odorant receptor protein OR1 |
| Amino acid sequence: | GWALRIMFLHLY |
| Organism: | Mosquito |



Source: RCSB Protein Data Bank

**Odorant receptor proteins allow organisms to detect specific chemicals in their environment.**

## Protein Facts

- Odorant receptor protein OR1 recognises a component of human sweat.
- This protein is present in the antennae of mosquitoes that carry malaria (*Anopheles gambiae* and *Anopheles stephensi*).
- Only female mosquitoes feed on human blood and transmit malaria (males don't!) and therefore only female mosquitoes have this protein.
- Production of this protein in the mosquito decreases after it has a blood meal.
- Knowledge of this protein and the compound it recognises could be used to battle malaria.

## How does this protein relate to me?

The mosquito OR1 protein binds to components of human sweat. Using this protein, a mosquito can identify a blood meal.

Mosquitoes such as *Anopheles gambiae,* are vectors of the parasite *Plasmodium spp* that causes malaria. When travelling to countries where the disease is endemic (naturally present) it is important to take anti-malarial drugs and to sleep under a mosquito net.

© Centers for disease control and prevention, Wikipedia commons

To view the protein in 3D visit: www.ebi.ac.uk/thornton-srv/databases/pdbsum/ and type **1obp** in the code box.
Find your protein sequence at www.uniprot.org and enter **Q8WTE7** in the search box.

For more information on genes and proteins visit www.yourgenome.org.

# Function Finders

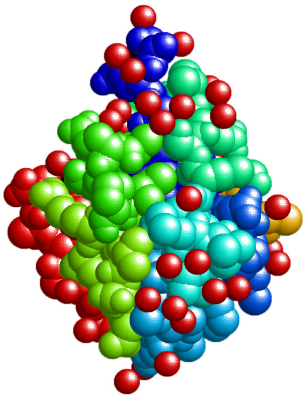
Protein Profile 4

## Fish antifreeze protein: A cool customer

Protein name: Antifreeze protein type III

Amino acid sequence: MKSAILTGLLFV

Organism: Atlantic wolffish



Source: RCSB Protein Data Bank

**Antifreeze proteins prevent organisms, like the Atlantic wolffish, from freezing in the icy seawater they live in.**

### Protein Facts

- Antifreeze proteins are also known as ice-structuring proteins because they bind to ice crystals and prevent them from growing.
- Antifreeze proteins are produced by the liver and are released into the blood. These proteins are also produced in the scales, fins and gills of fish and act as the first line of defence against freezing.
- The 3D structures of the many different proteins that act as antifreeze agents show very little similarity to each other.
- The evolution of these proteins is thought to have happened more than once: 10-30 million years ago in Antarctica and again 1-2 million years ago due to glaciation in the northern hemisphere.
- Other organisms that also use antifreeze proteins include bacteria, fungi, plants and insects.

### How does this protein relate to me?

Inspired by the way antifreeze proteins help prevent fish like the Atlantic wolffish (shown right) from freezing, some manufacturers are investigating the use of antifreeze proteins in ice cream and frozen yoghurt. The presence of the proteins prevents significant ice crystalisation and therefore creates a smoother ice cream or frozen yoghurt.

© DOAA, Wikipedia commons

To view the protein in 3D visit: www.ebi.ac.uk/thornton-srv/databases/pdbsum/ and type **1gzi** in the code box.
Find your protein sequence at www.uniprot.org and enter **P12416** in the search box.

For more information on genes and proteins visit www.yourgenome.org.

# Function Finders

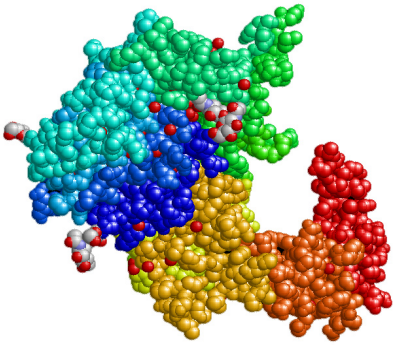
Protein Profile 5

## Cell switches: To divide or not divide?

Protein name: HER2

Amino acid sequence: MELAALCRWGLL

Organism: Human



Source: RCSB Protein Data Bank

**HER2 acts as a "molecular switch", instructing cells when to divide and when not to divide.**

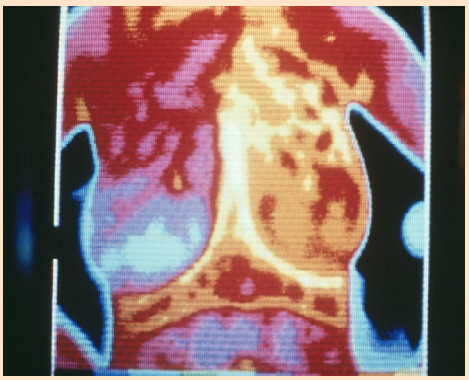
## Protein Facts

- HER2 stands for Human Epidermal growth factor Receptor 2. It helps cells to develop and divide. It is also known as ERBB2.
- The protein has relatives in worms and flies that do exactly the same thing.
- Approximately 15-20% of breast cancer patients have an amplification of the *HER2* gene or over expression of its protein product.
- Herceptin is a drug used to treat some types of breast cancer, it attaches to HER2 and stops it from working. Blocking HER2 with herceptin stops cancer cells from dividing and might also stop tumours from growing a good blood supply.

## How does this protein relate to me?

This protein is one of the 20-25,000 proteins in the human genome. In some cancers HER2 is produced much more than usual, helping cells to grow and divide faster and develop into cancerous tumours. Research has shown that women with *HER2*-positive breast cancer have a more aggressive disease, greater likelihood of recurrence and decreased survival compared to women with *HER2*-negative breast cancer.

Tests for *HER2* can help determine which course of treatment is best to treat the cancer.





To view the protein in 3D visit: www.ebi.ac.uk/thornton-srv/databases/pdbsum/ and type **1n8y** in the code box.
Find your protein sequence at www.uniprot.org and enter **P04626** in the search box.

For more information on genes and proteins visit www.yourgenome.org.

# Function Finders

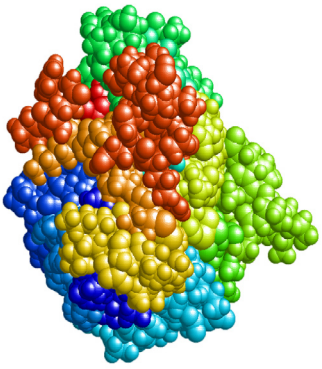
Protein Profile 6

## Snake venom toxin: Small but deadly

Protein name: Alpha-bungarotoxin

Amino acid sequence: PGENLCYRKMWC

Organism: Many banded krait (snake)



Source: RCSB Protein Data Bank

**Snake venoms contain a multitude of biologically active toxins that work together to overpower and kill prey, one of which is alpha-bungarotoxin.**

## Protein Facts

- Alpha-bungarotoxin is a deadly venom produced by a snake called the 'Many-banded Krait' from South-East Asia.
- The snake venom kills the prey by paralysing their muscles.
- Other venomous snakes, including cobras and sea snakes, have similair proteins that work in the same way.
- 30 milligrams of venom, less than a tenth of a teaspoon, is enough to kill a 100 kg man in 24 hours.
- Anti-venom is a treatment for snake bites; it attaches to the protein and stops it from working which can save the lives of snake bite victims.

## How does this protein relate to me?

Alpha-bungarotoxin, from the Many-banded krait, blocks signals between nerve and muscle cells and is used by scientists to study these signals. Along with other snake venoms, alpha-bungarotoxin has been studied as potential treatments for a range of conditions including arthritis, thrombosis and even some types of cancer. Research is still ongoing in this field.

© B. G. Fry, Australian Venom Research Unit, Melbourne, Australia

To view the protein in 3D visit: www.ebi.ac.uk/thornton-srv/databases/pdbsum/ and type **1abt** in the code box.
Find your protein sequence at www.uniprot.org and enter **P60615** in the search box.
For more information on genes and proteins visit www.yourgenome.org.

# Function Finders

Protein Profile 7

## Plant haemoglobin

Protein name: Leghaemoglobin

Amino acid sequence: MEFTLRQEALVN

Organism: Broad Bean

Source: RCSB Protein Data Bank

**Haemoglobin binds oxygen in animal blood, but few people know that plants can have a similar protein.**

## Protein Facts

- Leghaemoglobin is found in leguminous plants, e.g. peas and beans, whose roots have been infected by a bacterium called *Rhizobia.*
- The protein is formed from two parts, one provided by the plant and one by the bacterium.
- Like haemoglobin (found in red blood cells), leghaemoglobin binds oxygen and is red.
- Nitrogenase is an enzyme found in the root nodules of leguminous plants which plays an important role in "fixing" nitrogen into the soil. If exposed to too much oxygen this enzyme cannot function properly. Leghaemoglobin with its ability to bind oxygen buffers the concentration of oxygen in the root nodules to ensure that nitrogenase functions and the bacteria can also survive.

## How does this protein relate to me?

No animal has the ability to convert nitrogen from the atmosphere into compounds that can be used within the cell. We depend upon the symbiotic relationship between certain bacteria and plants to fix atmospheric nitrogen into the soil.

© Tuinboon, Wikipedia commons

To view the protein in 3D visit: www.ebi.ac.uk/thornton-srv/databases/pdbsum/ and type **2gdm** in the code box.
Find your protein sequence at www.uniprot.org and enter **P93848** in the search box.

For more information on genes and proteins visit www.yourgenome.org.

# Function Finders

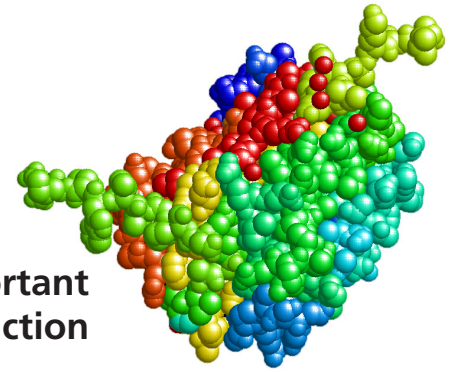
Protein Profile 8

## Caspases: Cellular self destruction

Protein name: Caspase 1

Amino acid sequence: EKRKLFIRSMGE

Organism: Human



Source: RCSB Protein Data Bank

**Caspases are enzymes that destroy or activate important proteins in the cell. The result is the deliberate destruction of targeted cells; a process called apoptosis.**

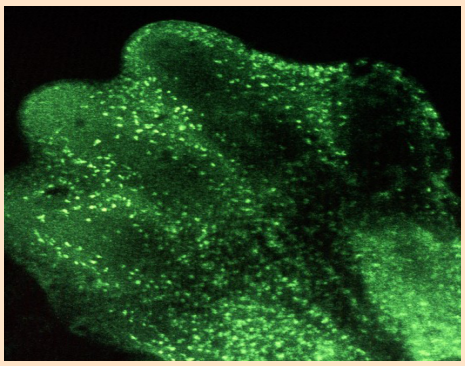
## Protein Facts

- Caspases cleave proteins at certain points to either destroy them or make them active.
- Some of our cells have to die for us to develop properly. For example, half of our nerve cells die as the nervous system develops – this leaves the correct neural wiring behind.
- Cell suicide protects against cells that may pose a threat to the rest of the organism, e.g. cells damaged by radiation or infected by viruses.
- Other organisms also use killer caspases to cause specific cells to die, e.g. when a tadpole loses its tail, the tail cells die due to a controlled, targetted process.
- Where cell death is blocked, cells can undergo uncontrolled division and result in certain cancers, e.g. types of leukaemia and melanoma (skin cancer).

## How does this protein relate to me?

Caspases are important for your development, e.g. the formation of your fingers and toes. As the foetus grows, caspases mediate the death of cells to form the spaces between your digits (shown in the picture, right).

Some cancer treatments also induce apoptosis in specific cancer cells to prevent their growth.



© Paul Martin, Wellcome Images

To view the protein in 3D visit: www.ebi.ac.uk/thornton-srv/databases/pdbsum/ and type **1ice** in the code box.
Find your protein sequence at www.uniprot.org and enter **P29466** in the search box.

For more information on genes and proteins visit www.yourgenome.org.

# Function Finders

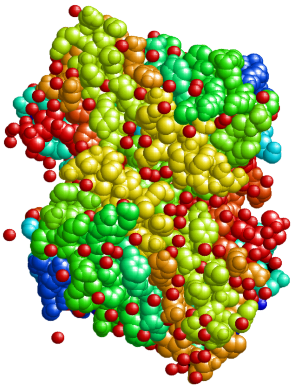
Protein Profile 9

## Fluorescent protein: Glowing green jellyfish

Protein name: Green fluorescent protein (GFP)

Amino acid sequence: MSKGEELFTGVV

Organism: Jellyfish *(Aequorea victoria)*



Source: RCSB Protein Data Bank

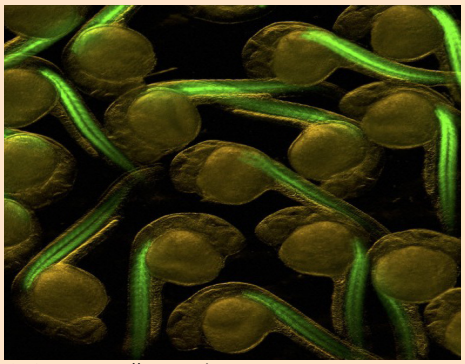
**GFP is used by certain organisms to communicate with one another. Genetic manipulation makes use of it to follow proteins in the cell.**

## Protein Facts

- Green fluorescent protein fluoresces green when exposed to blue light. The jellyfish that produces GFP is usually colourless, but flashes green in response to external factors.
- GFP is used by biologists in research. It can be introduced into an organism to watch its development. The picture below shows transgenic zebrafish embryos expressing green fluorescent protein in the muscle precursor cells (myotomes). When coupled with other genes *GFP* can be used as a marker for successful genetic manipulation – the modified cells glow green to indicate that the gene has been successfully inserted.
- The *GFP* gene has been deliberately mutated to fluoresce at other wavelengths producing more than eight colours, including cyan and yellow.
- The artist Eduardo Kac made a 'transgenic artwork' in 2000, a rabbit that was genetically manipulated with the jellyfish *GFP* gene so that it glowed green under blue light.

## How does this protein relate to me?

*GFP* can be used to understand human development and proteins. By attaching *GFP* to specific proteins in a cell, researchers can track their production and their path in our bodies. For this reason *GFP* is known as a 'reporter' gene as it tracks (or reports) the location of a second protein.



© S.Roy, Wellcome images

To view the protein in 3D visit: www.ebi.ac.uk/thornton-srv/databases/pdbsum/ and type **1gfl** in the code box.
Find your protein sequence at www.uniprot.org and enter **P42212** in the search box.

For more information on genes and proteins visit www.yourgenome.org.

# Function Finders

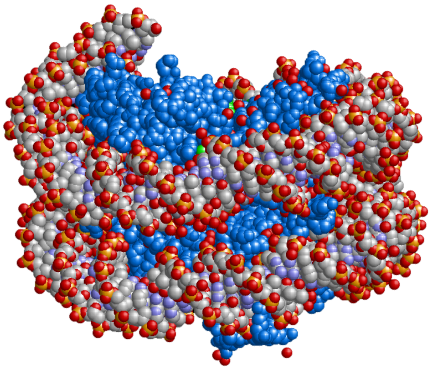
Protein Profile 10

## Histones: It's a wrap

Protein name: Histone H2B

Amino acid sequence: PREIQTAVRLLL

Organism: Pea



**Histones are used to wind DNA up so that it can be condensed and fit into the nucleus of each cell.**

Source: RCSB Protein Data Bank

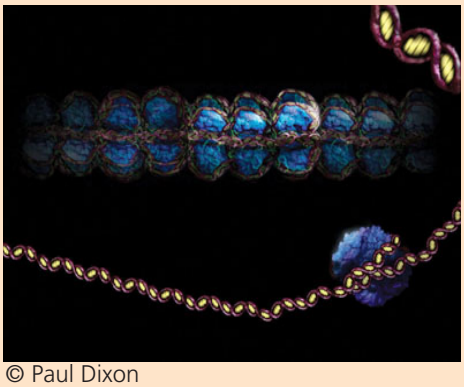
## Protein Facts

- Histones play an important role in all cells by providing a robust structure around which two metre lengths of DNA can wrap to create gigantic coils which form the basis of chromosomes. This coiling mechanism significantly reduces the space required in the nucleus to store unwrapped DNA.
- DNA is wrapped around four clumps of histone proteins named H2A, H2B, H3 and H4. The histones act like a cotton reel around which the DNA thread is wrapped.
- Histones don't only work to give a structure to DNA, they also have a role to play in gene expression: the activity of genes. Histone modifications can lead to the formation of heterochromatin; a tightly packed form of DNA which has limited transcription properties which results in gene silencing, or "switching off".

## How does this protein relate to me?

Histones are one of the most-conserved proteins in our bodies and are essential to life. If researchers remove histone H2B from yeast cells, the yeast dies.

Our cells modify histones by adding or removing chemical groups: these changes can be associated with diseases such as leukaemia, Hodgkin's lymphoma and Huntingdons disease.



© Paul Dixon

To view the protein in 3D visit: www.ebi.ac.uk/thornton-srv/databases/pdbsum/ and type **2cv5** in the code box.
Find your protein sequence at www.uniprot.org and enter **Q99285** in the search box.

For more information on genes and proteins visit www.yourgenome.org.

# Function Finders

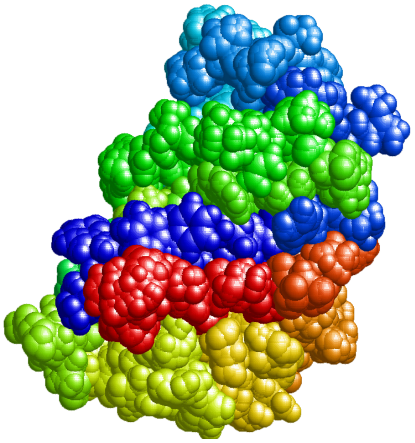
Protein Profile 11

## Mucins: Cell slime

Protein name: Mucin-1

Amino acid sequence: PGGEKETSATQR

Organism: Human



Source: RCSB Protein Data Bank

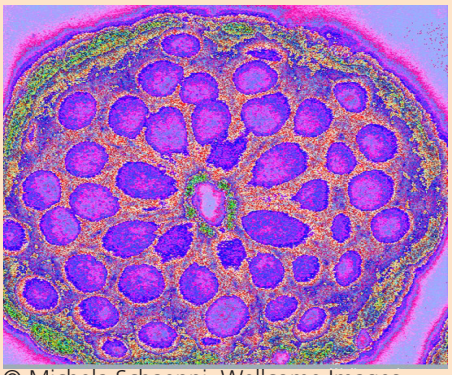
**Mucins are large proteins that form a protective film on the surface of certain cells in our bodies including the intestines and nasal passages.**

## Protein Facts

- A film of mucins can protect cells from micro-organisms and particles that may damage the cell (see image below of intestinal mucin-producing cells).
- To date 19 human mucin genes have been identified.
- There are two classes of human mucin proteins – membrane-bound and secreted. Mucin-1 is classed as a membrane-bound mucin.
- Mucin-1 molecules are secreted from the cell and bind to each other using regions at each end of the protein. This creates an uninterrupted protective layer on the surface of the cell.
- The central region of each mucin-1 protein is covered in sugar molecules which protects it from digestion by enzymes.

## How does this protein relate to me?

Mucins are involved with certain cancers. High levels of mucin-1 cause a reduction in the amount of a protein (named P53) involved in slowing cell division. As a result, cells divide more rapidly which can lead to certain types of cancer, e.g. breast and ovarian carcinoma.



© Michela Schaeppi, Wellcome Images

To view the protein in 3D visit: www.ebi.ac.uk/thornton-srv/databases/pdbsum/ and type **2acm** in the code box.
Find your protein sequence at www.uniprot.org and enter **P15941** in the search box.

For more information on genes and proteins visit www.yourgenome.org.

# Function Finders

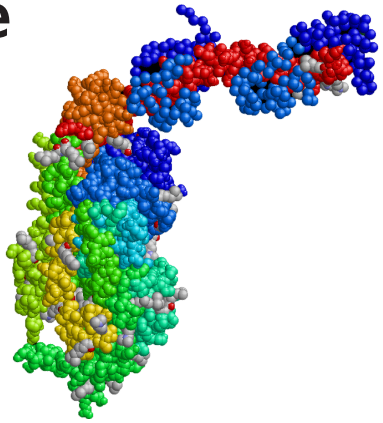
Protein Profile 12

## Myosin: Making muscles move

Protein name: Myosin 1

Amino acid sequence: MSSDSEMAIFGE

Organism: Human



Source: RCSB Protein Data Bank

**Myosin is a major protein found in muscle fibres. Without it you would not be able to move.**

## Protein Facts

- Myosin is attached to the end of the muscle fibre and makes it contract by binding to and pulling on a scaffold of another protein called actin. By 'walking' along this scaffold, the ends of the muscle fibres are brought closer together, thereby shortening the length of the muscle.
- It takes about two trillion myosin molecules working together to provide enough strength to hold an apple.
- Myosin is also used to transport other molecules around the cell. Myosin coats bags of different molecules and pulls them along an interconnecting network of 'paths' in the cell.
- Different forms of myosin are found in organisms from yeast, to sunflowers, to hamsters.
- Chimpanzee myosin differs to human myosin by only two genetic letters, or base pairs.

## How does this protein relate to me?

Almost all of your movements involve the action of myosin.

Myosin uses chemical energy in the form of a molecule named ATP. If you are running flat out, the supply of ATP is exhausted rapidly and muscle switches the energy supply to glycogen. This results in the production of lactic acid causing a burning sensation in the muscles.

© Paul Foot, Wikipedia commons

To view the protein in 3D visit: www.ebi.ac.uk/thornton-srv/databases/pdbsum/ and type **2mys** in the code box.
Find your protein sequence at www.uniprot.org and enter **P12882** in the search box.

For more information on genes and proteins visit www.yourgenome.org.