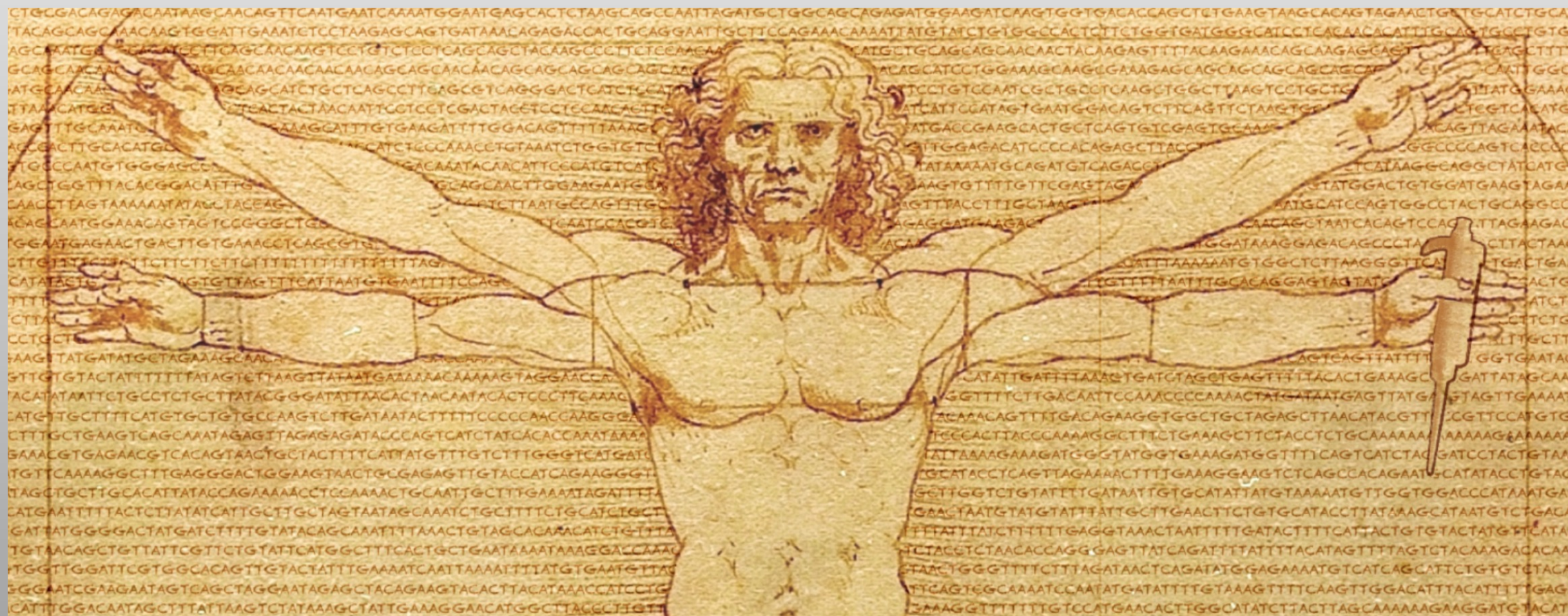


UNLOCKING LIFE'S CODE



Essential Knowledge for Every Human Being



Before using this resource, you can use this space to record current ideas you might have about DNA genes and proteins. You can add comments/questions as you go, either here or on the four pages at the end of this file (you don't *have* to fill up these spaces).



A large rectangular box containing 20 horizontal grey lines, intended for taking notes or recording ideas.

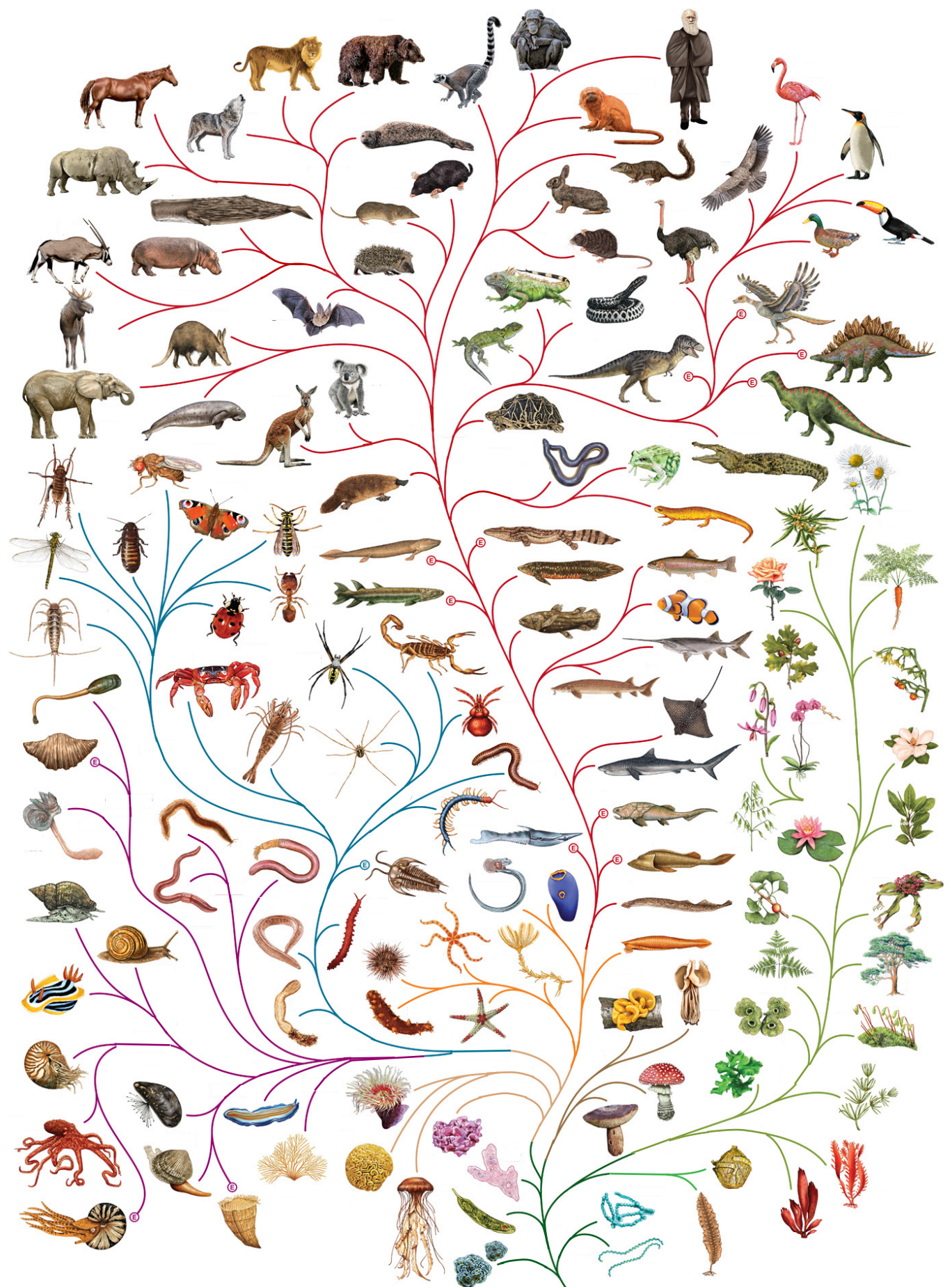


Naruto/David Slater (Wikimedia Commons – Public Domain)

This female Celebes Crested Macaque, who has been named Naruto, is over 90% identical to you! Not only does she enjoy taking the occasional selfie, but her DNA is very similar to that found in humans.

However, even the plants in the picture and the bacteria on Naruto's teeth have something in common with us. This is because all life on earth uses a genetic code made from four molecules that are the building blocks of DNA.

To understand how the DNA code works, we are going to start with a few fun exercises using an emoji code instead.



Origin of DNA, RNA and life, around 3 billion years ago



Aroha



Grin

Aroha and Grin belong to the same emoji family. We will call them the **Purine emoji**, because cats love to purr.

Emoji codon chart

(a codon is a group of 3 symbols)



Important! Look at how the red arrows on the chart point from the large emoji in the centre to the small emoji in the outer ring. This is the order you will read them on the chart.

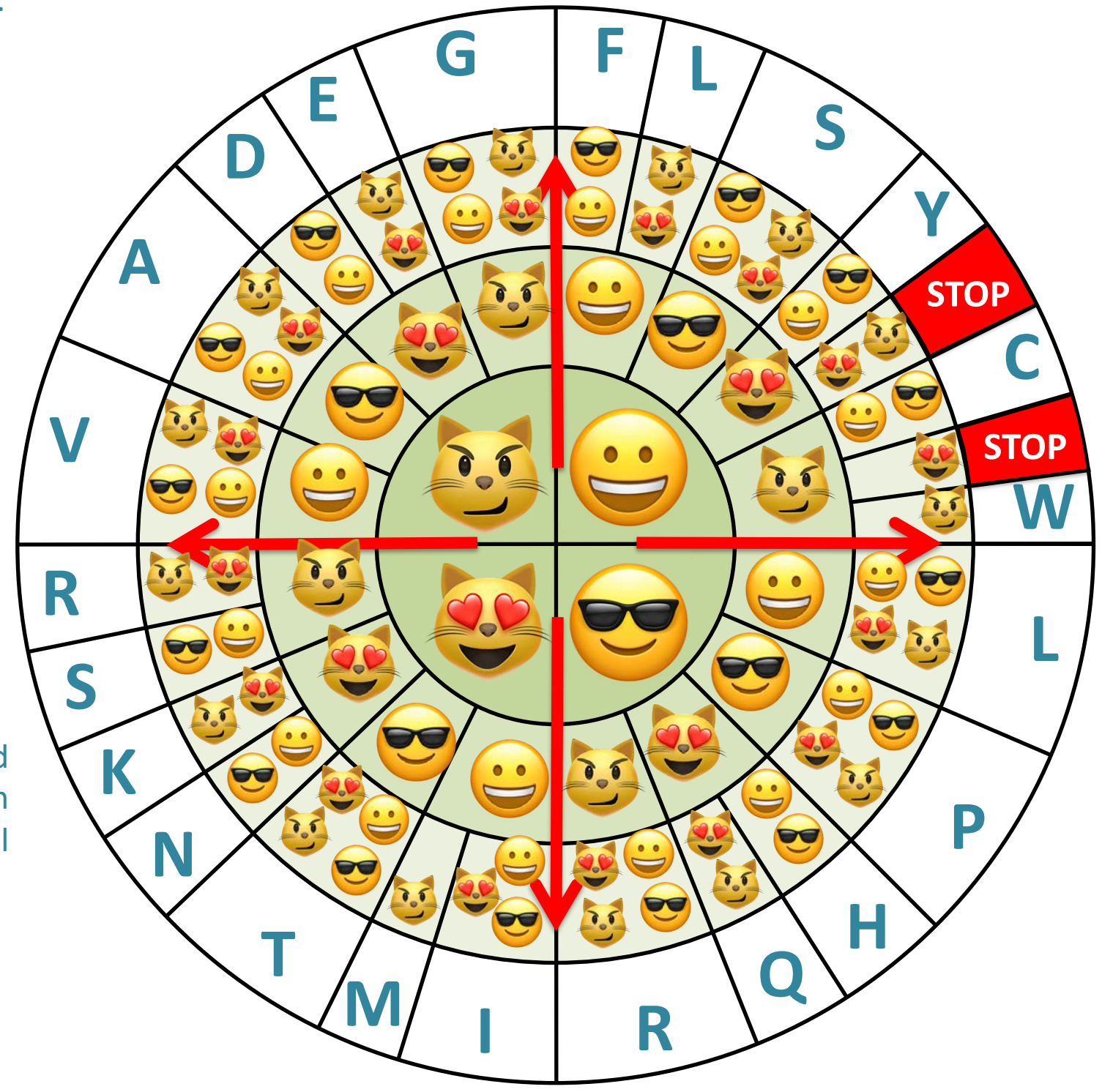


Teeth



Cool

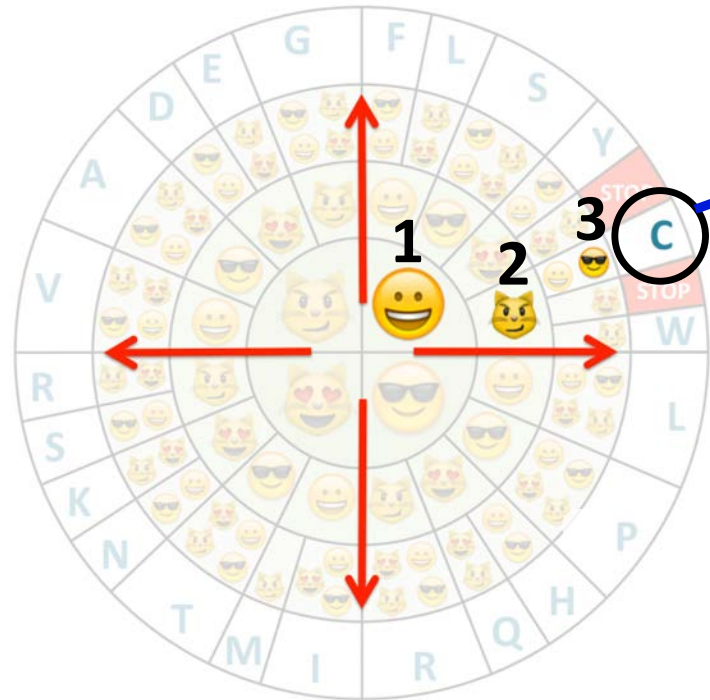
Teeth and Cool belong to another emoji family and have a sunny outlook. Because the ancient Egyptian sun god Ra was worshipped at the **Pyramids**, we will call them the **Pyrimidine emoji**.



Can you use the emoji code to figure out the words?

Group the emoji into sets of three (making codons) and use the circular emoji chart to figure out which letter each group stands for. On the chart, go from the largest emoji to the smallest, following the direction of the red arrows from the centre to the outer ring. Sometimes more than one emoji will be possible in a position. In the example shown, for the letter C, the third emoji could be either Teeth or Cool (in this case it is Teeth).

How to use the emoji codons



Put the letters here

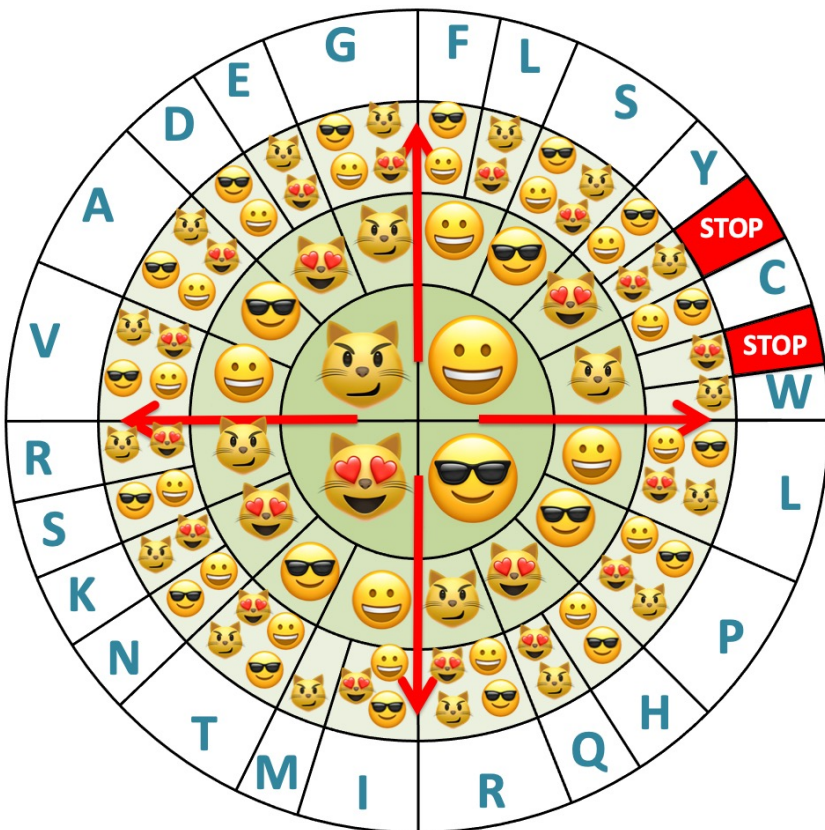
=

Now try to decode the meaning of these emoji

=

=

=



DNA isn't made from emoji, of course

Pyrimidines



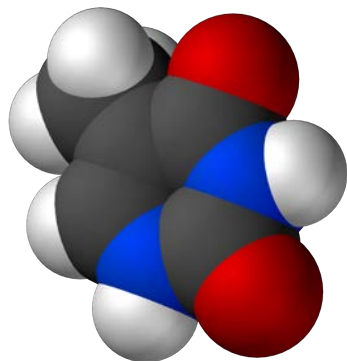
Teeth



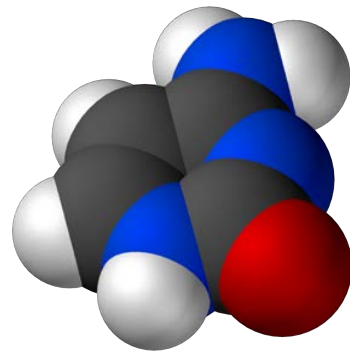
Cool

In Your DNA

Thymine



Cytosine



Purines



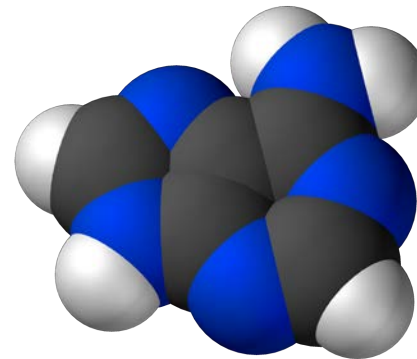
Aroha



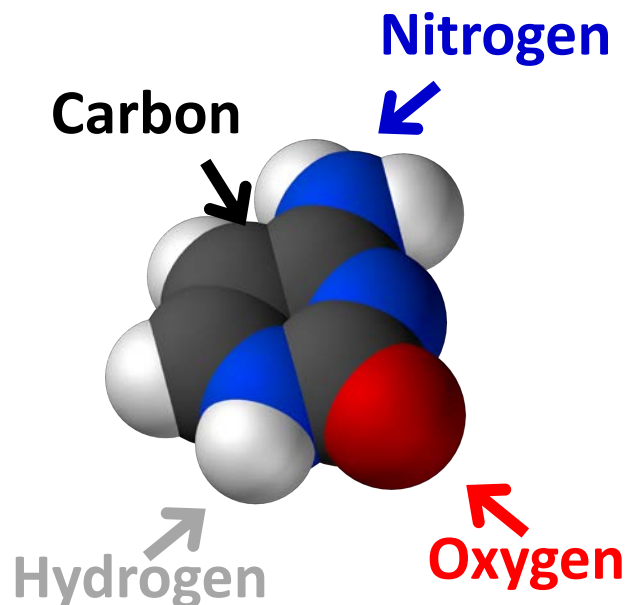
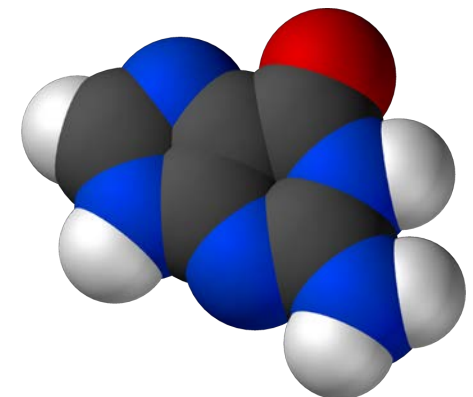
Grin

In Your DNA

Adenine



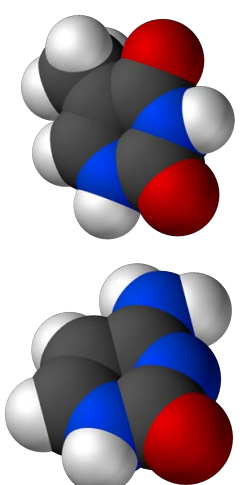
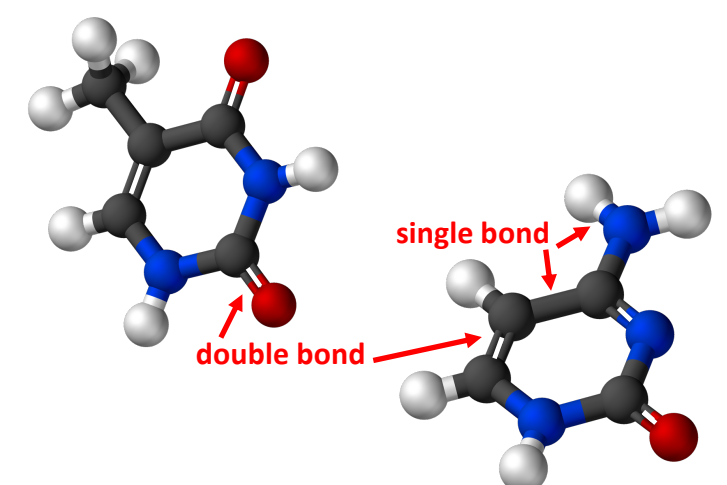
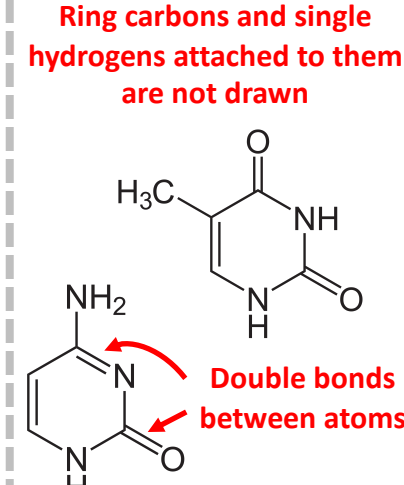
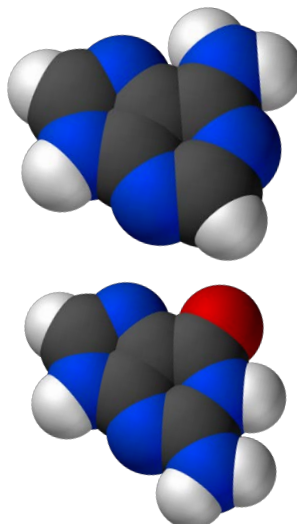
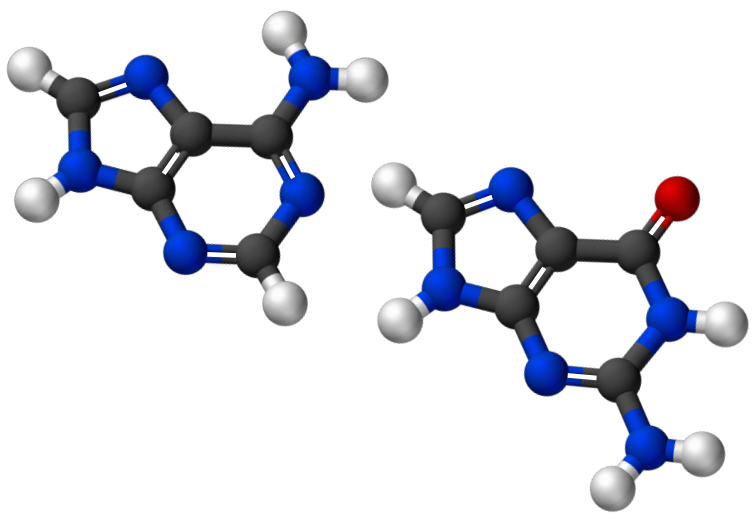
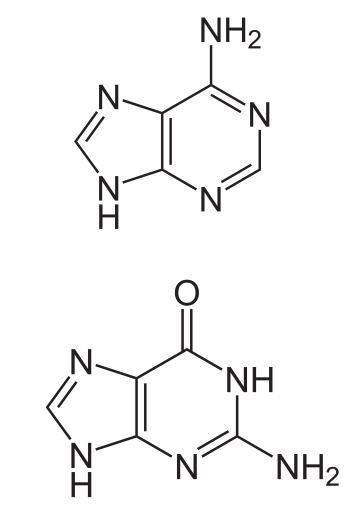
Guanine

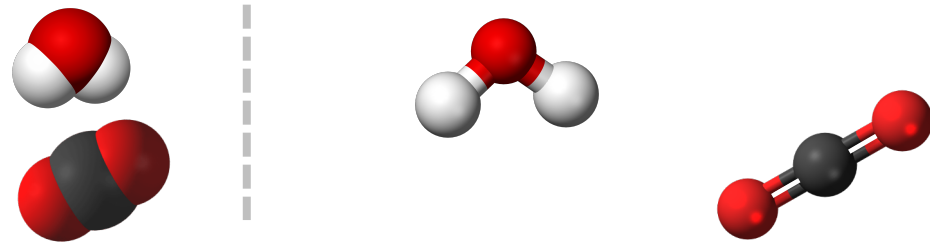


Thymine, Cytosine, Adenine and Guanine are key pieces in the building blocks of your DNA (which stands for Deoxyribonucleic acid). They are made from atoms of Carbon, Oxygen, Hydrogen and Nitrogen. Because they have such an important role in the structure and function of deoxyribonucleic acid, we sometimes refer to Thymine, Cytosine, Adenine and Guanine as 'nucleobases'.

Hint for remembering which atoms get which colour: Carbon is the basis for charcoal and graphite, which are black. Oxygen is needed for fire (combustion) so is red, like a fire truck. Nitrogen in the atmosphere makes the sky blue because of the way it scatters light, so its atom is coloured blue. Hydrogen is a colourless gas, so is shown in white (or gray).

There are different ways to show the same thing

	Space filling model	Ball & stick model	Chemical drawing	Chemical Formula	Common name	Letter
Pyrimidines			<p>Ring carbons and single hydrogens attached to them are not drawn</p> 	<p>Count the atoms on the models to check these</p> <p>C₅H₆N₂O₂</p> <p>C₄H₅N₃O</p>	<p>Thymine</p> <p>Cytosine</p>	<p>T</p> <p>C</p>
	Purines				<p>C₅H₅N₅</p> <p>C₅H₅N₅O</p>	<p>Adenine</p> <p>Guanine</p>



If working on paper, you could put the chemical drawings here:

Formula

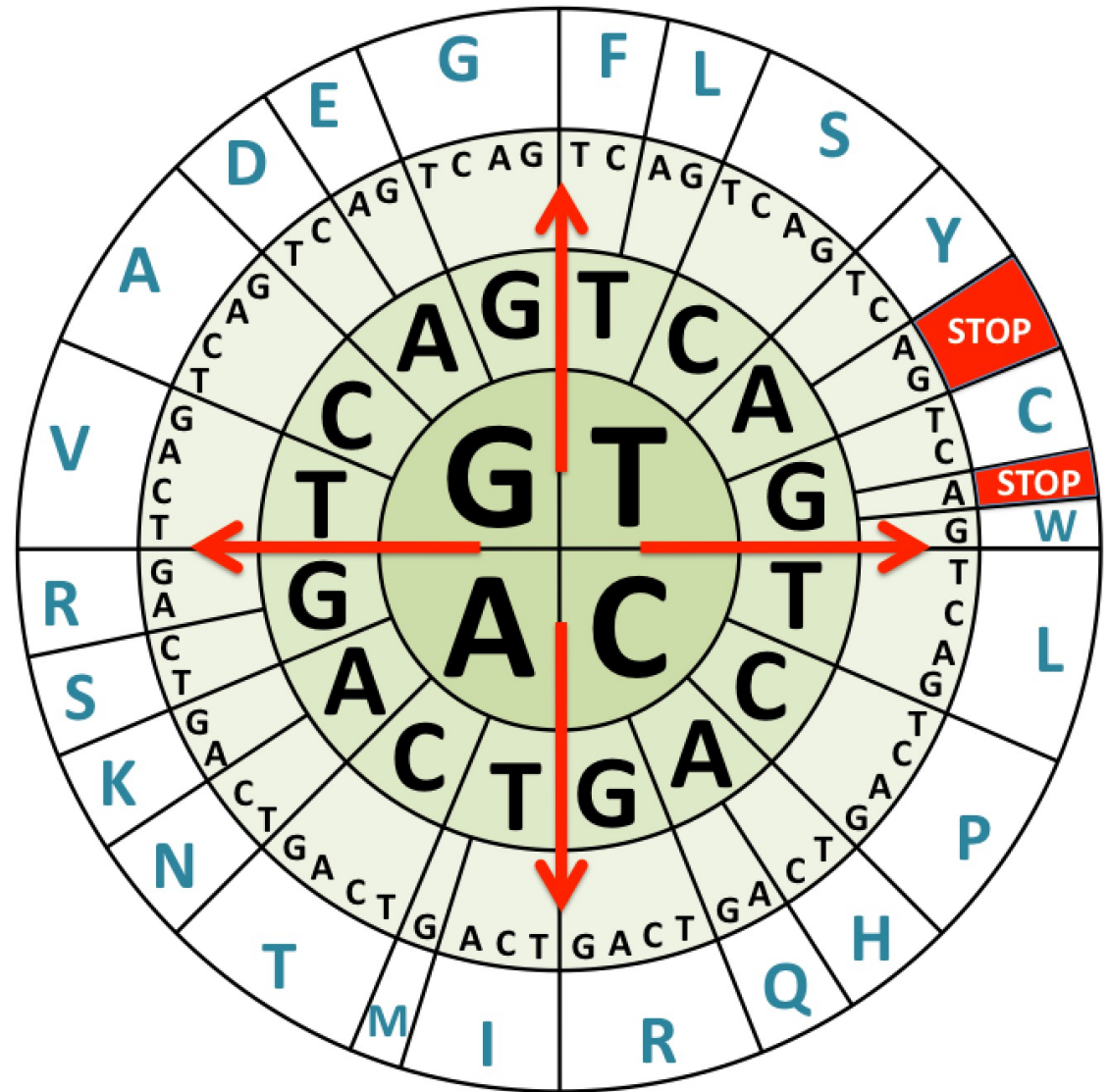
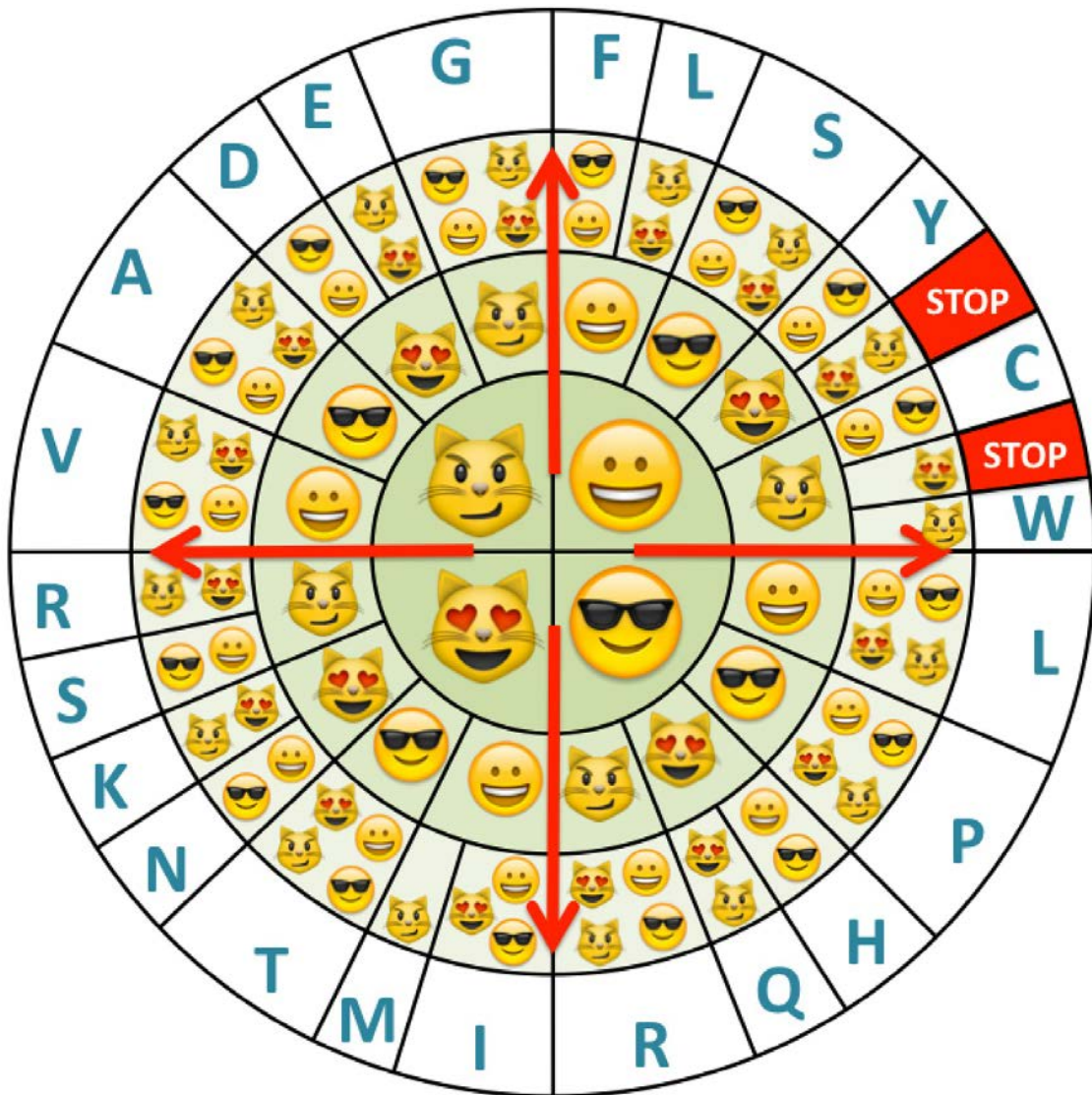
Common name

Hint for writing formulas: Follow the alphabetical order of atom names (Carbon then Hydrogen, etcetera). If there is only one of a particular atom in a molecule, you don't need to write the number 1 next to it.

Can you fill in the empty boxes?

Looking at the ball & stick models, find a pattern in the number of bonds that C, H, N and O atoms make and record your observation:

Replacing our emoji with the letter symbols for DNA nucleobases (T,C,A,G) that scientists use



Can you figure out the secret message?

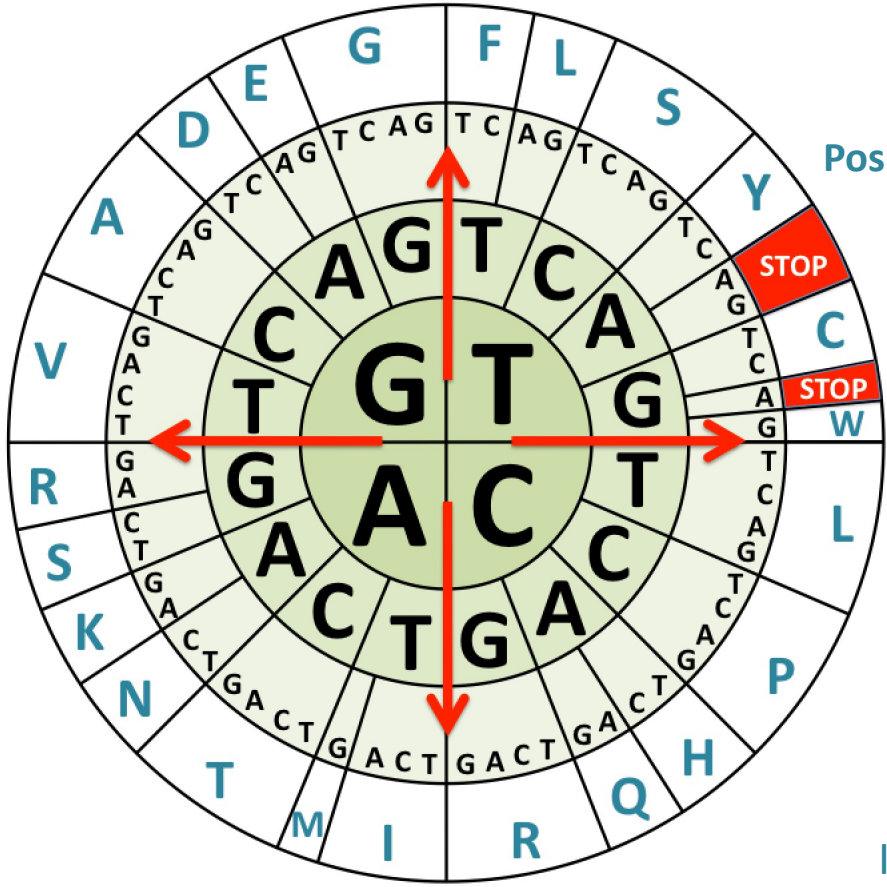
GACGAGTGTATTCCGCATGAACGTATTAACGGC CTGATAAAGGAA GCT TCGTGCATAGAGAATACGATTAGCACT

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--



If you are not familiar with the first word of the message, you might want to look up its meaning and type it down here:

Mutating words



Position 1 Position 8
 ↓ ↓
GTAGCGCTCATTGCAAACACC

Type the 7-letter word it makes here

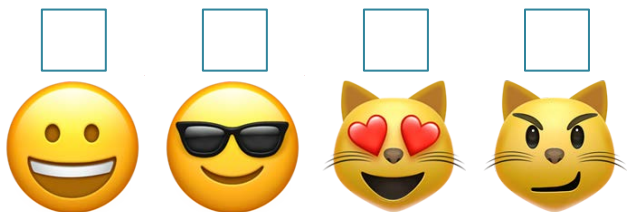
What happens if the nucleobases at Positions 3 and 6 mutate to the other purine and the nucleobase at position 8 mutates from thymine to guanine?

Type the new word it makes here

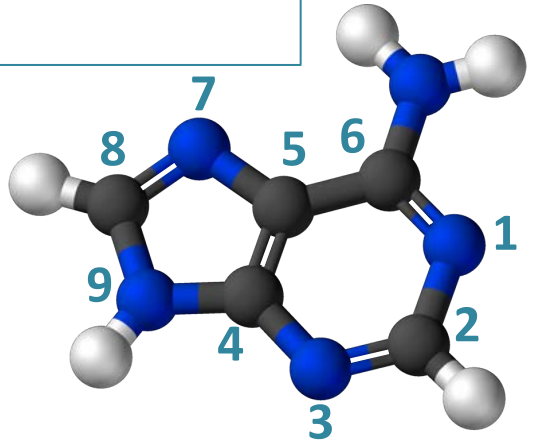
 Type the mutated sequence of bases here

In general, what do you conclude about the relative importance of nucleobases occupying positions 1, 2 or 3 of a codon? Put your answer in the space below:

Click to tick the emoji corresponding to the molecule below and write its scientific name

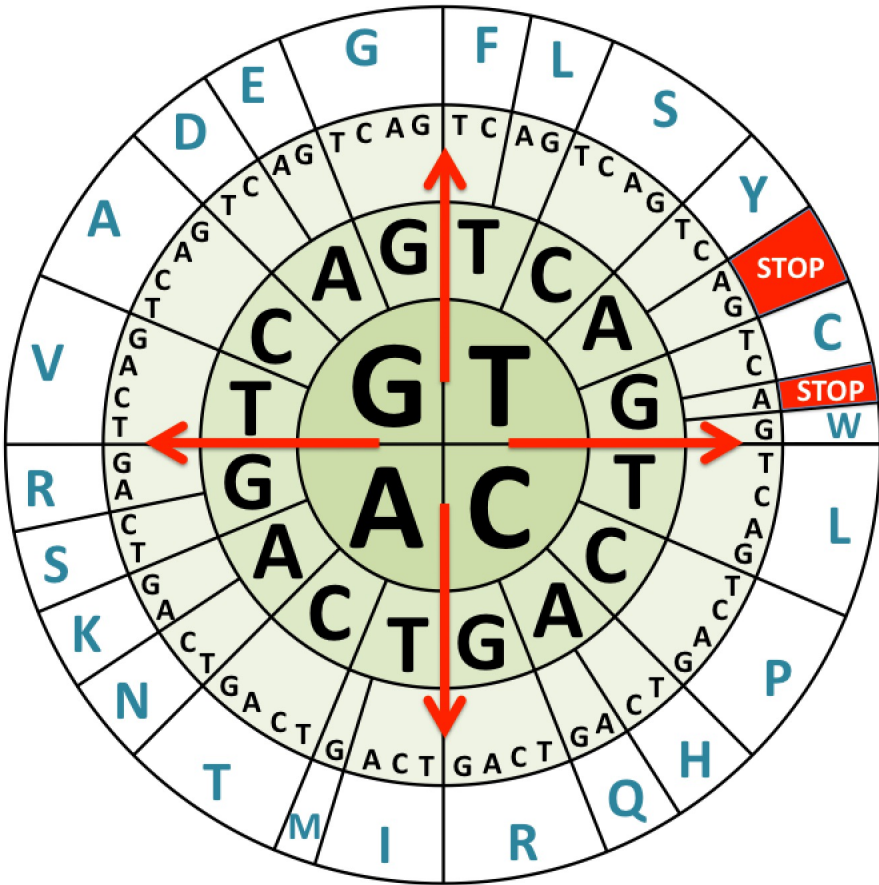


This nucleobase is:



Helpful hints for remembering which bases are purines and which are pyrimidines: Our emoji names give you a good clue, but learning the names of the four bases is also useful. **C**ytosine and **T**hymine are both **p**rimidines and both have the letter **Y**. Therefore, Adenine and Guanine must be the purines. If you want a way to remember whether the purines or pyrimidines are the larger nucleobases, notice that **Adenine** and **Guanine** have **nine** atoms that make up their two rings.

Make secret messages using life's code



Use the codon chart to convert your first or last name into a sequence of bases. If your name has the letters **B, J, O, U, X** or **Z** in it you will need to select another word instead (maybe a family member's, pet's or famous person's name, e.g. Einstein).

Type each letter of your name into a box

--	--	--	--	--	--	--	--	--	--

Then type a set of three bases corresponding to each letter

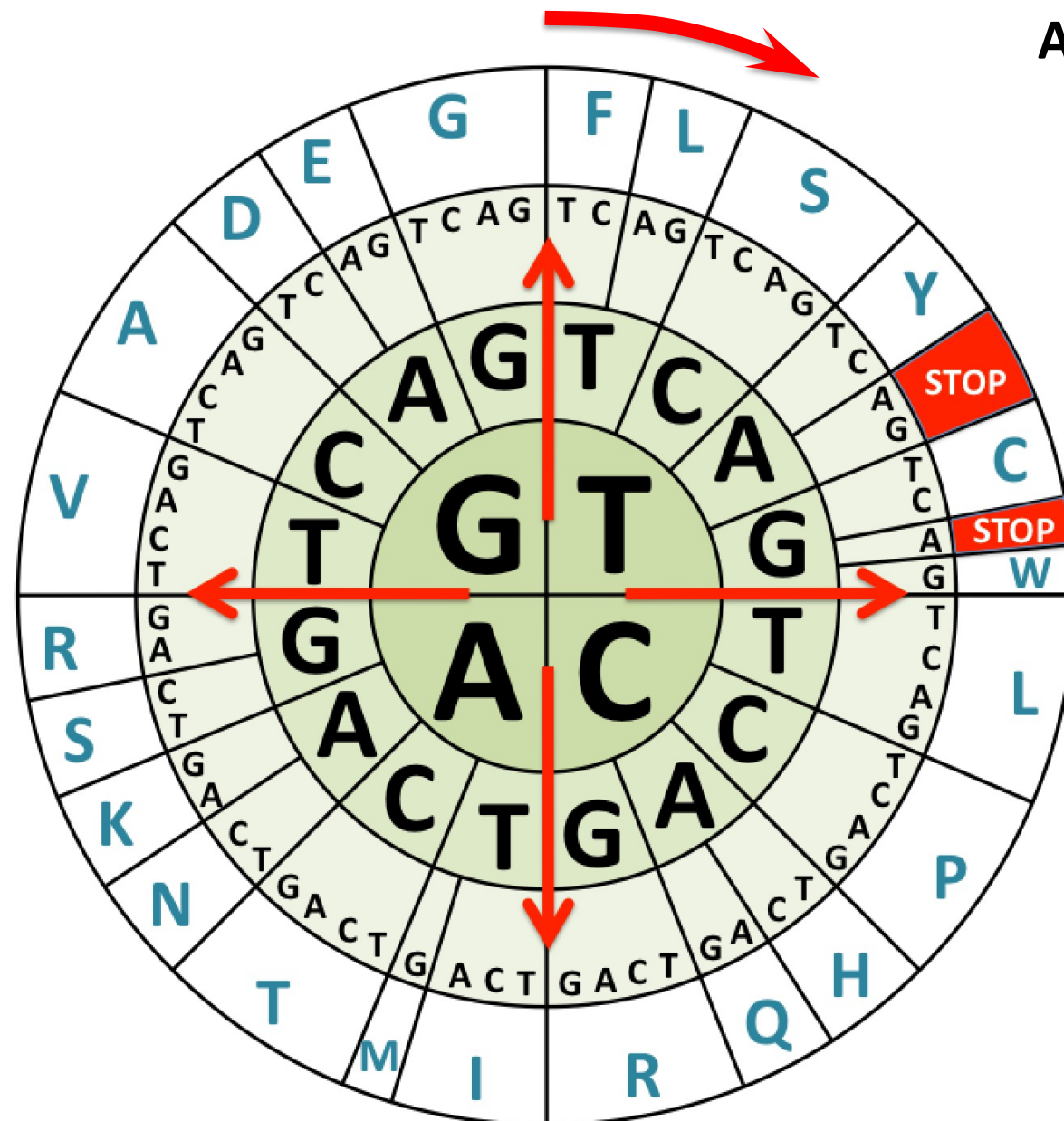
Without using letters B,J,O,U,X and Z, **write a message on a piece of scrap paper** using a maximum of 29 letters. Convert it into the DNA code on the dashed line below, ***without filling in the boxes above it.*** Don't include spaces between words, but **finish with a STOP codon.** The three stop codons are TAG (nicknamed Amber), TAA (nicknamed Ochre) and TGA (nicknamed Opal). Give your secret message to a friend and see if they can decode it.

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

The 'alphabet' letters on the outside ring of the codon chart represent amino acids, the building blocks of proteins

Just as scientists use the letters A, C, G and T when talking about the nucleobases Adenine, Cytosine, Guanine and Thymine, they sometimes also use single letters to represent the amino acids that make up proteins. A protein is a long string of amino acids joined together. The twenty common amino acids are shown here*.



Amino Acids, going clockwise around the chart

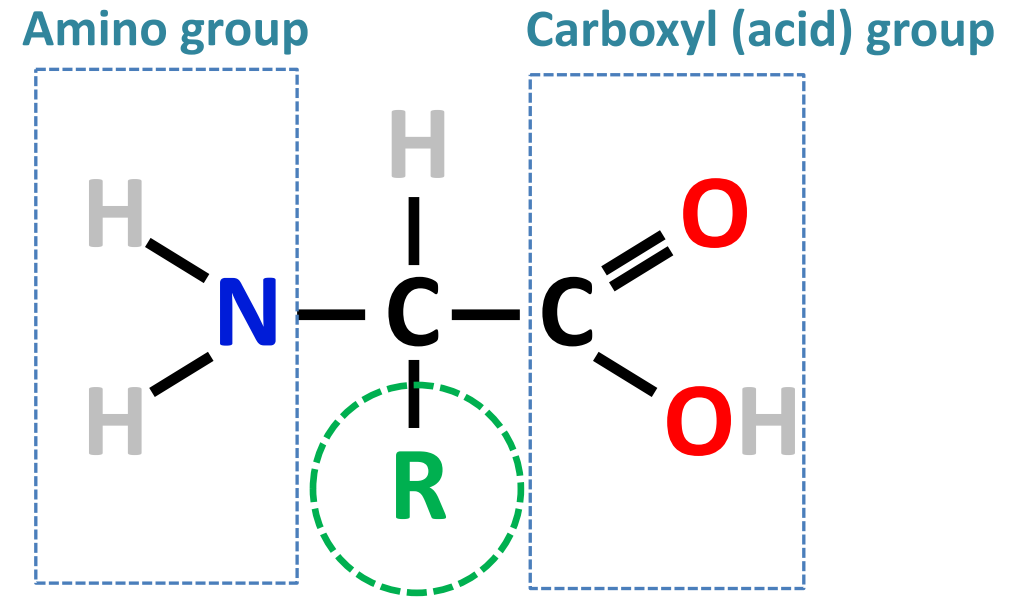
F	Phe	Phenylalanine
L	Leu	Leucine
S	Ser	Serine
Y	Tyr	Tyrosine
C	Cys	Cysteine
W	Trp	Tryptophan
L	Leu	Leucine (<i>again</i>)
P	Pro	Proline
H	His	Histidine
Q	Gln	Glutamine
R	Arg	Arginine
I	Ile	Isoleucine
M	Met	Methionine
T	Thr	Threonine
N	Asn	Asparagine
K	Lys	Lysine
S	Ser	Serine (<i>again</i>)
R	Arg	Arginine (<i>again</i>)
V	Val	Valine
A	Ala	Alanine
D	Asp	Aspartic Acid
E	Glu	Glutamic Acid
G	Gly	Glycine

*There are a few rarer amino acids, such as **selenocysteine (SEC)**, but for simplicity we haven't included them in our codon chart.

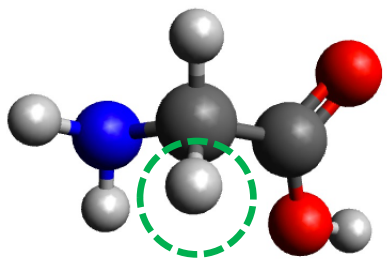
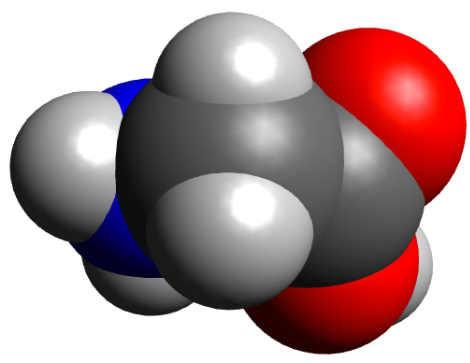
Amino acids have the same backbone, but different side chains

As with the DNA nucleobases, amino acids are made up from C, H, N and O atoms. However, Cysteine and Methionine also contain the sulfur atom.

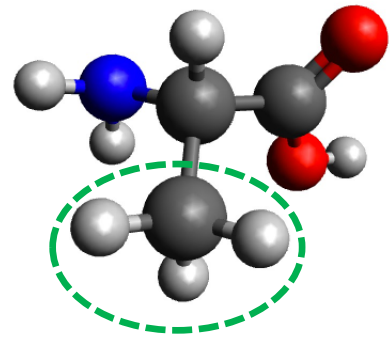
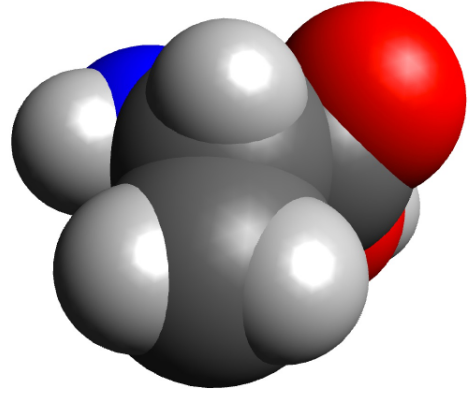
All amino acids have a similar general structure, with an amino group, a carboxyl group and a **side chain (R group)**, which is **different** for every amino acid.



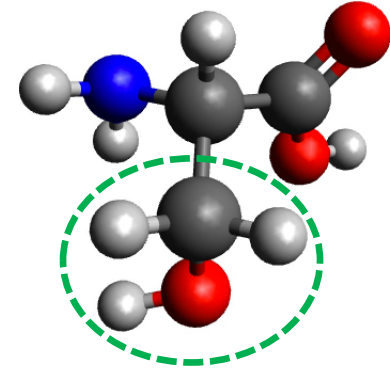
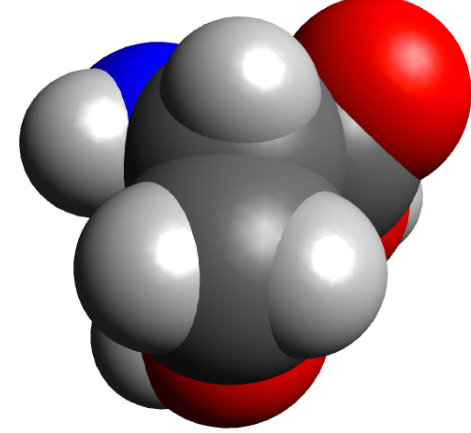
Glycine = Gly = G



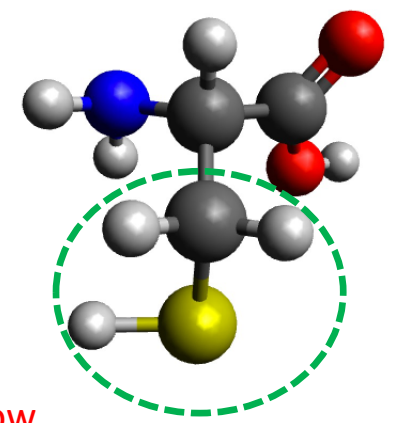
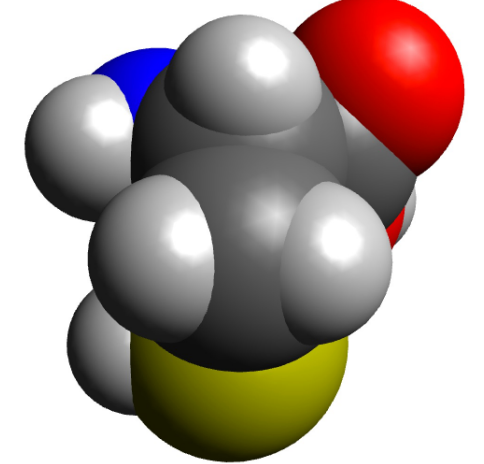
Alanine = Ala = A



Serine = Ser = S



Cysteine = Cys = C

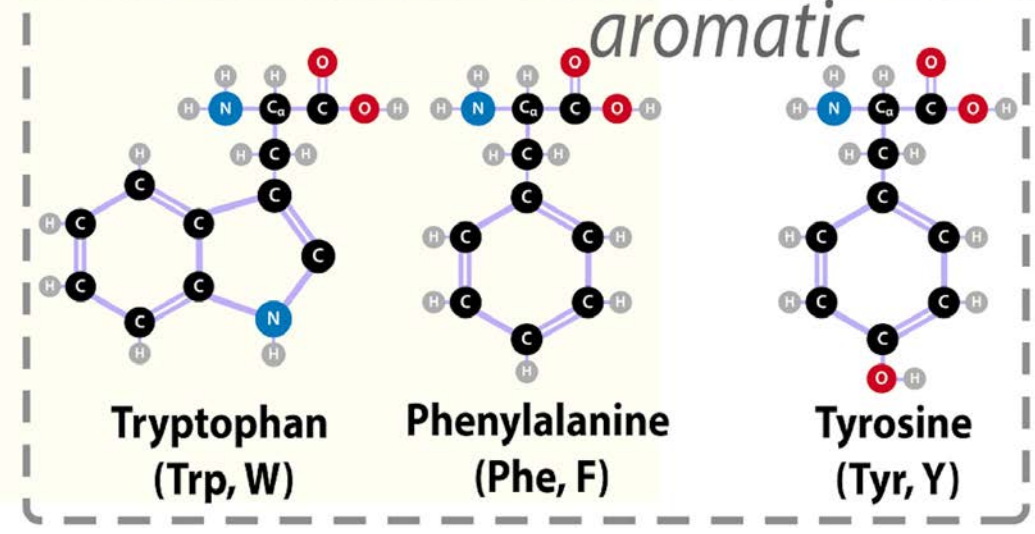
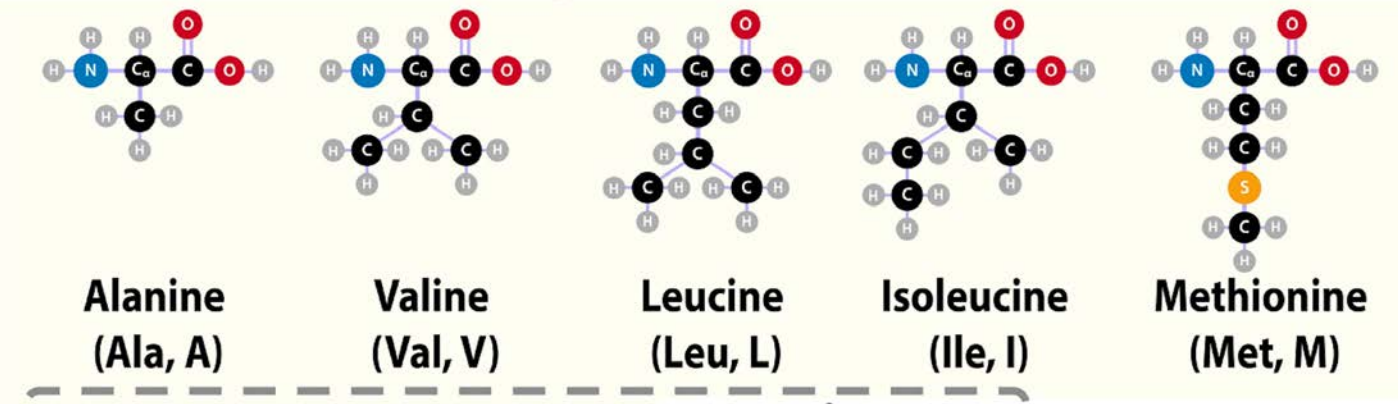


Type in the chemical formula for each of these four amino acids, G, A, S and C, in the boxes below

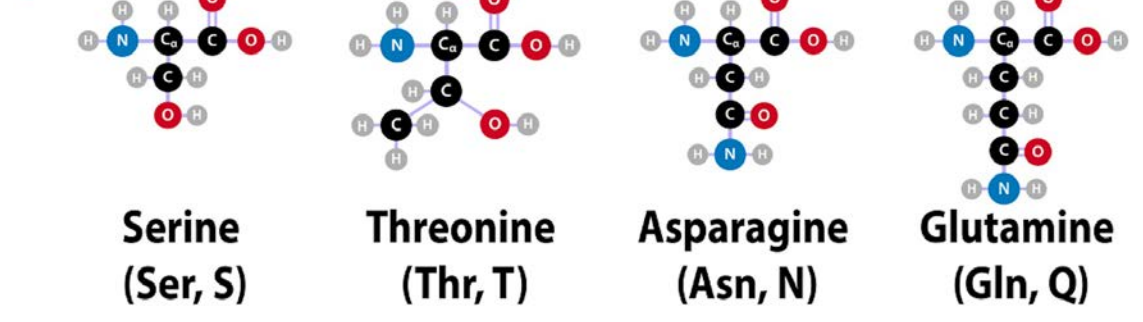
The side chains give amino acids different properties

Here are the 20 amino acids found on your codon chart. The **nonpolar** uncharged amino acids tend to avoid water, so they are **hydrophobic** and are often found on the inside of folded-up protein molecules. On the other hand, **polar** and **charged** amino acids are happy to be exposed to a watery environment, so they are **hydrophilic** and often found on the outside surface of protein molecules. **Interactions between the side chains (R groups) of amino acids in a protein are super-important to the folding and function of that protein.**

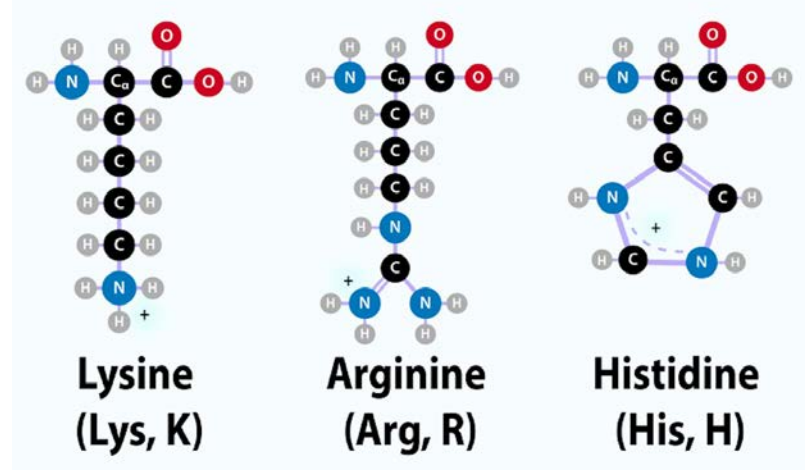
nonpolar, uncharged



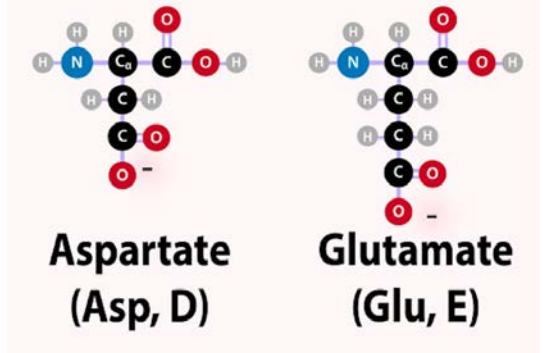
polar, uncharged



sometimes + (Positive)

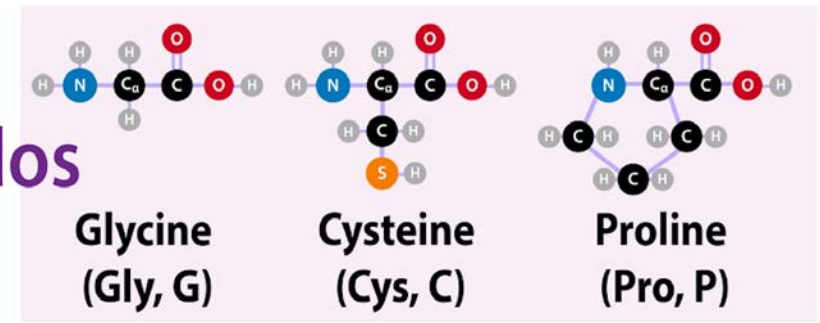


sometimes - (Negative)



You don't need to memorise these structures! The main thing to know is that some amino acids are small, some are bulky, some are hydrophobic (non-polar), some hydrophilic (polar), some have a positive charge and some a negative charge.

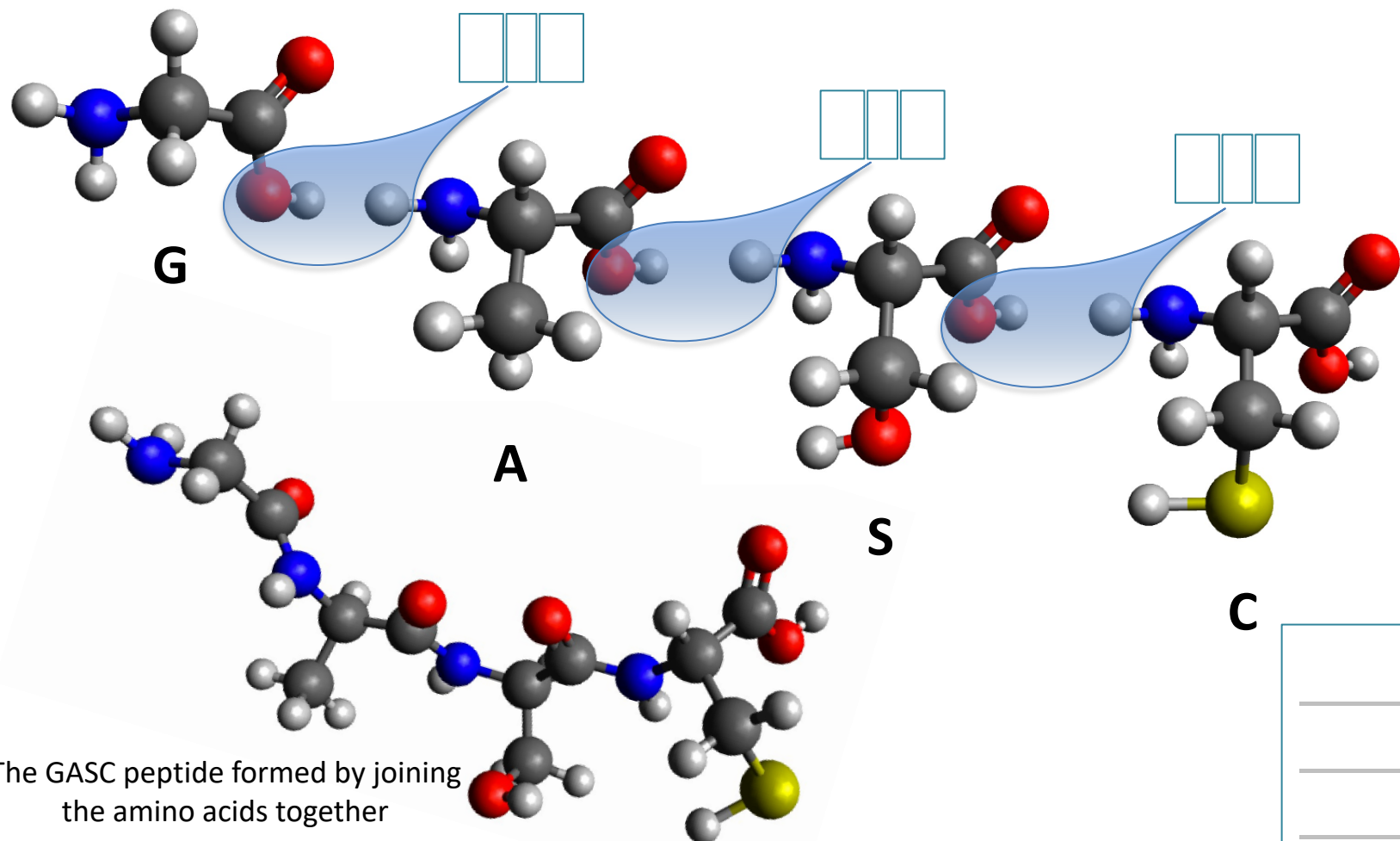
weirdos



Proteins are formed from a chain of amino acids

Proteins are made by complex biological machines called ribosomes. These translate information from a sequence of codons into a sequence of linked amino acids. The codon code used by ribosomes is read from **ribonucleic acid (RNA)** rather than DNA. This is because in order to make a protein, the genetic instructions in DNA (stored in the cell's nucleus) are first converted into an RNA copy (messenger RNA, or mRNA), which can move from the nucleus to the ribosomes, which are located in the **cytoplasm** and on the **rough endoplasmic reticulum**. There are three main differences between DNA and RNA. 1) DNA is made of two strands (as we will discuss later) whereas RNA is single stranded. 2) In DNA, the backbone joining nucleobases together uses the sugar **deoxyribose**, whereas in RNA the sugar is **ribose**. 3) Instead of the Thymine found in DNA, RNA uses a similar nucleobase called **Uracil**. In a famous experiment in 1961, Marshall Nirenberg and Heinrich Matthaei made an RNA from a long chain of uracils, then fed this to ribosomes. When they purified the protein that was translated, they figured out what it was made of and so identified the very first codon. **What do you think they found?**

The diagram below depicts the chemical reaction that occurs in the ribosome to join amino acids together. (the four amino acids here are those on p.11).



DNA GGTGCAAGTTGT
 RNA GGUGCAAGUUGU
 Protein GlyAlaSerCys

As each amino acid is added, the atoms in the tear drop shapes are removed. **Can you type the formula for this product in the boxes?** When this happens, the carbon in the carboxyl group of one amino acid forms a bond with the nitrogen in the amino group of the next amino acid. **This is called a dehydration synthesis reaction. Why do you think that is?**

The GASC peptide formed by joining the amino acids together

Let's look at three important peptides

Now we have a picture of how peptides and proteins are made, based on DNA sequence (with RNA as an intermediate), we can explore protein structures and the important biological processes they control. Small **neuropeptides** are easiest to examine, so we will start with a few examples of these. Have fun exploring these peptides structures, looking at them from different angles.

Met-Enkephalin

This is a pentapeptide that regulates how we feel pain and respond to stress. Click [Met-Enkephalin](#) to explore in 3D (or go to <https://tinyurl.com/ULC-Enk>). You can also scan the QR code on page 40. Click the structure and drag to rotate it, and identify each amino acid using the letter sequence at the top of the view as a guide. What is the met-enkephalin sequence?

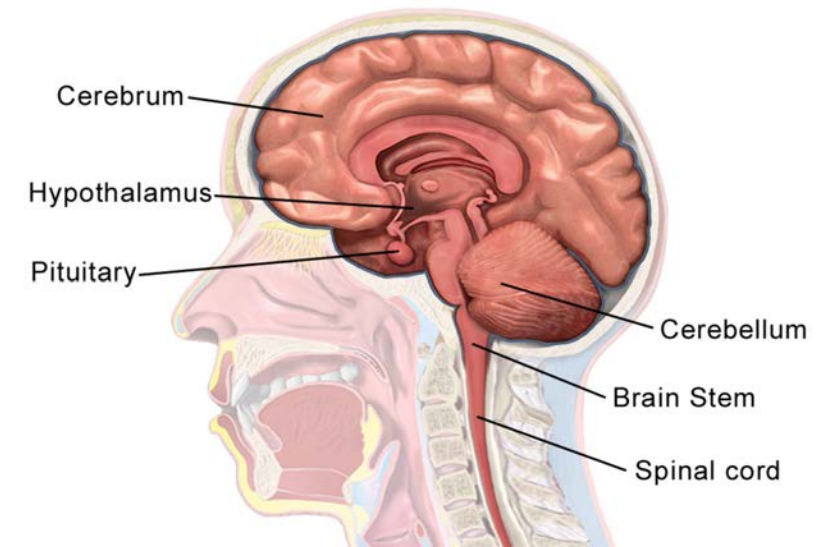
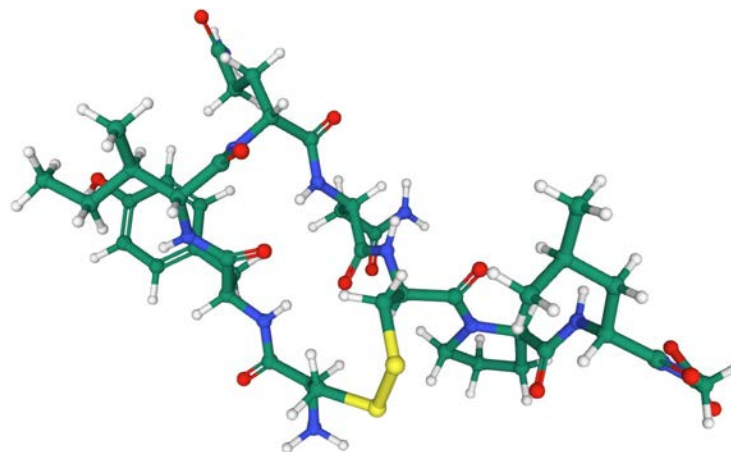
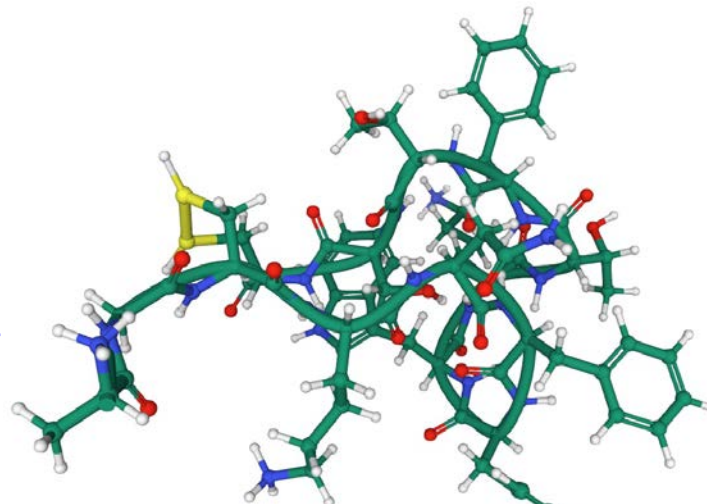
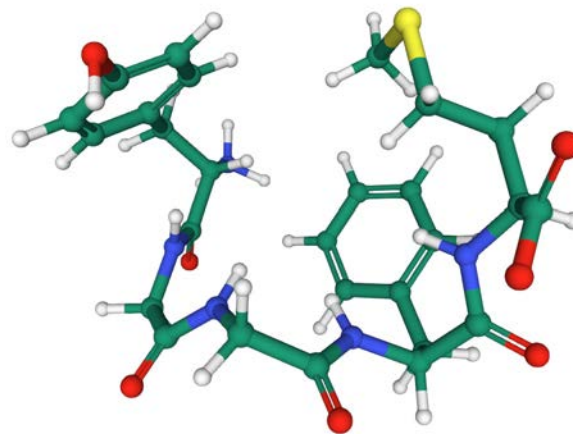
Somatostatin

This peptide has the sequence AGCKNFFWKFTSC. Somatostatin from the hypothalamus inhibits the pituitary gland's release of growth hormone. Click [Somatostatin](#) or go to <https://tinyurl.com/ULC-Som> (or use the QR code on p.40). How many aromatic side chains do you see sticking out from the structure?

Oxytocin

Oxytocin (CYIQNCPLG) is produced in response to love and during labour and has an important role in reproduction and childbirth. Click [Oxytocin](#) to explore (QR or go to <https://tinyurl.com/ULC-Oxy>). What do you notice about the cysteines in oxytocin (and somatostatin)?

Note: carbon atoms are shown in green rather than black here

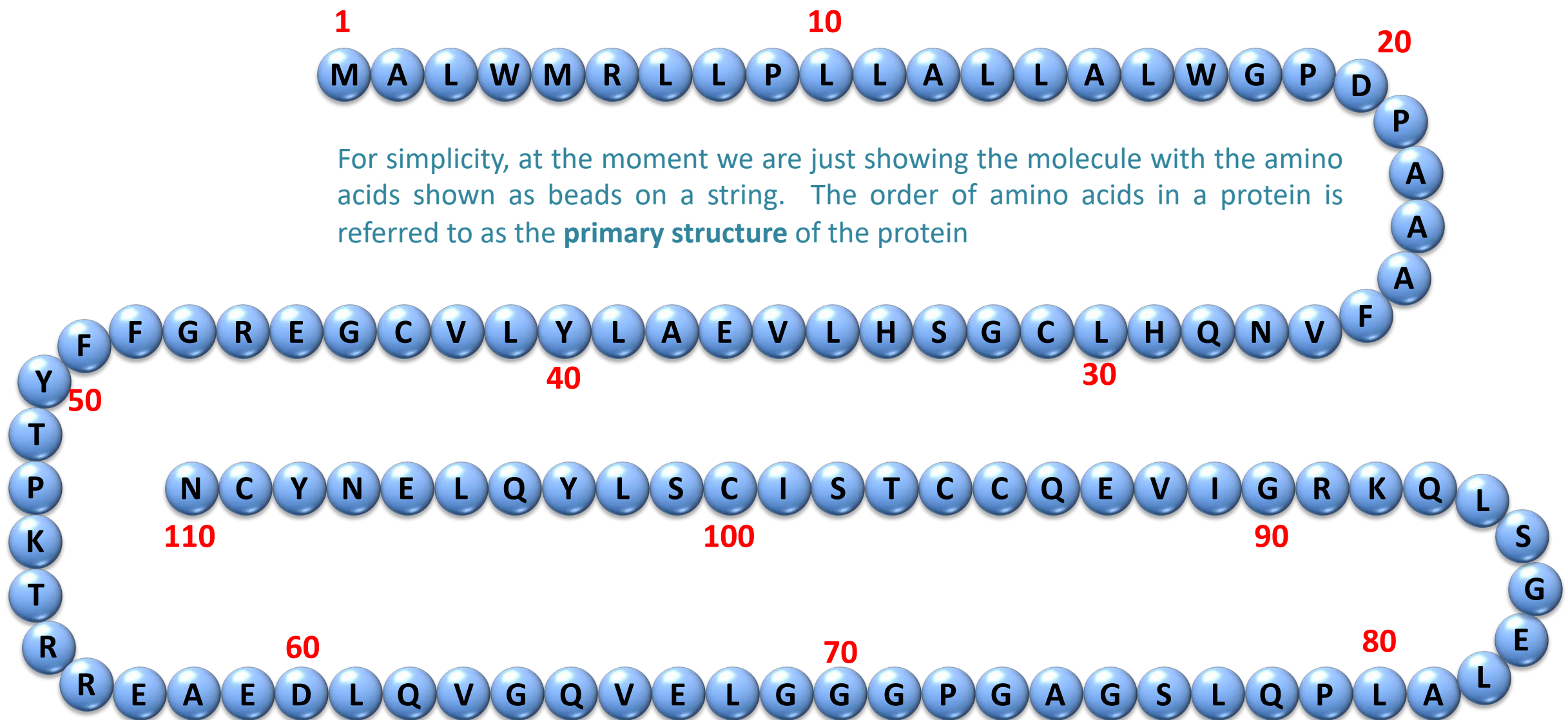


Wikimedia CC-BY-SA-4.0 BruceBlaus 2016

The hypothalamus and pituitary have important roles in **homeostasis** (the process of maintaining a relatively constant internal environment by regulating temperature, thirst, feeding, growth and behaviour). They release the neuropeptide hormones shown here, which act on other neurons in the brain as well as cells of other organs. The **Hypothalamus** and **Pituitary** also work together with the **Adrenal** glands (the **HPA axis**) to regulate the body's response to stress.

The insulin protein also functions as a hormone

This is the sequence of amino acids in the protein initially made by the human insulin gene. **Insulin is a hormone that is essential for regulating the level of a sugar, glucose, in your blood.** The version of the protein shown below is called pre-pro-insulin. All proteins made by the ribosome begin their sequence with a methionine. The reason that the neuropeptides on the previous page didn't start with a methionine is because they are actually fragments of larger proteins that did start with a methionine. As we will see, the insulin hormone is also a processed version of the pre-pro-insulin molecule shown below.



How many DNA nucleobases are needed to code for the pre-pro-insulin protein?

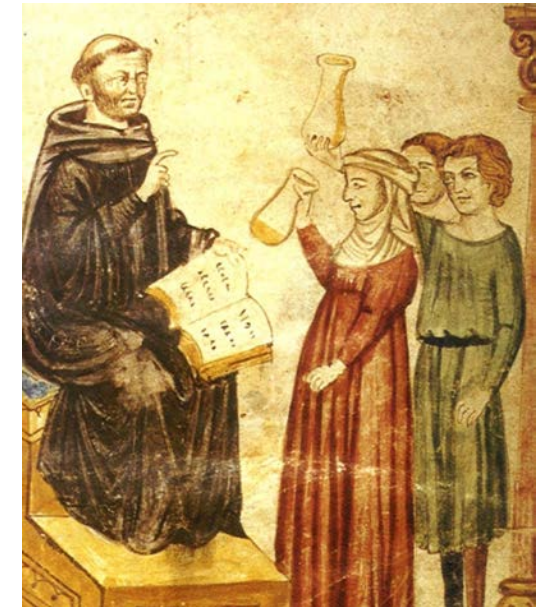
Why is insulin important?

Your body tightly regulates your blood glucose levels. However, **if someone has diabetes, then their body cannot control blood sugar (glucose) levels properly** and they have to be very careful about their diet. Although diabetes is now more frequent than in the past, this disease was known to ancient Egyptian, Greek and Hindu medical experts. Over the centuries, countless people died from diabetes, so there was a lot of interest in its cause.



Fhgitarre, Ants on Sugar, Flickr, CC BY-2.0

Early physicians from ancient Egypt and Greece noticed that the urine of people with diabetes was attractive to ants. This was because their urine contained lots of sugar. Diabetics produce a lot of urine that is described as tasting as though it was mixed with honey (!!!). In the Middle Ages people were employed as ‘water tasters’, testing people suspected of having the disease by tasting their urine.



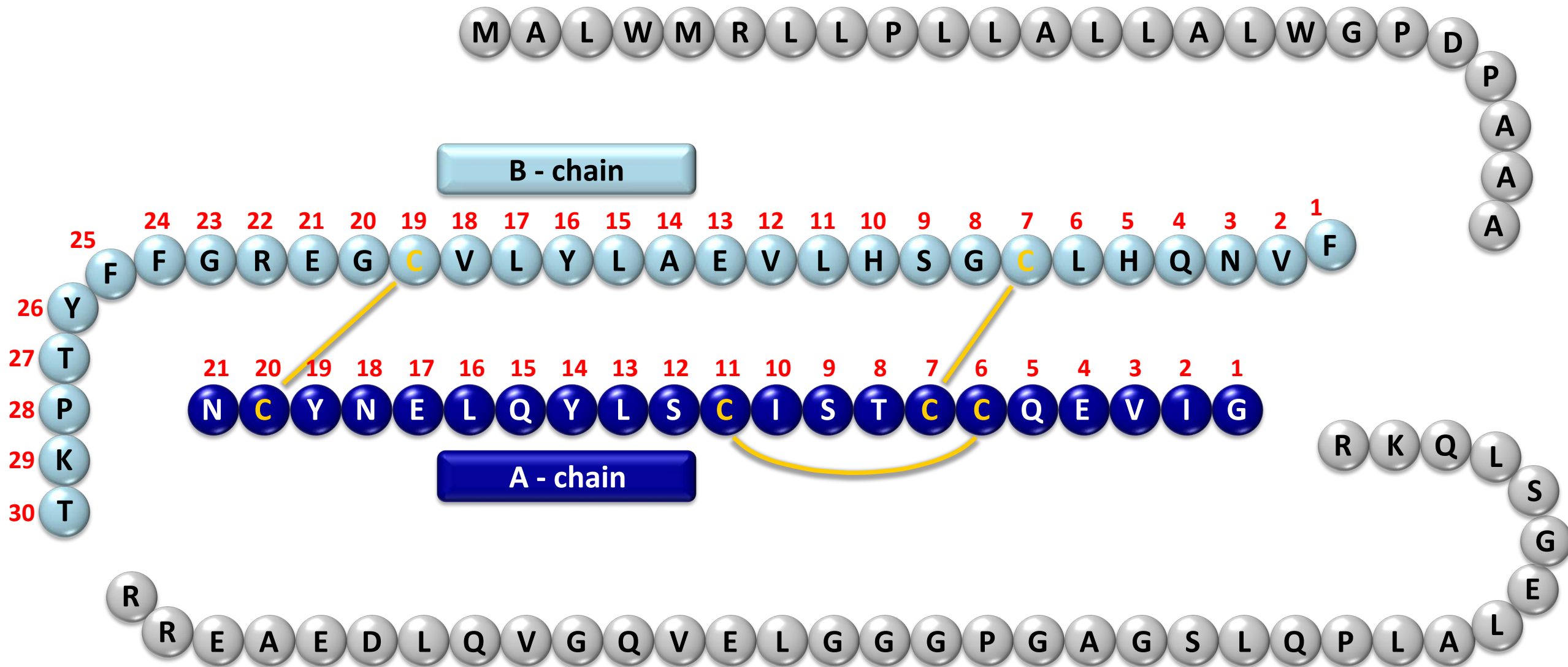
Wikimedia Commons, Public Domain

In 1889 two European scientists, working in France, discovered that the pancreas was an important organ for regulating blood sugar levels. In the 1920's, the researchers Frederick Banting, Charles Best, John Macleod and James Collip used this information to purify the insulin protein from pancreas. They found it could be used to treat diabetic patients, saving millions of lives. Before that, one of the main treatments for **Type I diabetes** was starvation. Two of the researchers, Banting and Macleod, were awarded the Nobel prize in 1923 for the discovery of insulin. A 24-minute long [documentary](#) describes the history of this major medical discovery (or go to <https://tinyurl.com/ULC-Ins>). **Please note:** *the video begins with a few disturbing images, from 100 years ago, of children that were ill with Type I diabetes, on the starvation diet and very thin.*

Insulin was also the first protein to have its primary structure identified, by Fred Sanger, who was awarded a Nobel prize in 1958 for figuring out how to sequence proteins, showing how proteins are made from ordered chains of amino acids.

Processing pre-pro-insulin into the mature hormone

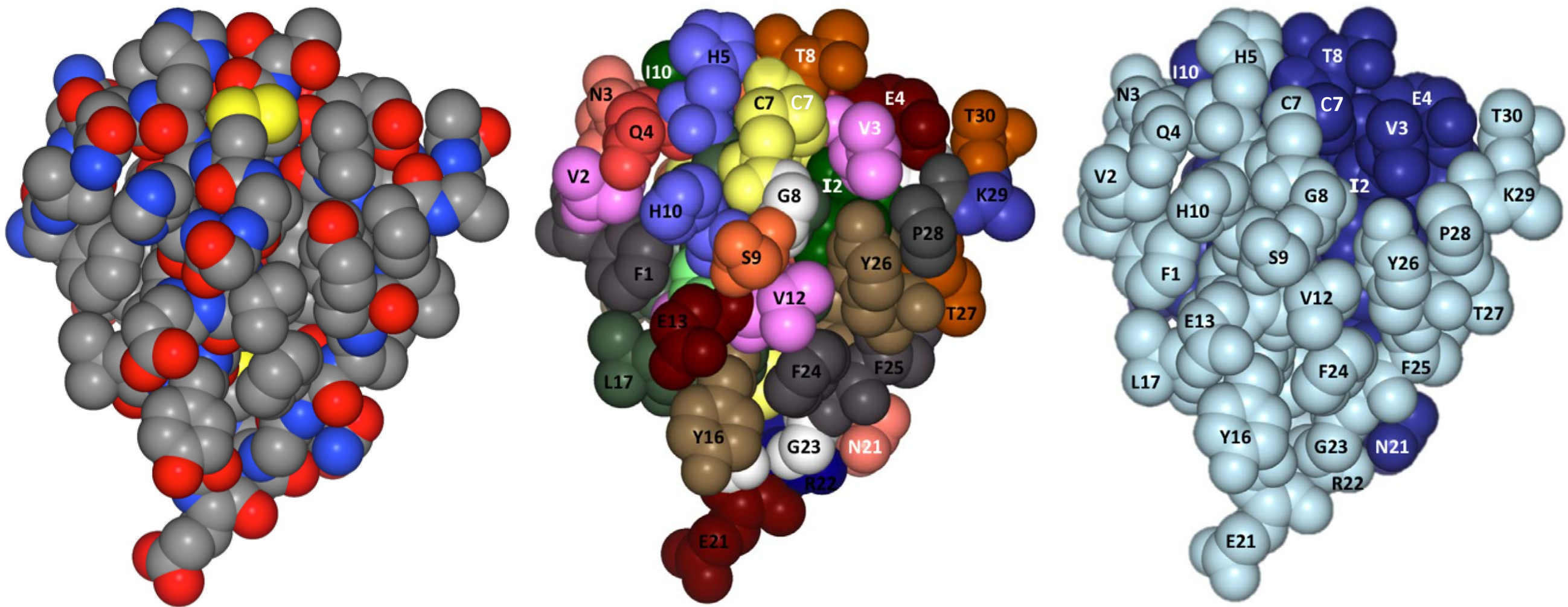
Before insulin can work, some of the amino acids of pre-pro-insulin (shown in gray) are chopped out. However, the A and B chains stick together by chemical links (**disulfide bonds**) that join the sulfur atoms of cysteines (shown as golden lines). Similarly, you might have noticed when exploring the 3D views of somatostatin and oxytocin that their two cysteine amino acids are connected, making the peptide chain loop onto itself in a circle (making them cyclopeptides).



For *other* proteins, a variety of different changes can be made after they are formed. For example, fats can be joined to proteins at their Glycine (G) and Cysteine (C) amino acids, sugars can be attached to Asparagine (N) and phosphorus atoms can be rapidly attached and removed from Serine (S), Threonine (T) and Tyrosine (Y) amino acids, turning a protein's function off and on, sort of like a light switch. As we will see later, histones, important proteins that bind DNA, can have methyl (CH₃) groups attached to their tails, or be acetylated (CH₃CO added), altering the packaging (chromatin structure) of DNA in the cell nucleus chromatin structure to turn genes on or off.

So what does the insulin protein look like?

These space-filling models of the A and B chain of insulin are a bit closer to the shape this protein actually takes. In the model on the left, you can see the atoms making up the amino acids in insulin. The hydrogen atoms have been left off to make it less complicated but you can still see carbon, nitrogen, oxygen and sulfur atoms (*notice how the sulfur atoms of the C7 amino acids of the A and B chains are adjacent to each other*). In the middle image, each type of amino acid is given a different colour - so now all the atoms that make up the cysteines (except hydrogens) are coloured yellow. On the right, the A and B chains of insulin are shaded blue, as on the previous page.



On the right hand model, click on 10 amino acids that occur in sequence in the B chain of insulin (the primary structure on the previous page might help). You can probably also find other short runs of amino acid sequence too.

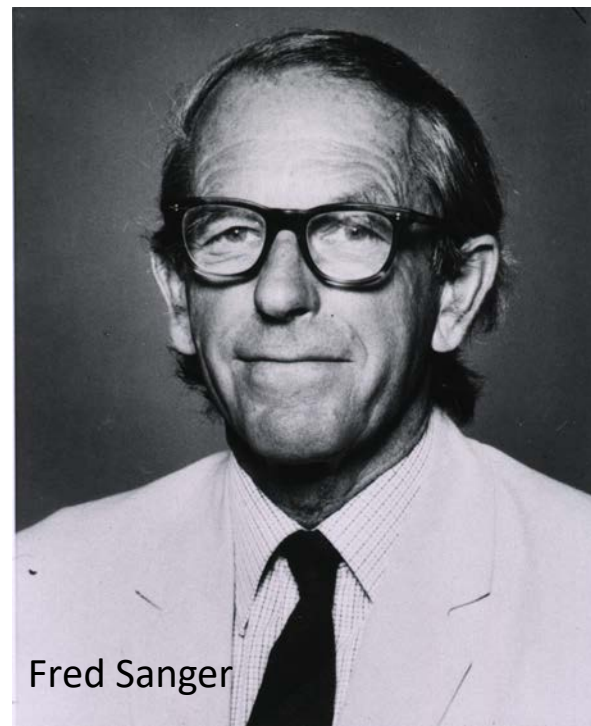
DNA and protein sequence of the B chain of human insulin



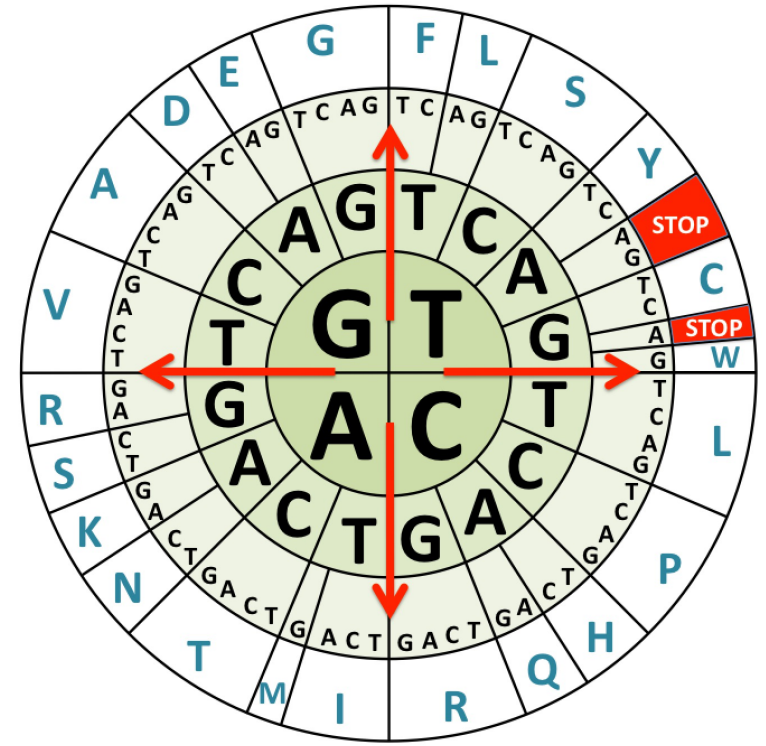
Some people with Type I diabetes have mutations in one of the two adjacent phenylalanines (F or Phe) in the insulin B chain. We call these **missense mutations** (variants) Phe24Ser (F24S) and Phe25Leu (F25L). Rather than use the term mutants, for people we use the term **genetic variants**. **Type on the dashed line missense mutations that would result in insulin proteins having a Serine at amino acid 24 or a Leucine at amino acid 25. Name another amino acid/position you think might important, giving a reason why.**

In 1951, Fred Sanger was the first person to determine the sequence of a complex biological molecule. He achieved this major advance by studying the B-chain of the insulin protein. He had to do this using a series of chemical reactions with acid to break the protein into smaller pieces that he then purified and identified (DNA sequencing wasn't invented at the time).

Sanger on science: "It is like a voyage of discovery into unknown lands, seeking not for new territory but for new knowledge. It should appeal to those with a good sense of adventure".



Fred Sanger
National Library of Medicine, Public Domain

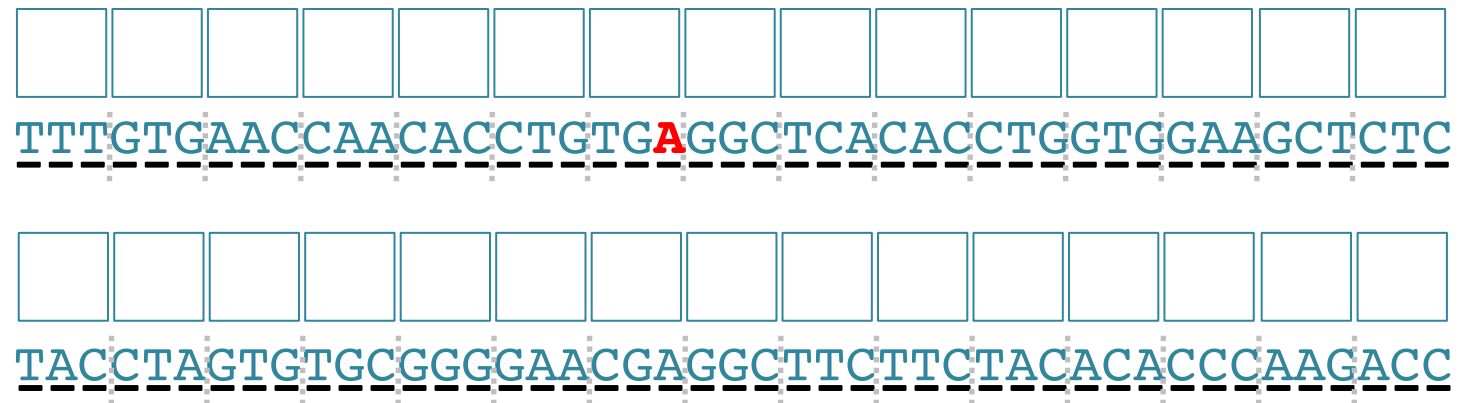


Some mutations dramatically alter proteins

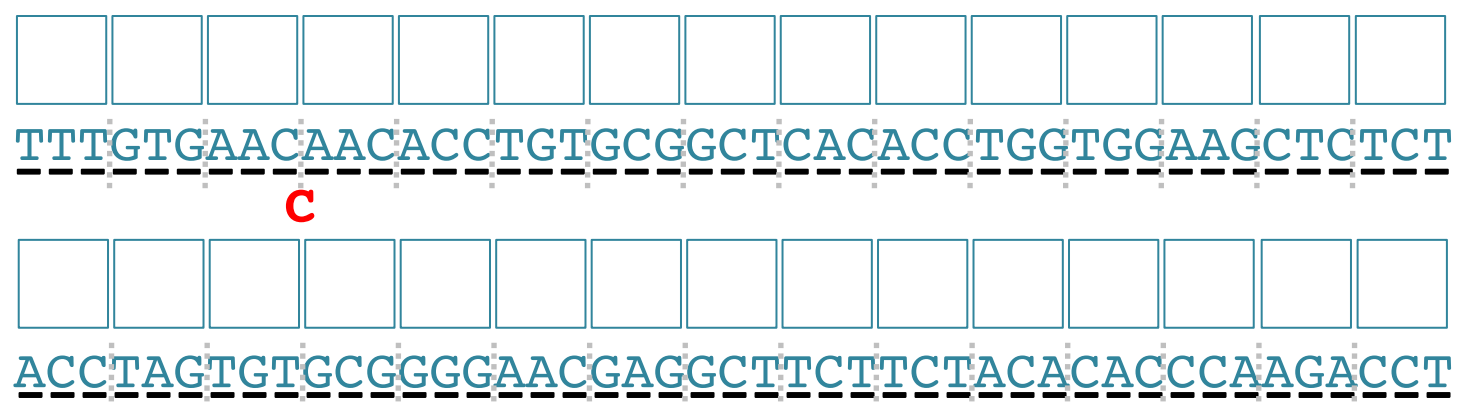
This is the normal Insulin B chain sequence you have already seen



In a **nonsense mutation**, one of the three stop codons is created and no amino acids are added after it (in the example here, a single base change at the third position of a codon for cysteine is mutated, resulting in the Opal stop codon, TGA). **What do you think happens to the Insulin A chain when this mutation occurs?**



In a **frameshift mutation**, one or two bases are deleted or inserted. Here, just the first base of the glutamine codon (Q) is deleted. **What do you think happens to the B and A chains when this mutation occurs?**



Comparing the genetic code and computer storage

So, now you know that the sequence of DNA bases in a gene provides the instructions for the sequence of amino acids in a protein, and that each amino acid is based on a three-base codon. In some ways, this is similar to how your digital devices work and store information. Information in a digital device is handled and stored through a binary code of 1's and 0's, called bits. The 1's and 0's are like the A,C,T or G of DNA. In the same way that three bases make up a codon, eight bits make up a byte. The particular order of the bits in a byte is very important and determines what character that byte stands for. Each of the letters on your computer keyboard is represented by a distinct set of eight bits. For example:

A = 01000001
C = 01000011
T = 01010100
G = 01000111

Can you figure out how many 1's and 0's would be in a 1 megabyte (1Mb) file?

Therefore, your computer would store the word CAT as:

010000110100000101010100 (3 bytes)

As you might have noticed, in the same way that a single base change in a codon can alter the amino acid in a protein, the switch of a single bit, from 0 to 1, can change a digitally encoded character from an A to a C.

On the dashed line above, show how you would change the bits to convert the word CAT into the word ACT?

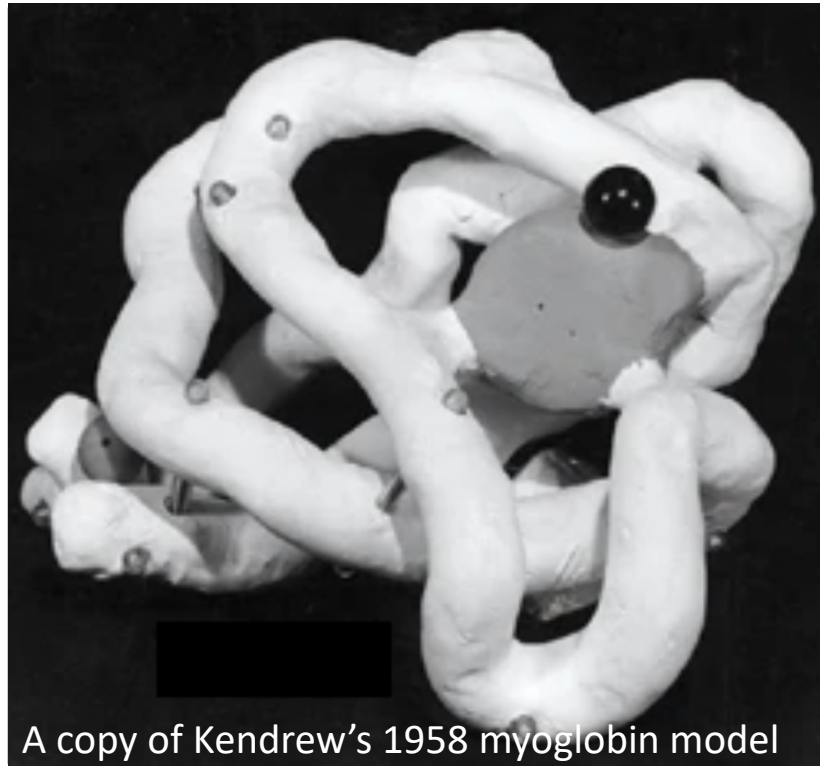
Some scientists are investigating whether we could use DNA as a way to store digital information in a test-tube, rather than on computer hard drives. They claim that 20 grams of DNA (the same amount a person has in their body) would be enough to store 4.2 exabytes* of data, and 2.6 kilograms of DNA could hold all the data currently stored in global data centres (where Google stores all its information is just a small fraction of this). However, DNA-based storage of digital information is currently too slow and expensive for practical use.

Listen to a scientist doing this work on this short video [DNA-based data storage](#) (or if you are reading this on paper, you could use the QR code on page 40, or type in this URL - <https://tinyurl.com/ULC-Data>).

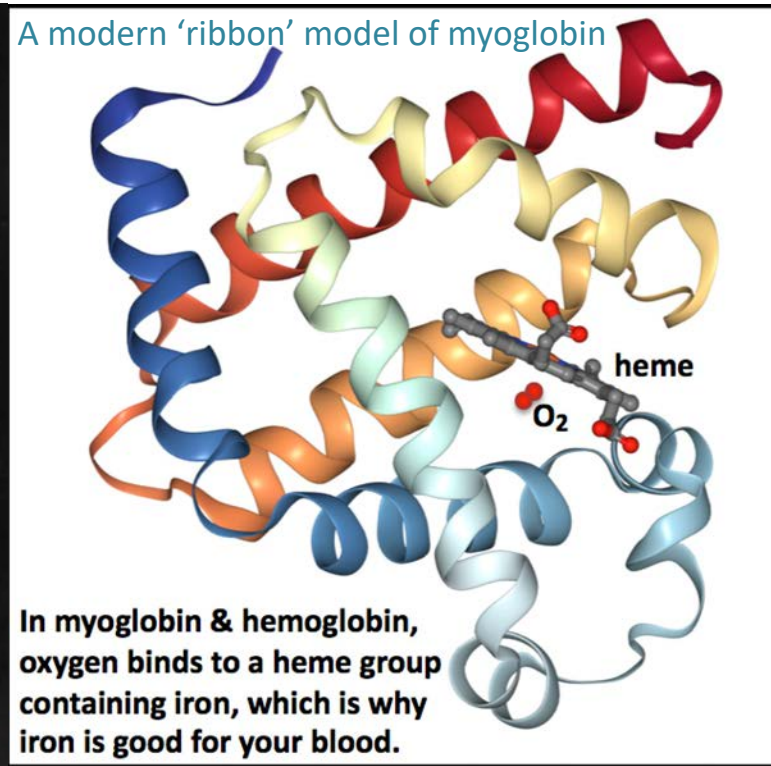
*One exabyte is equal to 1,000,000,000,000 megabytes (or one million terabytes)

Myoglobin: the first 3D protein structure to be discovered

In 1945 John Kendrew joined the research unit of Sir Lawrence Bragg, in Cambridge England. Bragg was a famous X-ray crystallographer and in 1913 had determined the crystal structure of table salt (sodium chloride, NaCl). Kendrew began using X-rays to look at protein structure, a task which many scientists at the time thought was almost certain to fail, because proteins seemed hopelessly complex. When an X-ray is blasted through a protein crystal and hits an atom, the X-ray gets deflected. Doing this lots of times produces an X-ray 'diffraction' pattern that can be used to calculate where atoms are in a molecule. It took Kendrew many years to figure out the structure of the 154 amino acid long myoglobin protein and he was the first person to use a computer to figure out a complex problem in biology. In 1962 he was awarded the Nobel Prize along with Max Perutz who identified the structure of hemoglobin. **Myoglobin is needed for storing oxygen in muscles (it is what makes muscles red) and haemoglobin is the protein in your red blood cells that transports oxygen around your body.**



A copy of Kendrew's 1958 myoglobin model
Alan Fersht / University of Cambridge



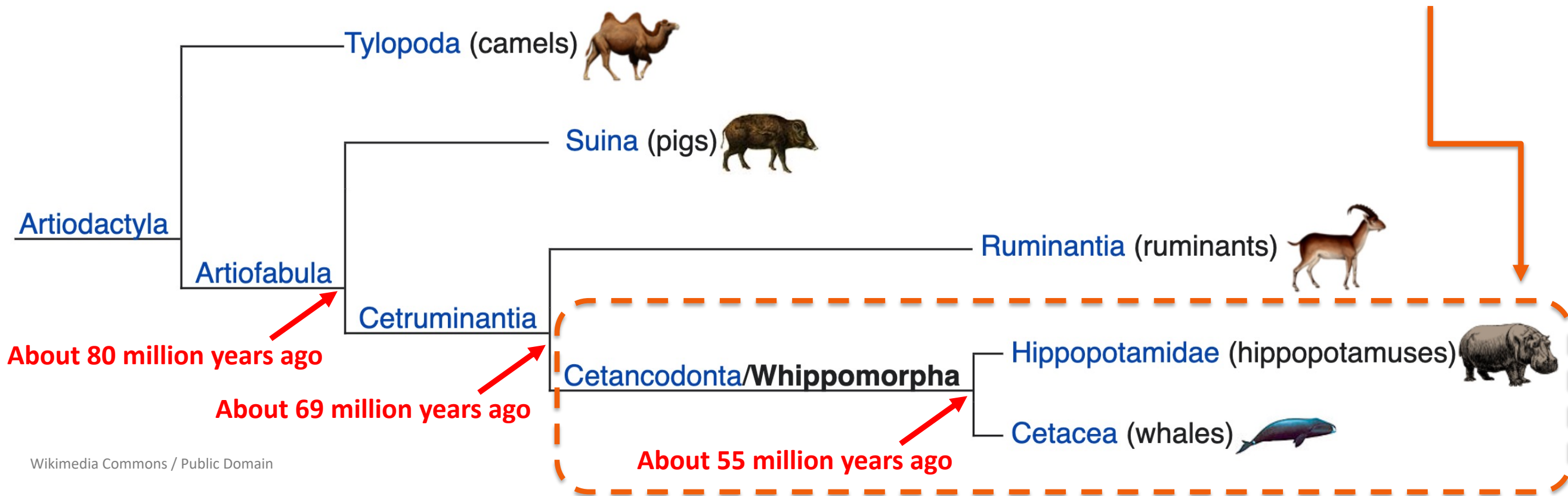
Max Perutz John Kendrew
Keystone/Hulton Archive/Getty Images

Which animal's myoglobin do you think Kendrew studied because the animal has a good supply of muscle myoglobin: cheetah, horse, sperm whale, falcon, penguin, human or yak? Record your reasoning below.

DNA change is the basis for evolution

Kendrew was able to make crystals from Sperm Whale myoglobin, this species (*Physeter catodon*) and other diving marine mammals can have 30 times more myoglobin in their muscles than land mammals. However, when there is a lot of myoglobin it can clump together, stopping it from working properly. During the evolution of whales, 55 to 35 million years ago (mya), some **random** DNA mutations occurred that helped stop myoglobin from clumping up. The prehistoric whales that had some of those mutations were therefore able to produce more myoglobin in muscle, store more oxygen and dive deeper searching for food. This gave those particular whales a **selective advantage**, and they reproduced to pass their version of the myoglobin gene onto the ancestors of the whales and dolphins swimming around in pods (groups) today. **Read the news item from the BBC and watch this video about [myoglobin evolution in diving mammals](#) (or go to <https://tinyurl.com/ULC-Myo>). What do the muscles of whales and seals look like? What type of changes occurred in their myoglobin to make their muscles look this way?**

This diagram (a **cladogram**) shows groups of animals (**clades**) that share a common ancestor. It shows the evolution of the Artiodactyla (mainly hoofed mammals). This branch of mammals separated from the one that led to humans, other primates, rodents and rabbits about 95 mya. On the next page, we focus on evolutionary changes in myoglobin in **Whippomorpha**.



The evolution of myoglobin in cetaceans

This is a **multiple alignment** showing myoglobin from different species. The shaded amino acids are the ones that in dolphins and whales (cetaceans) are **different** from the hippopotamus, the most closely related land mammal to cetacea.

Hippopotamus	MGLSDGEWQLVLNVWGKVEADVAGHGQEVLI R LFTGHPETLEKFEKFKNLKTEDEM RACE
Orca (Killer whale)	MGLSDGEWQLVLNVWGKVEADLAGHGQDILIRLFKGH PETLEKFDKFKHLKTEADM KASE
Bottlenose dolphin	MGLSDGEWQLVLNVWGKVEADLAGHGQDVLIRLFKGH PETLEKFDKFKHLKTEADM KASE
Bowhead whale	MVLS DGEWQLVLNIWAKVEADVAGHGQDVLIRLFKGH PETLEKFDKFKHLKTEAEM KASE
Humpback whale	MVLS DAEWQLVLNIWAKVEADVAGHGQDILIRLFKGH PETLEKFDKFKHLKTEAEM KASE
Sperm whale	MVLS EGEWQLVLHVWAKVEADVAGHGQDILIRLFKSH PETLEKFDKFKHLKTEAEM KASE
Pygmy sperm whale	MVLS EGEWQLVLHVWAKVEADIAGHGQDILIRLFKHH PETLEKFDKFKHLKSEAEM KASE

▲ 10 ▲ 20 ▲ 30 ▲ 40 ▲ 50 ▲ 60

Hippopotamus	NLKKHGNTVLTALGGILKKKGHHEELKPLAHS HATKHKIPIKYLEFISEAIIHVLQSKH
Orca (Killer whale)	DLKKHGNTVLTALGAILKKKGHHD AELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRH
Bottlenose dolphin	DLKKHGNTVLTALGAILKKKGHHD AELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRH
Bowhead whale	DLKKHGNTVLTALGGILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISDAIIHVLHSRH
Humpback whale	DLKKHGNTVLTALGGILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISDAIIHVLHSRH
Sperm whale	DLKKHGNTVLTALGAILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRH
Pygmy sperm whale	DLKKHGNTVLTALGAILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRH

▲ 70 ▲ 80 ▲ 90 ▲ 100 ▲ 110 ▲ 120

Hippopotamus	PGDFGADAQGAMNKALELFRNDIAAKYKELGFQ G
Orca (Killer whale)	PAEFGADAQGAMNKALELFRKDIAAKYKELGFH G
Bottlenose dolphin	PAEFGADAQGAMNKALELFRKDIAAKYKELGFH G
Bowhead whale	PGDFGADAQGAMNKALELFRKDIAAKYKELGFQ G
Humpback whale	PADFGADAQAAMNKALELFRKDIAAKYKELGFQ G
Sperm whale	PGDFGADAQGAMNKALELFRKDIAAKYKELGYQ G
Pygmy sperm whale	PADFGADAQGAAMNKALELFRKDIAAKYKELGYQ G

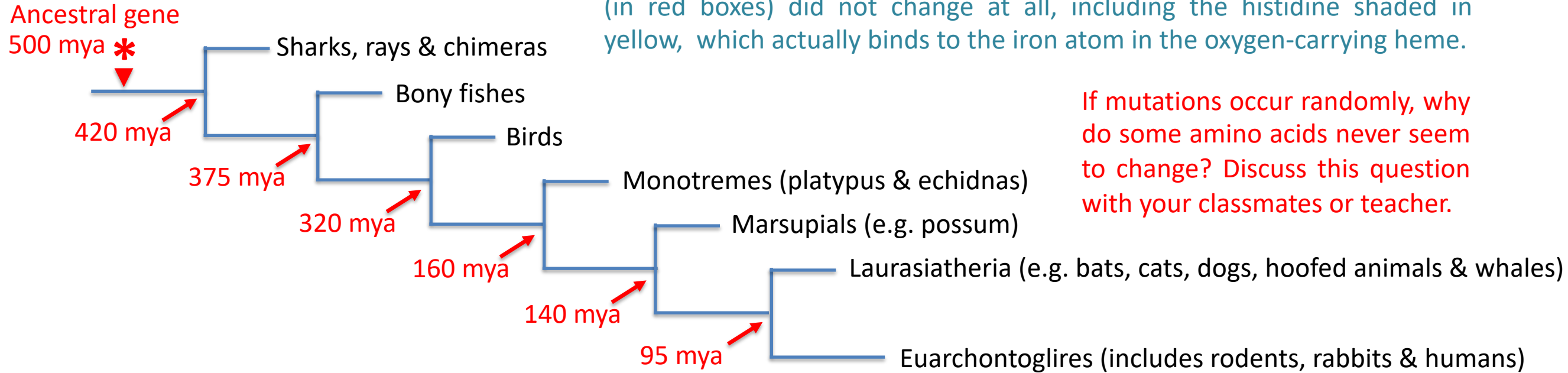
▲ 130 ▲ 140 ▲ 150

Which of the species shown here do you think is most closely related to Orca? Why?

Based on the BBC news story on myoglobin *and the amino acid structures on page 12*, Name changes in myoglobin you think might have enabled the muscles of all dolphins and whales to store lots of myoglobin. What other changes might have been important to deep-diving Sperm Whales? What mutations millions of years ago might have resulted in the Q117H change?

A deeper look into myoglobin evolution

Here we are just looking at amino acids 61-120 of myoglobin. Scientists have predicted the protein made from the ancestral gene for myoglobin and haemoglobin, around 500 million years ago (mya) in the Cambrian era. Notice that some amino acids (in red boxes) did not change at all, including the histidine shaded in yellow, which actually binds to the iron atom in the oxygen-carrying heme.



If mutations occur randomly, why do some amino acids never seem to change? Discuss this question with your classmates or teacher.

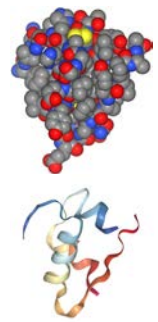
Human	DLKKH	GATV	LTAL	GGIL	KKKG	HHEA	EIKPL	AQSH	ATKHK	IPV	KYLE	FISE	CIIQ	VLSK	H
Dog	DLKKH	GNTV	LTAL	GGIL	KKKG	HHEA	EELKPL	AQSH	ATKHK	IPV	KYLE	FISDA	IIQV	LQSK	H
Orca	DLKKH	GNTV	LTAL	GAIL	KKKG	HDAE	EELKPL	AQSH	ATKHK	IPI	KYLE	FISE	AIH	VLSR	H
Possum	DLKKH	GATV	LTAL	GNIL	KKKG	HHEA	EELKPL	AQSH	ATKHK	ISI	QFLE	FISDA	IIV	LQSK	H
Koala	DLKKH	GTTV	LTAL	GNIL	KKKG	QHEA	EELKPL	AQSH	ATKHK	ISV	QYLE	FISE	AIQ	VILQ	H
Platypus	DLKKH	GGTV	LTAL	GNIL	KKKG	QHEA	EELKPL	AQSH	ATKHK	ISI	KFLE	FISE	AIH	VLSK	H
Echidna	DLKKH	GGV	LTAL	GSIL	KKKG	QHEA	EELKPL	AQSH	ATKHK	ISI	KFLE	FISE	AIH	VLSK	H
Kiwi	DLKKH	GVTV	LTL	LGKIL	KQKGN	HHEA	EELKPL	AQSH	ATKHK	IPV	KYLE	FISE	VI	LKVI	AEKH
Kea	DLKKH	GATV	LTL	LGKIL	KQKGN	HHEA	EELKPL	AQT	HATKHK	IPV	KYLE	FISE	VI	LKVL	AEKH
Tuna	AVSAH	GATV	LKKL	GELL	KAKG	SHAA	ILKPL	ANS	HATKHK	IPIN	NFKL	LISE	VLV	KVM	HEKA
NZ Spotty	AISAH	GATV	LKKL	GELL	RAKG	SHAD	ILKPL	ANS	HATKHK	IPIN	NFKL	LITE	VI	LKVM	EKA
Great White	DLRKH	ANTV	FRAL	GNIL	KQKGN	HGAN	VKEL	ADTH	INKHK	IPP	KNFT	LITN	VAV	KVL	TEMY
Ghostshark	DVKTH	GNTV	FKAL	GDVV	KQKGN	HASNL	QALAT	THINKHK	IPP	QNFT	LITN	VILK	VFA	EKF	
* Ancestral M/H	QVKAH	GKRV	MSAL	GDVV	QHLD	NLSS	VLKPL	AEKH	HANKHK	VDP	HNFK	LLSD	VIL	AVL	AEKF

Annotations below sequence: * * * * * * * * * * * * * *

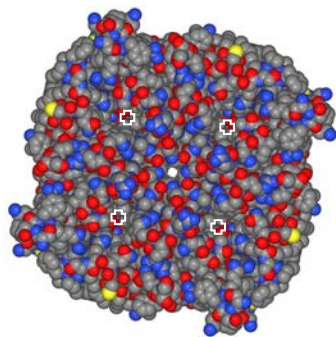
Position markers: 70 80 90 100 110 120

You have more than 20,000 different proteins in your body

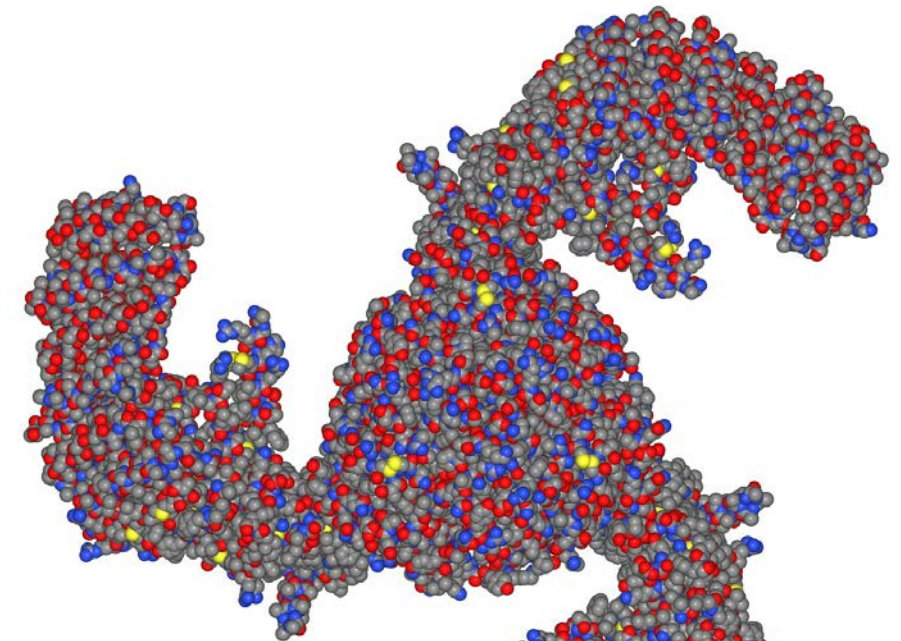
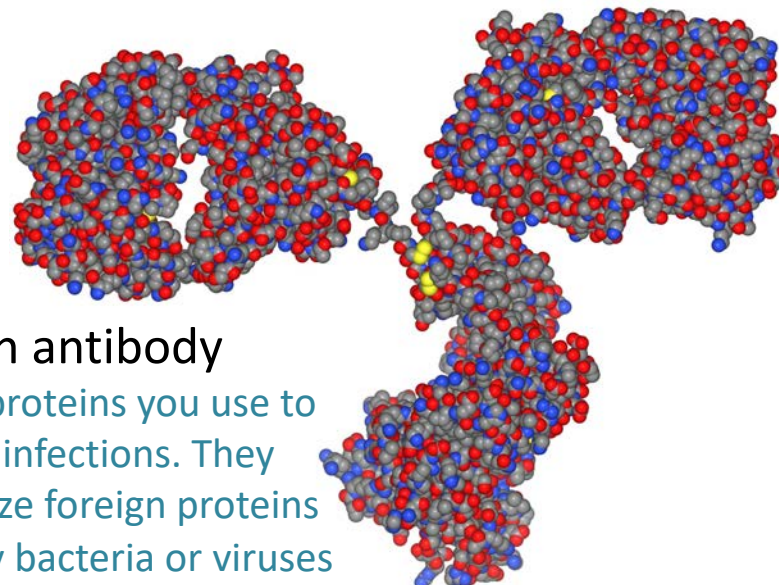
You now know that proteins are chains of amino acids coded for by the order of DNA bases. Because there are 20 different types of amino acid, they can be arranged in many different ways to make proteins with very different sizes, shapes and functions. Scientists still have no idea what one third of all human proteins do. If you have a sense of adventure, perhaps you fancy a career helping to figure that out.



Insulin
Regulates blood sugar levels

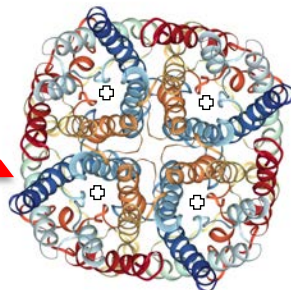


an antibody
These proteins you use to fight infections. They recognize foreign proteins made by bacteria or viruses

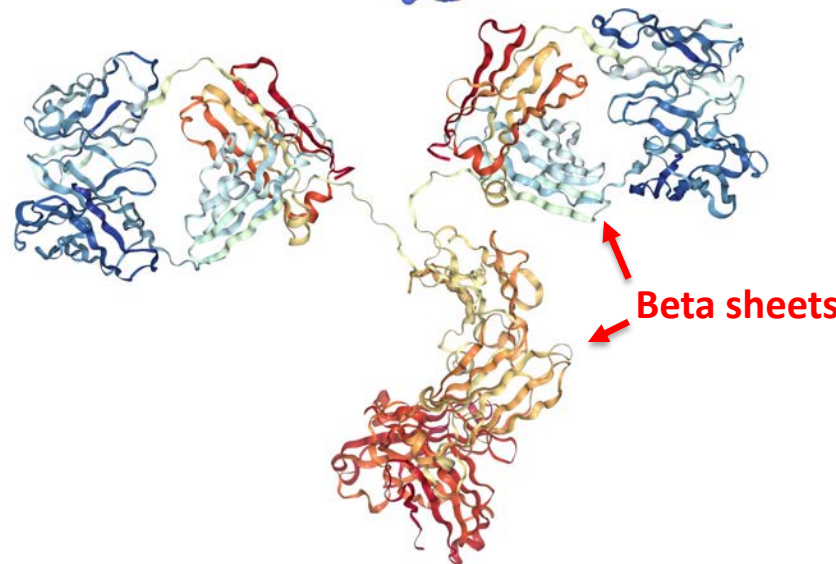


Piezo1
This propellor-shaped protein is found in neurons that end in your skin and is required for your sense of touch

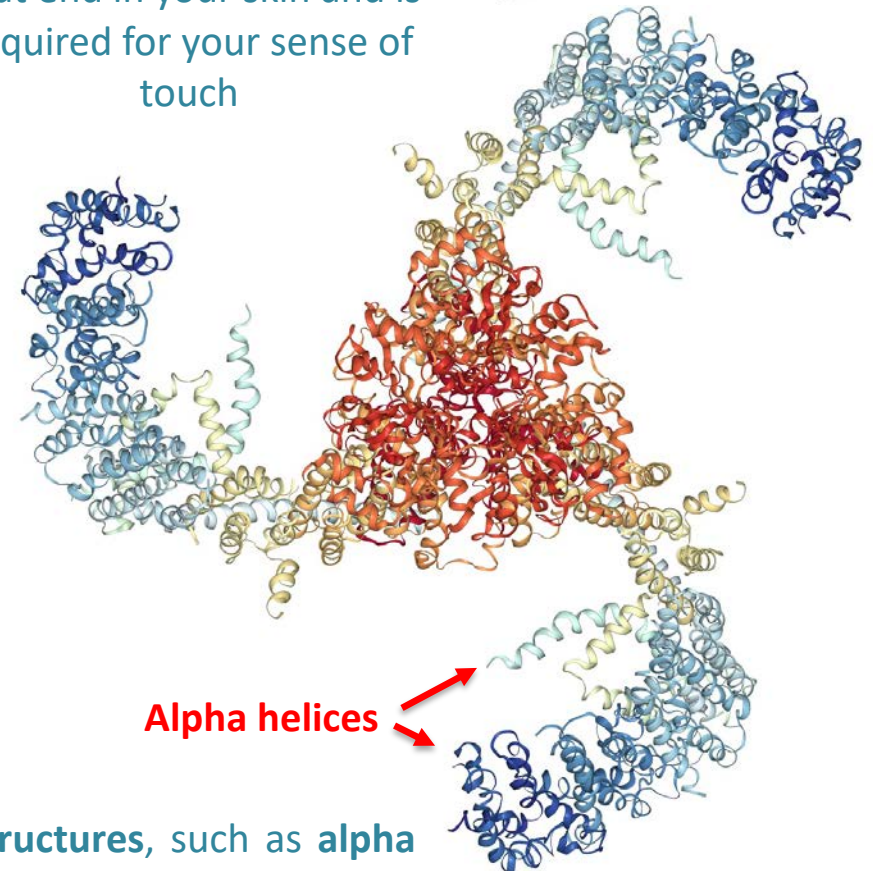
Alpha helix



Aquaporin
This is an important protein in your kidneys. It allows movement of water molecules into cells (through the four places marked with crosses)



Beta sheets



Alpha helices

The 'ribbon' models show the protein backbone and make it easier to see protein **secondary structures**, such as **alpha helices** and **beta-pleated sheets**. In addition, the start of each chain in these proteins is shown in blue and the end in red

Proteins have many, many, functions and can be very large

Every process in your body depends on proteins. You now know that **insulin** is essential to control your blood glucose levels and **myoglobin** stores oxygen in your muscles. **Haemoglobin** transports oxygen in your blood, taking it from your lungs, through the heart and out to all the other tissues in the body. **Antibodies** in your blood fight infections and **aquaporin** regulates how water moves into and out of cells. You have learned that even small fragments of proteins (such as the peptide hormones **somatostatin** and **oxytocin**) can regulate growth, reproduction and social behaviour.

Many proteins, termed enzymes, speed up chemical reactions. Some enzymes help break down food, providing energy and nutrients needed for growth. For example, in saliva the enzyme **amylase** digests bread and other carbohydrates and, in the gut, the enzyme **lactase** helps baby mammals digest the sugar lactose in mother’s milk. As a general rule, proteins with names ending with **-ase** are enzymes. Other enzymes make new copies of DNA before cells divide (**DNA polymerase**), or make RNA copies of genes so that cells of the body can produce the proteins necessary to perform their specific functions (**RNA polymerase**). **Does this raise a “chicken or the egg” problem? What do you think this means for the origin of life?**

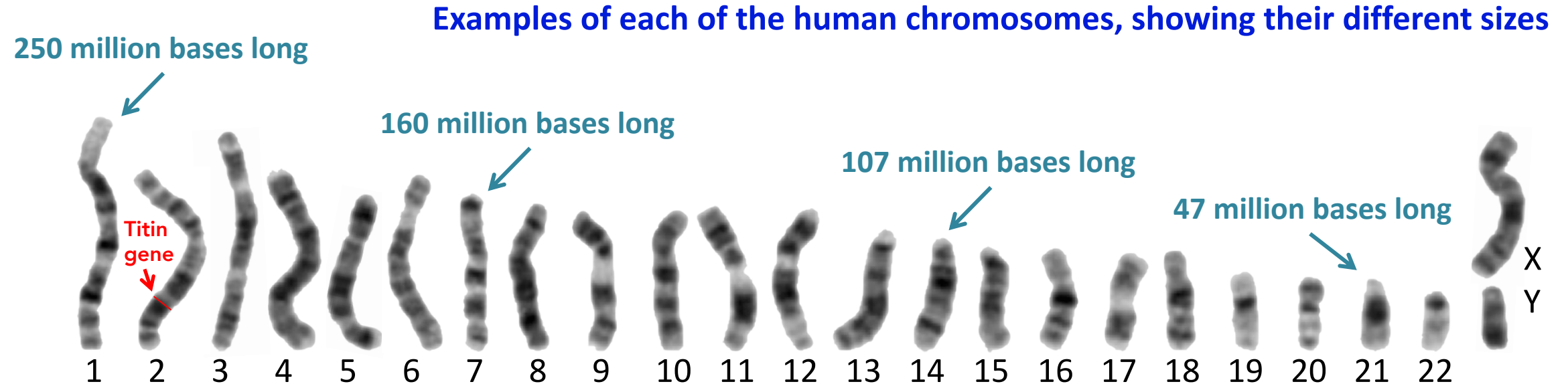
Other proteins have more of a structural purpose. For example, **keratin** proteins make up your hair and nails and **collagen** makes up over a quarter of all the protein in your body and is important for skin, tendon, bone and teeth. Some structural proteins can be very large. Even collagen, with chains of about 1400 amino acids, is small compared to the largest protein, **Titin**, which is up to 38,138 amino acids long. **Titin** has a very important function as a spring in muscle cells (making them elastic) and mutations in titin can cause heart and other muscle diseases. Since each amino acid in a protein is coded by a 3-base codon, the longest possible form of the titin protein would need a DNA coding sequence of 114,414 bases (3 x 38,138).

Congratulations! You have by now learned a huge amount about the genetic code and how it contains the instructions for the many thousands of proteins that every microsecond perform the functions that enable all life to exist. As a little reminder and to learn a bit more, watch this summary video on [protein structure](#) (or use <https://tinyurl.com/ULC-Proteins> or the QR code on page 40).

How is your DNA organised?

You have learned a lot about how DNA codes for proteins, how amino acids are joined together in the ribosome to make those proteins, and how interactions between amino acids determine their shape and function. You also have learned about the atomic structure of nucleobases and the pattern of bonds made by the major atomic elements of life, Carbon, Hydrogen, Nitrogen and Oxygen (although many other elements, including sulfur, iron and phosphorus are very important). In the next section, you will learn about how the nucleobases are joined in DNA and the fascinating history of how DNA was discovered and its structure revealed.

But first we will start with an overview. Within the nucleus of your cells is your **GENOME** – that is, your complete DNA sequence of over three billion DNA letters long. This is divided up into large chunks called **chromosomes**, which are numbered from 1-22 based on their size. Generally, apart from red blood cells and eggs or sperm, **each of the cells in your body has two copies of chromosomes 1-22 (one from each parent) and, if you are a boy, an X and Y chromosome and if you are a girl, two X chromosomes.** Along each chromosome, there are many hundreds, if not thousands, of genes. Each gene is a stretch of DNA, typically several thousand nucleobases long, that codes for a protein.



Even though titin is the largest known gene (like the protein), it still only makes up about one-thousandth of chromosome 2, where it is located (shown by the thin red line). In fact the 20,000 genes that code for all the proteins in your body only cover about 60 million bases.

Scientists are still trying to work out what most of the other 98% of our DNA does.

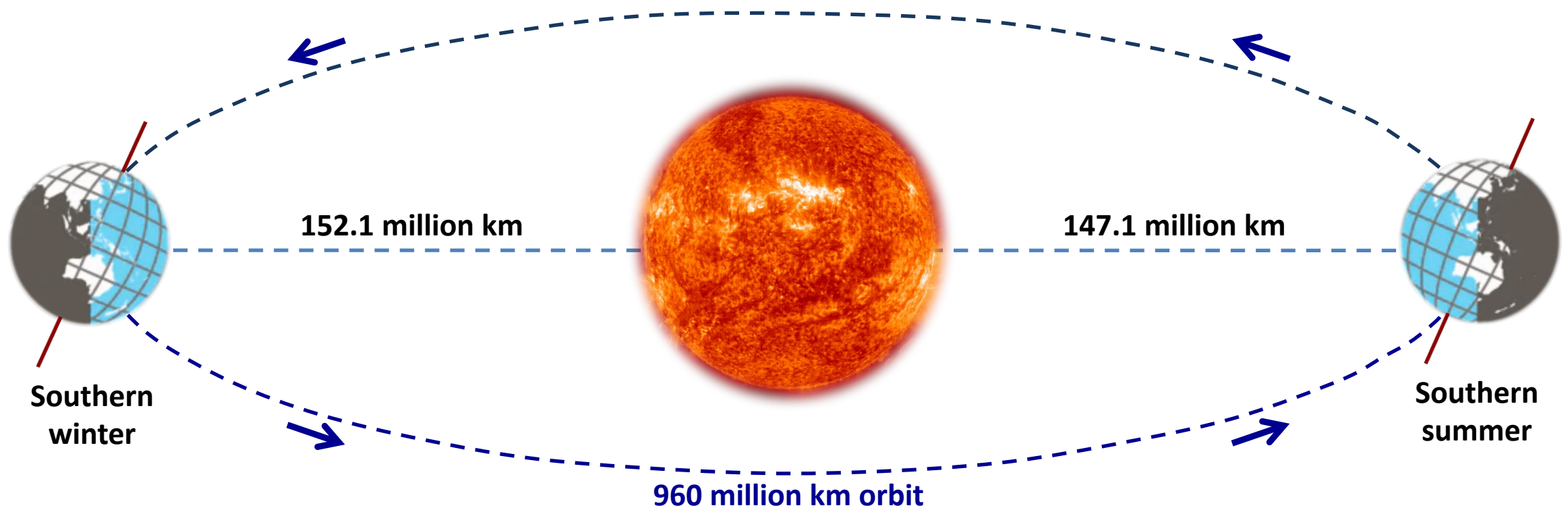
The amount of DNA in you is astronomical – literally!

The chromosomes shown on the previous page are what your DNA looks like when cells divide. A group of proteins, called histones, help to coil the DNA up so that it is tightly packed. Scientists know a lot about the structure of DNA, which we will discuss shortly. Before doing that, an amazing fact is that if the histones were removed and the DNA in just one of your cells was stretched out, it would be about two metres long. In a few pages, we will see whether we can confirm this fact.

How many cells containing your genome do you think you have in your body (click one)?

- One million 1,000,000
- One billion 1,000,000,000
- One trillion 1,000,000,000,000
- Three trillion 3,000,000,000,000

For an adult male, the answer is close to three trillion. If the DNA in each cell could be stretched out to two metres, that would mean a person’s total body DNA could cover a distance of up to six trillion metres, or six billion kilometres (your mileage may vary). This is further than six times around earth’s orbit, or around 40 Astronomical Units (one Astronomical Unit is 150 million kilometres, the average distance from the earth to the sun).



How was DNA discovered?

The story starts with a Swiss Doctor named Friedrich Miescher, who was partially deaf and didn't want to work with patients. Instead, Miescher became interested in leukocytes, a type of white blood cell found in pus. He collected used bandages from the local surgical clinic and use these as a source of leukocytes. This was in 1869, when antiseptic methods in surgery were only starting and before antibiotics were discovered, so he had a good supply of pus to work with, enabling him to make one of the most important discoveries in science.

Miescher carefully washed the cells from the bandages, then broke them up and separated different parts of the cells to test the chemical make-up of the substances he purified. To his surprise, he found some material that contained large amounts of phosphorus, an element uncommon in proteins. He called this new substance **nuclein**, because it came from a special structure inside the cells, called the nucleus.

Other chemists became interested in nuclein. During the 1880's in Berlin, Albrecht Kossel isolated compounds from nuclein and named them **adenine, cytosine, guanine and thymine** (he received the Nobel prize for this). Phoebus Levene, from the Rockefeller Institute of Medical Research in New York City figured out that the nucleobases were linked to a phosphate and sugar, forming what he called a **nucleotide**. The phosphate and sugar part of nucleotides formed a 'backbone' joining together the nucleobases. The sugar he identified was deoxy-ribose and so the substance named by Miescher as nuclein came to be called Deoxyribose Nucleic Acid, or DNA.



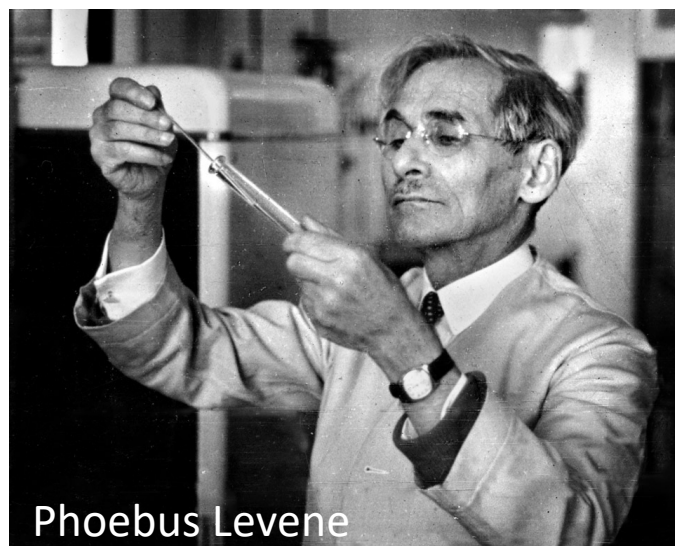
Friedrich Miescher

Wikimedia Commons, Public Domain



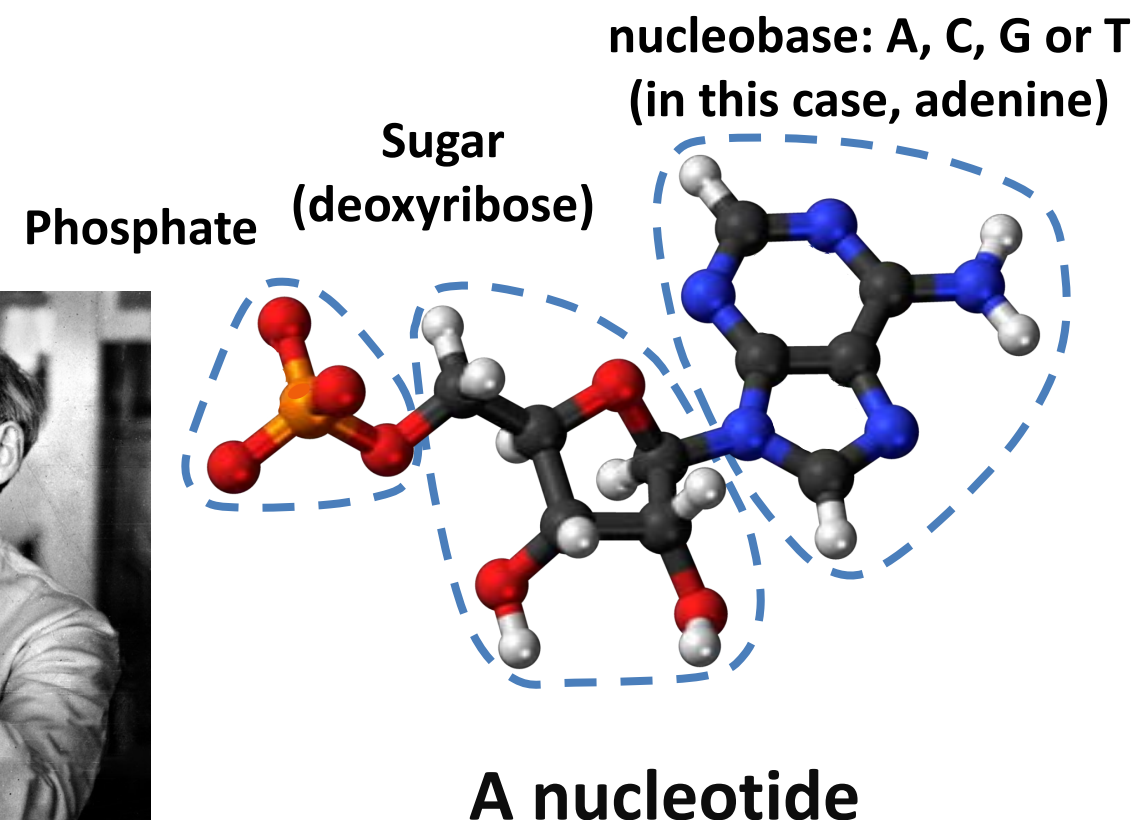
Albrecht Kossel

Library of Congress, Public Domain



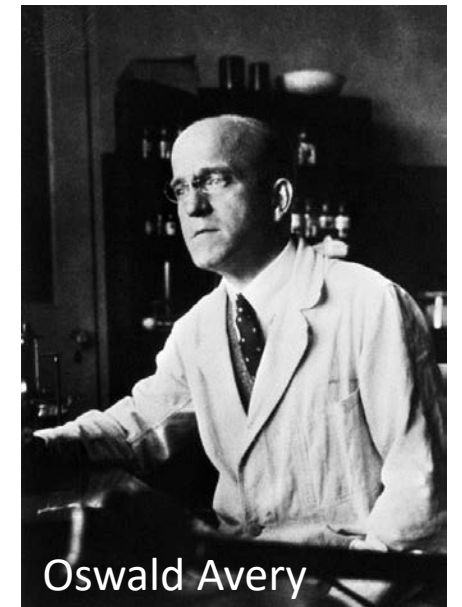
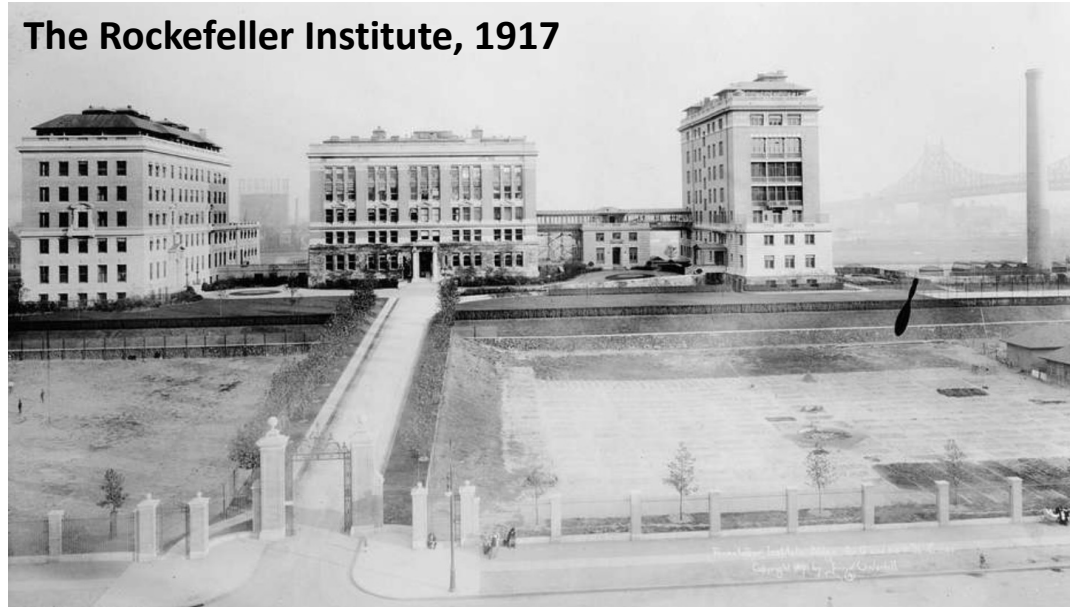
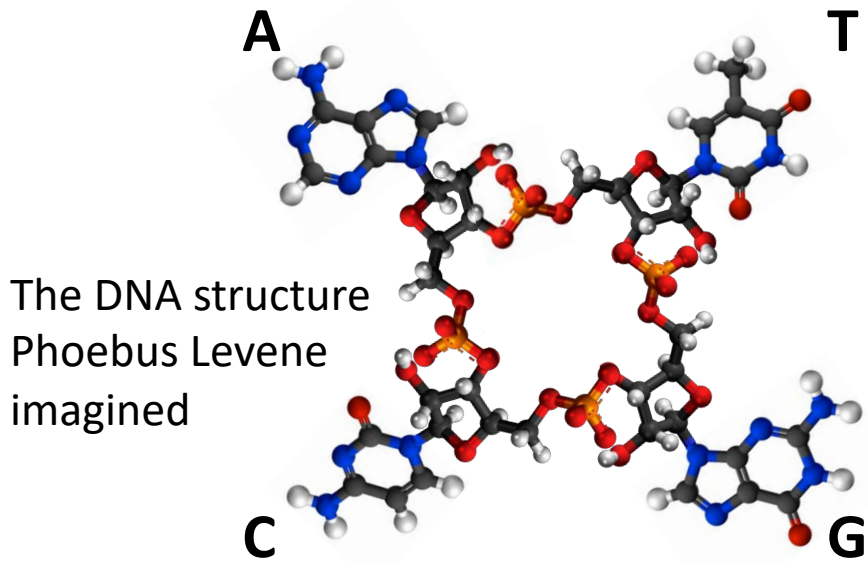
Phoebus Levene

Digital Commons @RU, The Rockefeller University



Could DNA be the material of heredity?

The work of Miescher, Kossel and Levene showed that DNA had an interesting chemistry, but proteins were still thought to be the basis of heredity. Levene thought the four nucleotide bases were arranged as a sequence, but in a ring structure that seemed far too simple to code for the complex genetic information needed for life (so proteins must instead be the genetic material).



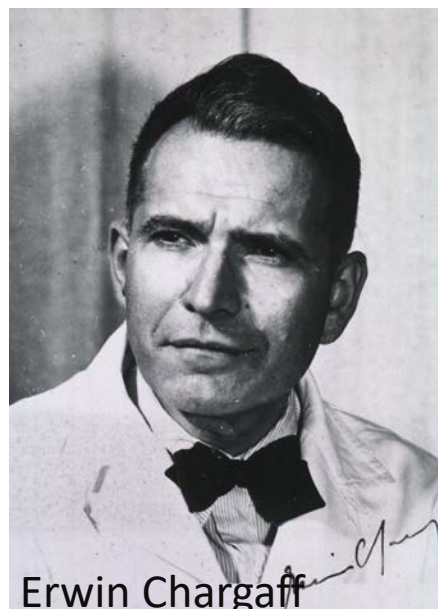
Library of Congress, Public Domain The Grainger Collection / Universal Images Group

Later at the Rockefeller Institute, in 1944, **Oswald Avery** figured out that **DNA was in fact the material in cells that contained genetic information**. This caught the attention of Erwin Chargaff, who was also in New York. He decided to investigate the percentage of A, C, G and T in different organisms. Although Levene's ring structure for DNA meant there should be equal amounts of A,C,G & T in DNA, Chargaff's results suggested something else.

Here are some of Chargaff's results

Organism	% A	% G	% C	% T
Chicken	28	22	22	28
Octopus	32	18	18	32
Wheat	27	23	23	27
bacteria	24	26	26	24

What pattern do you notice in Chargaff's data?



Erwin Chargaff

The race to figure out the structure of DNA

Although Chargaff found that, for any organism, the percentage of A = T and G = C, this information didn't seem very helpful for understanding how DNA functioned as the basis of heredity, allowing genetic information to pass from one generation to the next. Only a few years later did that become clear.

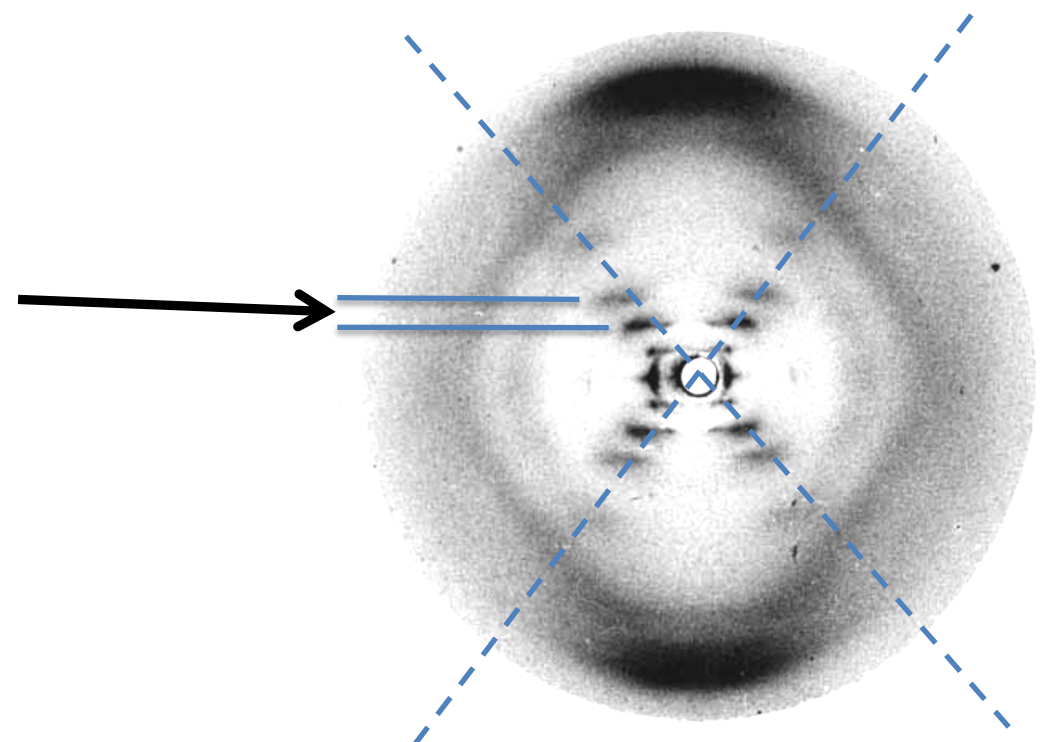
In the early 1950's a few scientists began to investigate the structure of DNA. In California, Linus Pauling, the world's leading chemist of complex molecules, was expected to be the first person to figure out DNA. Pauling was famous for identifying two structures in proteins, the **alpha-helix** and **beta-sheet**. You have already seen examples of these. **The alpha helix is a corkscrew shape that sequences of amino acids make** (as seen in the ribbon model of aquaporin on page 25). **The beta-sheet is made of parallel flat ribbons formed by amino acid chains** (as in the ribbon model of an antibody).

Pauling's main competitor at the time was Sir Lawrence Bragg, in Cambridge, England. You might recall that Bragg was the boss of Perutz and Kendrew, who you know worked on the protein structure of haemoglobin and myoglobin. A graduate student working with Perutz, Francis Crick, and a student working with John Kendrew, James Watson (an American), heard Pauling was interested in DNA and decided to get in on the action.

To figure out the structure of DNA, these groups needed to get their hands on good X-ray diffraction images of DNA (similar to those used to figure out protein structure, as mentioned earlier). Pauling had some not-very-good X-ray data and rushed to publish his DNA model, which turned out to be completely wrong. Watson and Crick obtained measurements of DNA from an excellent image that Rosalind Franklin and her co-worker, Raymond Gosling, produced while working with Maurice Wilkins at King's College, in London.

Franklin and Gosling's famous Photo 51 of DNA has been described as the most important photograph ever taken. To an experienced X-ray 'crystallographer' the X-shape here suggests that DNA has a helical shape (corkscrew-like). The spacing between the smudges indicates how far apart the bases are in DNA.

As a [classroom demonstration](#) experiment, your teacher might be able to show, with a laser pointer and a spring, how a diffraction pattern like Photo 51 can be produced (the video is also here: <https://tinyurl.com/ULC-Dif>).



Discovery of the Amazing Double Helix

Linus Pauling's incorrect model of DNA had three helical strands of sugar-phosphate backbone in the middle of the DNA, with the bases pointing outwards (a bit like Phoebus Levene's old model, but with a long corkscrew helix rather than just having each of the four nucleotides in a simple ring). However, Watson and Crick realized Pauling's model couldn't be right. The phosphates on the DNA backbone have a negative charge, so if you had three sugar-phosphate backbones wound tightly around each other, all the negative charges would repel (in the same way that opposite ends of magnets attract, but like ends repel). So the DNA would just fall apart. Another problem with Pauling's model was that it didn't explain Chargaff's strange result of how the number of A's and T's were always the same, as were the number of G's and C's.



Rosalind Franklin

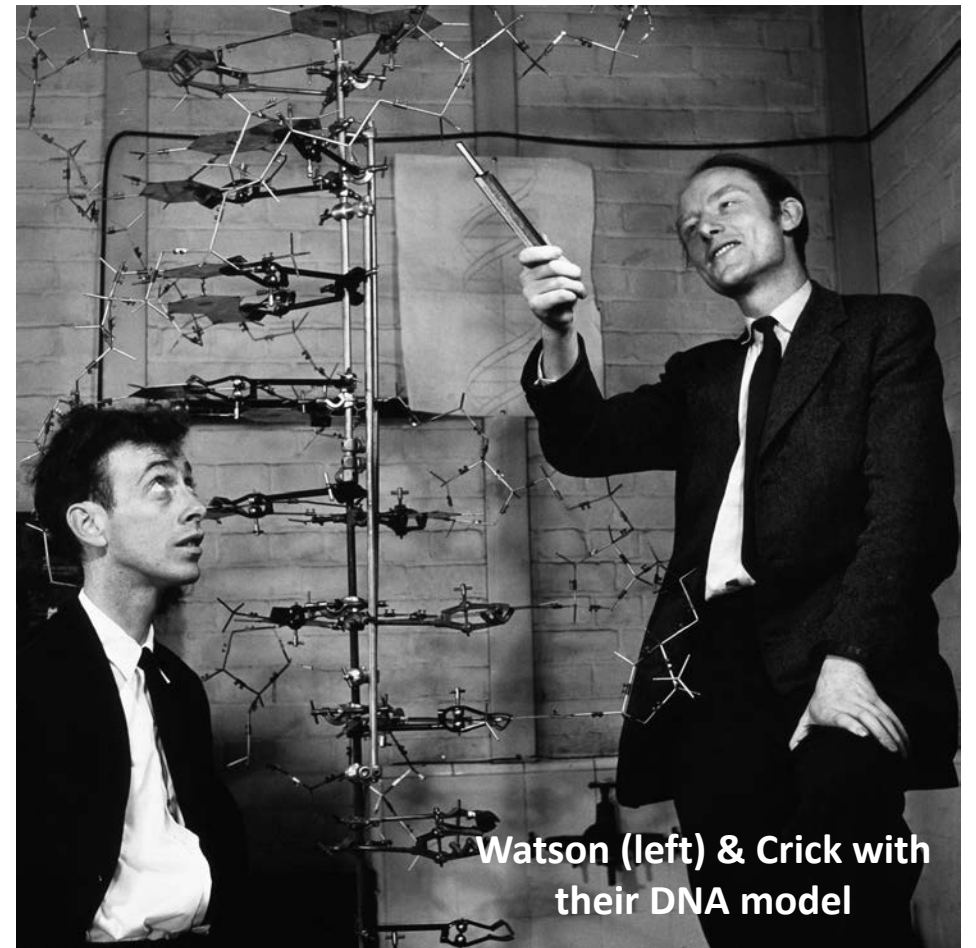
Elliott & Fry / National Portrait Gallery, London

Rosalind Franklin's calculations showed that, instead, the phosphate groups were on the outside of a double helix, and from photographs of other preparations of DNA, she figured out that the two strands of DNA forming the double helix ran in opposite directions to one another. Using this information, Watson and Crick started building models, but initially they couldn't figure out how the bases fitted in to the DNA structure. In 1952 they met with Chargaff and although they didn't enjoy each other's company, Chargaff told them all about his results.



Maurice Wilkins

National Library of Medicine

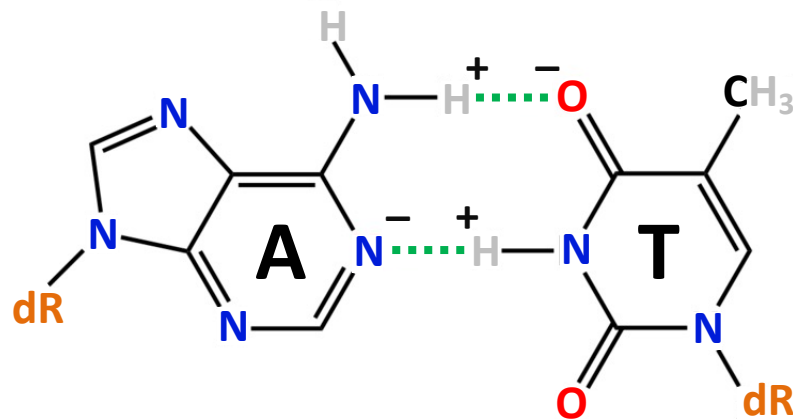
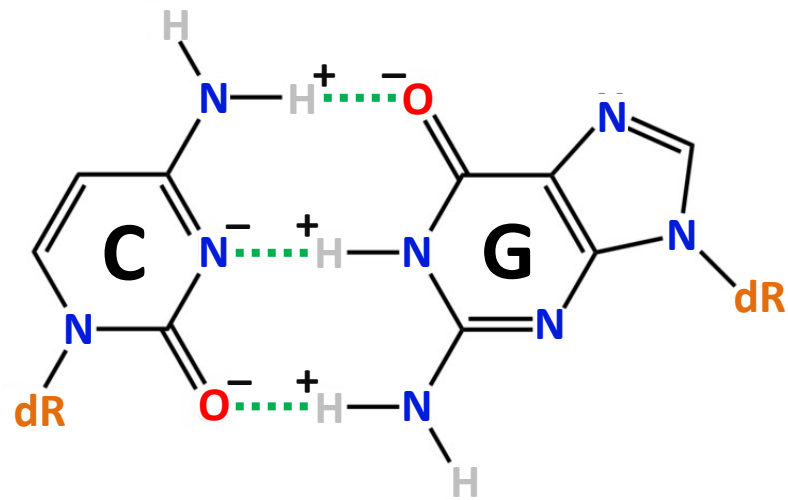


Watson (left) & Crick with their DNA model

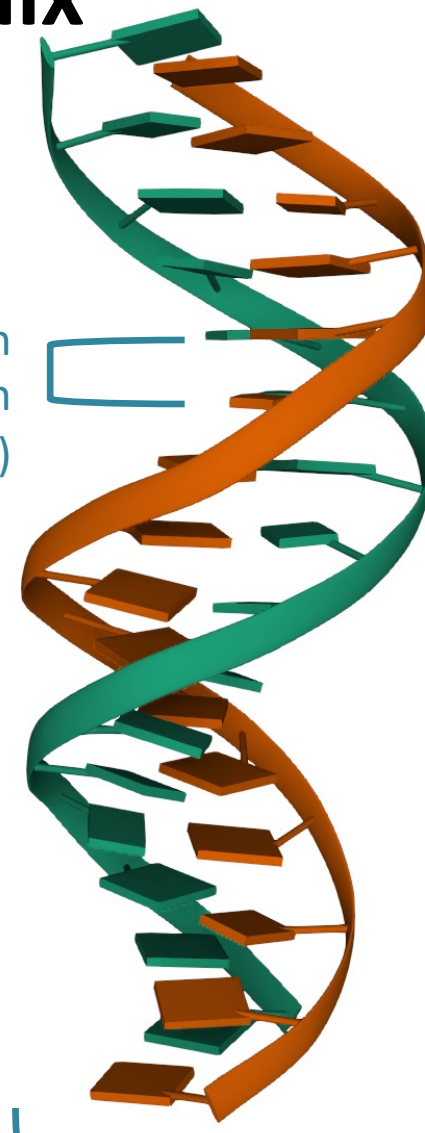
A. Barrington Brown, © Gonville & Caius College / Science Photo Library

Watson and Crick started making models with cardboard cutouts of the bases, then had a Eureka! moment. **Pairing an A (purine) and T (pyrimidine) produced a similar shape to pairing together a G (purine) and C (pyrimidine)**, as shown on the next page. **Each pair fitted nicely in the distance between the DNA backbones that Franklin had calculated.** Watson, Crick and Wilkins were awarded the Nobel Prize in 1962. Tragically, Franklin died in 1958 from ovarian cancer. Nobel Prizes are not awarded posthumously (after a person dies), or shared by more than three people, but many believe she deserved the Nobel prize too.

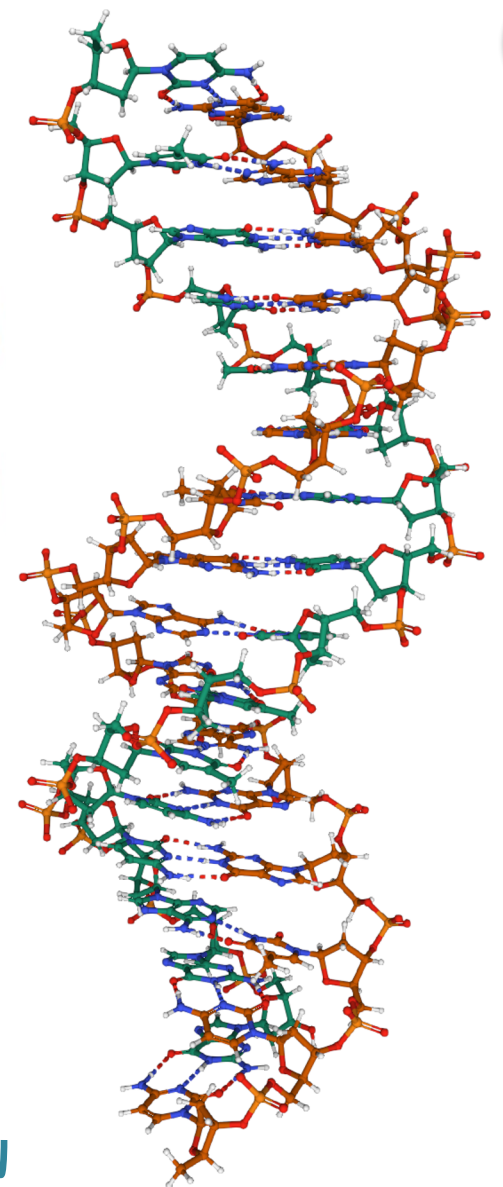
The Double Helix



The distance between each adjacent step on the ladder (base-pair) is 0.34 nanometres

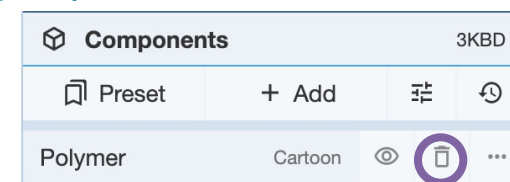


width = 2.0 nanometres (two-billionths of a metre)



As you can see here, when C pairs with G, or A pairs with T (consistent with Chargaff's rule), then the width across these two sets of base-pairs is the same. These base-pairs are stacked on top of each other, like steps on a ladder (with the ring structures laying flat, instead of as shown here), between the two strands of the DNA double helix. The strands are made of alternating phosphate (not shown) and deoxyribose (dR) groups. The two strands are held together by hydrogen bonds (green dashes) between the nucleobase pairs. This is a weaker bond than the single and double bonds shown by the solid lines. Hydrogen bonds are due to the attraction of positive and negative charges on some of the H, N & O atoms.

Go to this [DNA model](#) (or use <https://tinyurl.com/ULC-DNA>, or the QR code on page 40). This model is of a 16 base-pair long piece of double-stranded DNA. To see the DNA as a ball-and-stick or space filling representation, click the ... icon on the RCSB page.



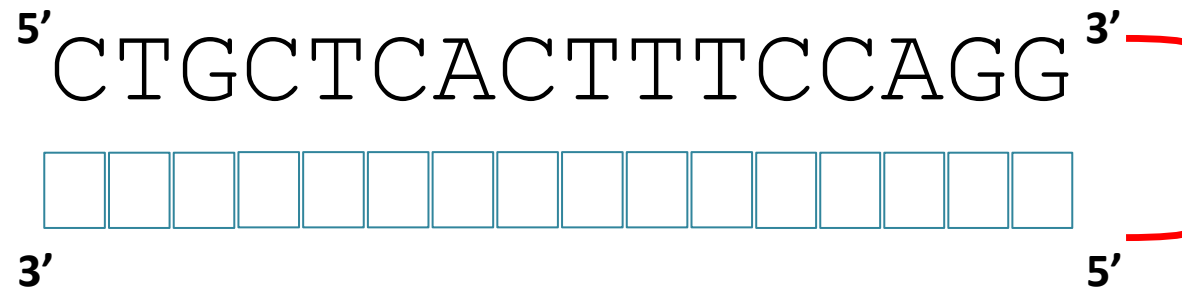
Select "Add representation". From the drop-down menu, choose a model you would like to see. You can go back to "Add representation" and click the trash icon to remove them.

How do cells make a new DNA copy before they divide?

Before the structure of DNA was figured out, one of the big mysteries of life was how a cell could make an accurate copy of its genetic instructions. After all, the trillions of cells that make up your body all originated from just one cell (the fertilized egg, or **zygote**) and each cell has a copy of your genome. Watson and Crick quickly saw that the structure of DNA provided a beautiful solution to this mystery. They concluded their famous 1953 paper with the comment:

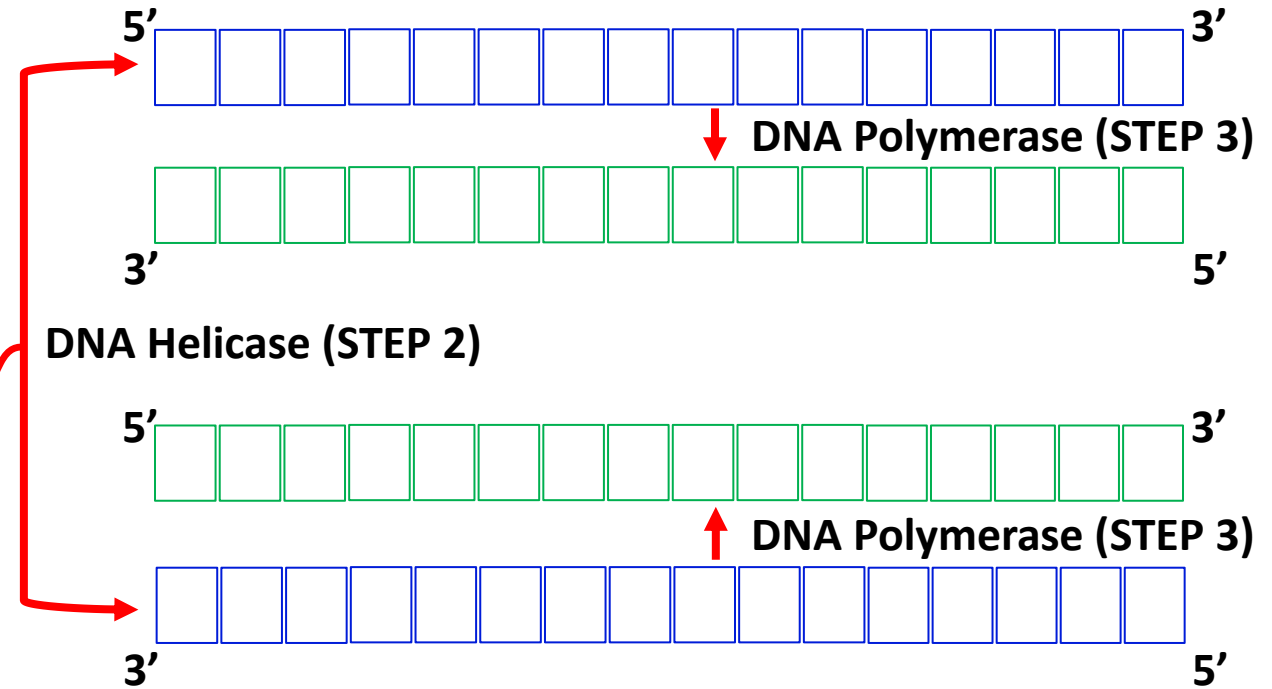
“It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material”

As an example of how this works, let’s imagine we are looking at just a short stretch of DNA in the genome, with the DNA sequence on one of the two DNA stands as given below.



STEP 1. Enter into the squares the sequence of the other strand of the DNA using the base pairing rule. Also, using a capital letter I, mark in either two or three dashes between each pair of nucleotide bases, to symbolize the correct number of hydrogen bonds they make.

You have now filled out the DNA sequences corresponding to the two strands of the double helix you explored on the previous page. The 5' and 3', (5-prime and 3-prime) is used to describe the direction of the DNA sugar-phosphate backbone, which goes in opposite directions on the two strands.



STEP 2. Cells make use of a protein called a DNA helicase. **What do you think it might do?** If you guessed that it must be an enzyme (ends with -ase) that unwinds the double helix, then very well done indeed! The helicase is able to separate the two DNA strands by breaking apart the hydrogen bonds holding together the DNA base-pairs. **Copy the sequence of each strand into the two rows of blue boxes.**

STEP 3. Cells make use of a protein called DNA polymerase. DNA polymerase uses the sequence information on one strand to synthesize a complementary copy *using the base pairing rules*. Fill in the newly synthesized sequence of each strand into the two rows of green boxes. **Note:** the DNA polymerase only works in the 5' to 3' direction, so that is the order you must add DNA bases here.

You have now made two identical copies of the original sequence. This is the process of DNA replication. Each copy would go to a different ‘daughter’ cell when a cell divides.

How does DNA fit into a cell and how are genes expressed? 36

Remember from the structure of DNA on Page 34 that there is a 0.34 nanometre gap between each of the DNA base-pairs? On Page 29 we mentioned that you have two copies of a genome made up of 3 billion base-pairs and made the surprising claim that each cell contained two metres of DNA. **Now you can verify that by multiplying 0.0000000034 metres x 6,000,000,000.** As mentioned on Page 29, histones tightly coil up the DNA so that it can fit into the nucleus of cells only a ten-millionth of a metre in diameter. **Take a look at the structure of a nucleosome (DNA coiled around histone proteins) - the first step of compacting DNA to package it into the nucleus** (see QR or <https://tinyurl.com/ULC-Nuc>).

The nucleosomes are themselves coiled into a tight bundle and this is further wound up to form a highly condensed fiber. **The complex of DNA and proteins together in the nucleus is termed chromatin.**

When the DNA is very tightly wound up it is not very accessible to RNA polymerase. **RNA polymerase is an enzyme that uses the DNA sequence of genes to make (transcribe) an RNA copy (called messenger RNA, or mRNA).** Once made, the mRNA moves from the nucleus to ribosomes, where it can be translated into protein.

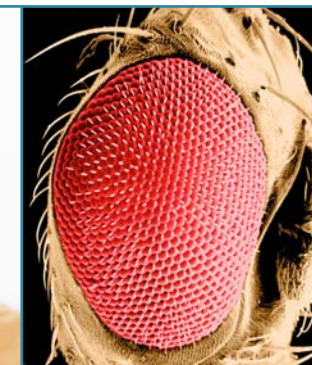
For each cell type, different regions of genomic DNA are opened up by methylating or acetylating the tail-end of histones (also see p.17). This process is central to epigenetics and is discussed more on p.53. **Watch a computer animation about epigenetics and chromatin modification.** (p.40 QR code or <https://tinyurl.com/ULC-Chromatin>).

One of the amazing things that research on DNA has revealed is that some genes (and the proteins they encode) have highly conserved roles, as you saw for myoglobin. This is especially true for genes encoding ‘transcription factors’, proteins that recognize short sequences of DNA near to genes. Transcription factors direct RNA polymerase activity to the genes necessary for a particular cell type to function.

For over 100 years, many genetic researchers have focused on the fruit fly, *Drosophila*. In the fly, scientists have found transcription factors that regulate eye and heart development and that tell cells whether they belong to the head or the tail. Even after more than 800 million years of evolution since our common ancestor (the urbilaterian), it seems that closely related genes control the development of insect and vertebrate hearts, eyes, heads and tails, too.



Sanjay Acharya, Wikimedia CC-BY-SA-4.0

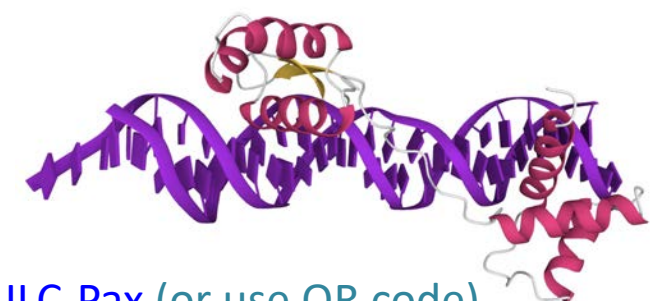


David Strutt, CC-BY-SA-4.0



Jean Beaufort, Public Domain Pictures CC-0

Although the fly and human eye are very different, their proper development requires the **Pax6** transcription factor. In the cartoon model below, alpha helices of **Pax6** bind to a DNA sequence.



<https://tinyurl.com/ULC-Pax> (or use QR code)

DNA sequencing and the genomic revolution

The 21st century is an exciting time to be a scientist. Biological and computational advances now allow protein [folding](https://tinyurl.com/ULC-Folding) predictions directly from DNA/RNA sequence information (<https://tinyurl.com/ULC-Folding>). These sorts of approaches have been critical in the race to produce vaccines against the virus responsible for COVID-19. Many important research discoveries are still to be made. These include figuring out ways to reduce human impacts on the environment, provide reliable crops to feed populations and produce cures for otherwise untreatable diseases. Genome sequencing is an essential tool in reaching these goals. For this, we must again thank Fred Sanger, who received his second (!) Nobel prize for developing an effective method of DNA sequencing. He shared that award with Walter Gilbert, who developed a different method.

This is Gilbert, reflecting on sequencing: “When I started it was one base a month, now its billions of bases in an afternoon. I find it unimaginable to have been a part of this entire effort”.

Today, it is possible to sequence a person’s genome for \$1000 (The first draft human genome cost \$300,000,000!). Also, portable DNA sequencing devices are now the size of a candy bar and plug into a laptop (or even cell phone).

What scientific question would you ask using genomics?



Svante Pääbo with Neandertal skull. Photo: Frank Vinken



David Eccles (Gringer) CC-BY-4.0

Genomics has helped understand the origin of modern man, our relationship with ancient related species, such as Neandertals, and the history of human migrations, for example as people discovered and settled across the Pacific.

A major effort, the Earth Biogenome Project, is underway to sequence the genomes of all known species. The [Darwin Tree of Life Project](https://tinyurl.com/ULC-Dar) (<https://tinyurl.com/ULC-Dar>) is one part of this, with researchers from elsewhere around the world also using genomics to help in species [conservation](https://tinyurl.com/ULC-Koala) (QR code or <https://tinyurl.com/ULC-Koala>). Genomics is helping us understand how species evolved, a question that has fascinated biologists since Charles Darwin developed his theory of evolution through natural selection.

“There is grandeur in this view of life, with its several powers, having been originally breathed into a few forms or into one; and that, whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved. (Darwin 1859; *The Origin of Species*).

CROSSWORD
1

1		2				3		4					5			6			7	8	
												9									
		10		11						12											
														13		14					
15									16				17								18
														19							
				20	21		22														
														23							
24																					
												25							26		
27				28																	
							29		30				31			32		33			
		34																			
35											36									37	
													38						39		
						40															
		41																		42	43
44					45						46			47		48					
														49							
50				51			52						53								
54			55			56								57					58		
					59				60	61											
62														63						64	

See how much of this crossword you can complete from memory, then use the page references to fill in the rest.

Across

1. The basic unit of a triplet code used for protein synthesis. p.7
4. The building blocks that make up nucleobases and amino acids. p.4
7. A digital message encoded by the binary 01010100 01011000 01010100. p.21
9. The shape of deoxyribonucleic acid (6,5). p.34
10. The person who identified that nucleic acid was the genetic material, at a time when most people thought the genetic material would be protein (6,5). p.31
13. A water transport protein with a tightly corkscrewed secondary structure. p.26
15. Name given to a person who identifies protein structures using X-rays. p.22
19. Tightly regulated sugar found in the blood. p.15
20. For ancient Egyptians, the greatest god, from which all creation unfurled. p.2
22. A property of some amino acids that causes them to be hydrophobic. p.12
23. An atom that typically forms three bonds in DNA nucleobases. p.5
24. Name given to molecules having a ring-like structure. p.12
25. CH₃ group that is attached to histone tails to regulate chromatin structure and gene expression. p.17
26. A paired-domain homeobox gene that has an important role in eye development throughout animal evolution. p.36
27. An amino acid that is required for the proper function of insulin. p.19
28. Negatively charged acidic amino acid. p.12
29. The photographic image that results when a crystal is bombarded with X-rays (11,7). p.22
35. Positively-charged (basic) proteins that interact strongly with the negatively-charged DNA sugar-phosphate backbone, coiling DNA up to form nucleosomes. p.36
36. A small polar amino acid that can have phosphorus added to it, making a protein active or inactive. p.17
38. A useful type of molecular model that highlights the bonding between atoms (4,3,5). p.5
40. With 37 down, a type of molecular model that represents what DNA or protein looks like at the atomic level. p.5
42. A general term used to describe a young person with an X and Y chromosome.
44. High-frequency electromagnetic radiation, with a wavelength of 0.01 to 10 nanometres, used to study protein crystals. Also used by radiologists to look at broken bones. p.22
45. Pyrimidine nucleobase containing a single oxygen atom. p.5
49. Cyclopeptide produced in the hypothalamus, with hormonal functions regulating reproduction and maternal behaviours. p.14
50. Female scientist who made major contributions to understanding DNA structure (8,8). p.33
54. A word for the forces that hold atoms together in a molecule. p.5
56. The four DNA bases. p.4
57. Light-sensitive sensory organs on the head. p.36
58. A bulky non-polar amino acid that is often found on the inside of proteins, avoiding interactions with water molecules. p.12
60. A strong fibrous structural protein that forms the basis for rhinoceros horns, bird feathers and cat hairballs. p.27
62. Scientist working in New York who found that in the DNA of all organisms he tested, the number of A's and T's was always the same, as was the number of G's and C's (5,8). p.31
63. A small iron-containing compound that binds oxygen and sits deep in a pocket within myoglobin and haemoglobin proteins. p.22
64. Marine mammals closely related to dolphins. p.24

Down

1. Group of marine mammals known for their breath-holding ability. p.24
2. The version of a sugar molecule found in the sugar-phosphate backbone of DNA. p.13
3. A nonpolar amino acid with a C₃H₇ R-group. p.12
4. The smallest of the non-polar amino acids. p.12
5. Term for a social group of dolphins or whales. p.23
6. The clade (animals with a common ancestor) that includes hippopotamus and whales. p.23
7. The largest known gene and protein in humans, important for muscle elasticity. p.27
8. The sex chromosome pair typical of human females. p.28
11. The common name for dihydrogen monoxide (two hydrogens and one oxygen). p.5
12. Enzyme that copies a gene from DNA into RNA, so that the RNA can be translated into protein by the ribosomes (3,10). p.36
13. A positively-charged amino acid with four nitrogen atoms. p.12
14. An RNA codon for phenylalanine (actually the first codon to be figured out, in 1961). p.13
16. Sections of DNA in the genome that code for proteins. You have about 20,000 of these. p.28
17. Structure at the anterior end of an organism. The opposite of the posterior (tail) end. p.36
18. The smallest atomic element, comprised of just one electron and one proton. This element participates in bonds that hold together DNA base-pairs. p.34
21. Social insect that has a liking for sugary substances. p.16
22. The chemical formula for common table salt. p.22
24. A frequent secondary structure in proteins (5,5). p.32
28. The term given for the entire DNA instructions of a living organism. p.28
30. A British biophysicist famous for being a co-discoverer of the structure of DNA (7,5). p.32
31. An atomic element capable of forming four covalent bonds, allowing it to have a central role in building the molecules of life. p.5
32. The nickname of the UGA stop codon (which would be TGA in the DNA sequence). p.8
33. Photosynthetic organisms (capturing energy from the sun to grow), that branched off from animals early in eukaryote evolution. p.1
34. Abbreviated name for an acidic (negatively charged) amino acid (also a sort of snake). p.12
37. With 40 across, a type of molecular model that represents what DNA or protein looks like at the atomic level. p.5
39. The distinguishing feature that makes amino acids different from each other (4,5). p.11
40. Feature of the environment that is usually light blue because of the way light of different wavelengths is scattered by N₂ molecules in the atmosphere. p.4
41. American scientist famous for being a co-discoverer of the structure of DNA. p.32
43. Word for people who have difficulty controlling their blood sugar (glucose) levels. p.16
46. The number of atoms in the ring structures of purine nucleobases, adenine & guanine. p.7
47. A type of mutation that results in a stop codon at a position in a gene that would normally have coded for an amino acid. p.20
48. The sulfur-containing amino acid capable of forming disulfide bonds that link different protein chains together, or cause a single chain to form a stable loop on itself. p.17
50. A simple 5-carbon sugar that is part of the backbone of RNA. In its deoxy-form, this sugar is part of the DNA backbone. p.13
51. Three-letter abbreviation for the smallest of the basic (positively-charged) amino acids. p.12
52. The biological macromolecule (macro = big) that contains the instructions for life. p.1
53. The position you would normally find a stop codon in a gene.
55. A dipeptide encoded by the sequence GACGCA. Also, the abbreviated title of a lawyer (e.g. on a US television show) who might use forensic DNA evidence in a courtroom. p.6
59. The three-letter abbreviation for an unusual amino acid containing a selenium atom. p.10
61. Name given to a long period in Earth's history. p.25

QR codes to scan for biochemical models and video links

Met-Enkephalin



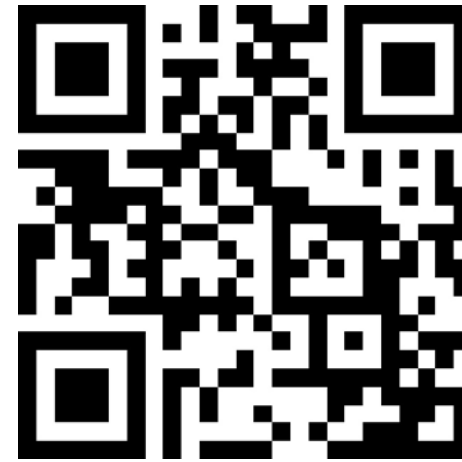
Somatostatin



Oxytocin



Insulin



DNA data storage



Myoglobin



Protein Structure



Diffraction



DNA



Nucleosome



Chromatin



Pax6



Folding



Darwin Tree of Life



Koala



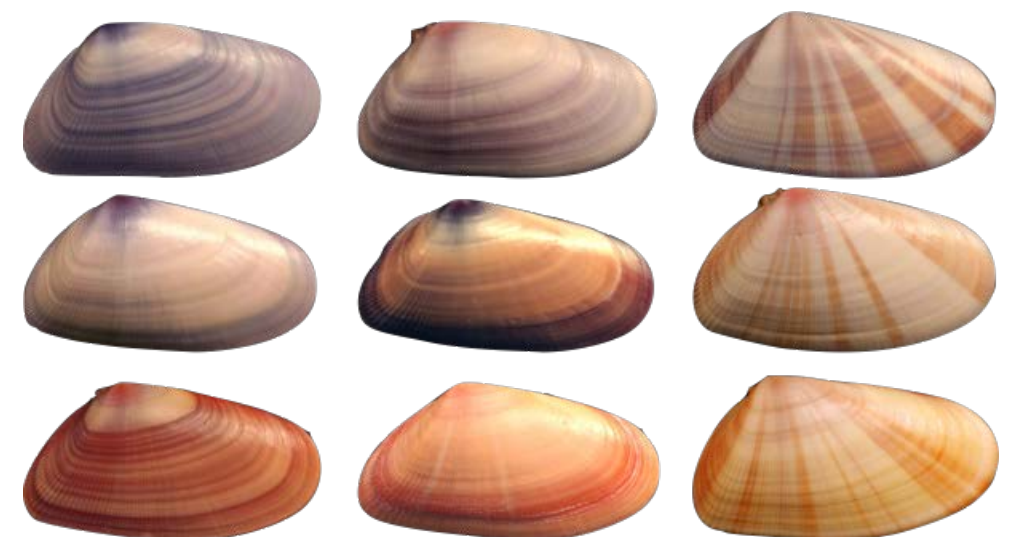
Genotypes, Alleles and Phenotypes

So, you now have a good understanding of how the sequence of DNA base-pairs in a gene determine the order of amino acids in a protein. You learned how variations in the DNA sequence of a gene can sometimes result in harmful disease (such as when the insulin gene carries a mutation, resulting in Type I Diabetes). But you also saw how DNA variations can be beneficial to an organism (for example, changes in the myoglobin gene allowed the ancestors of modern whales to dive deeper for food). You also know that in our genome, we have around 20,000 different genes, which code for proteins with a huge range of different structures and activities. These help the trillions of cells in our bodies grow, divide (replicate) and perform their essential functions.

Knowing all this, we can introduce you to a few more scientific terms. The term **genotype** has two meanings. It can refer to the complete set of genetic instructions provided by the genome. **Genotype** is also used to refer to the particular versions of a specific gene in an individual organism. Remember from page 28 that we have two copies of each chromosome, one from each parent. These two copies of each chromosome carry slightly different versions of the hundreds (or thousands) of genes they encode. This is because, in a normal population of organisms (or people), there are many small (single base-pair) DNA variations in each gene. These variants of each gene are called **alleles**.

The word allele comes from the Greek root word *allos*, meaning *other, or different*. You might recognize *allos* in a different situation. If someone has an allergy, or allergic reaction, it is because their body's immune system is reacting strongly to something that is *different* from them, such as peanuts, bee-stings, pollen (hay-fever) or pets. The important thing to learn here is that **allele means a different version of the same gene**.

Phenotype is another really important term in genetics (it comes from the Ancient Greek root word, *phaínō* (φαίνω), meaning "to shine, to show, or to appear"). This is the word used for what we can observe about an organism. This might be the organism's colour, shape, size, how it develops, what is inside it (for example, blood type), or even the way it behaves. For example, some cats are very friendly and some are timid. So, social behaviour in cats (or other animals) is also a **phenotype**.



Debivort, Wikipedia CC BY-SA 3.0 (adapted)

Phenotypic variation in the colour pattern of shells from different individuals of the mollusc, *Donax variabilis*.

Characteristics and Traits

The **phenotype** of an organism is therefore based on many different **characteristics**. For example, easily observed human characteristics include height, body shape, hair colour, skin colour, eye colour, ear lobe shape, foot width, hand span and thumb flexibility.

Geneticists use the term *trait* to indicate how a characteristic appears in a particular individual. For example, for the eye colour characteristic, you might have the trait of brown eyes. For the ear lobe shape characteristic, you might have the trait of detached ear lobes and for thumb flexibility, you might have the trait of hitchhiker’s thumb and be able to bend it back 90 degrees.

Traits can also be microscopic or molecular. One fascinating example is whether a person continues to express the lactase gene into adulthood. The characteristic of **lactase persistence** determines whether someone exhibits the trait of lactose tolerance or intolerance (**QR code on page 56**, or <https://tinyurl.com/ULC-Lactase>).

Another example, is the **Sickle Cell Trait**, which occurs because of a DNA variation in one copy (allele) of a haemoglobin gene, causing a person’s red blood cells to change shape when deoxygenated. Remember how diving mammals evolved a version of muscle myoglobin that doesn’t clump together at high concentrations? Well, in each red blood cell there is over 200 million haemoglobin molecules. Normally these don’t clump together, even at this high concentration. However, the haemoglobin allele (called *HbS*) in Sickle cell trait codes for a haemoglobin protein with a Glu6Val (E6V) change that causes it to clump into rods, distorting the red blood cells into a sickle shape.



A hand sickle



sickle cell



normal red blood cell

For most people with sickle cell trait, this generally doesn’t affect their health much. However, it can be a big problem if someone inherits two copies of the *HbS* allele (one from each parent). This causes **sickle cell disease**, which can be very painful and harmful to different organs of the body, causing strokes and other complications. (or access video at <https://tinyurl.com/ULC-Sickle> or via the QR code).

This also raises an important question for scientists. If the *HbS* allele can actually be harmful, why are there so many people who have it? Wouldn’t evolution have caused that allele to be lost from the population? **Have a think about this, then watch this video** to find out the **surprising answer** to this question (<https://tinyurl.com/ULC-HbS>).

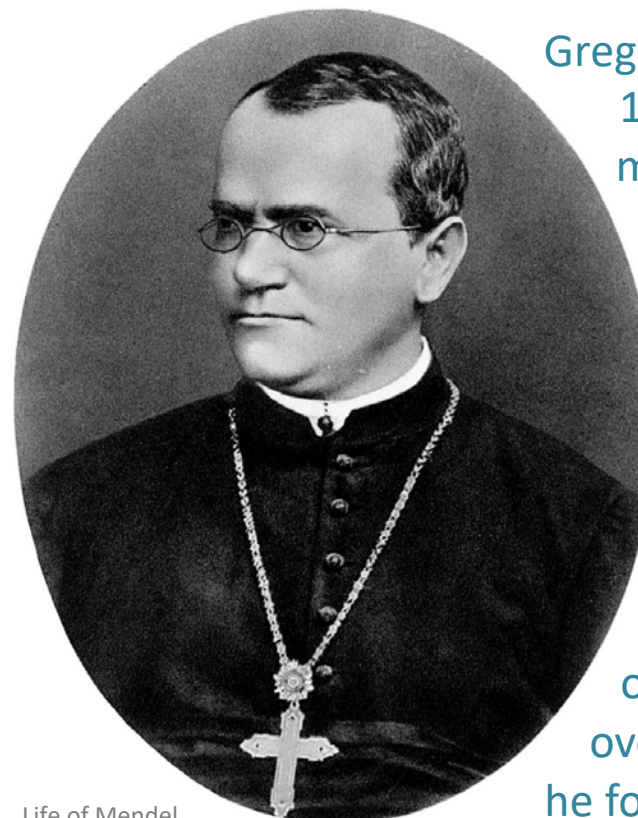
Use the space below to type down the reason the *HbS* allele is so common in some regions of the world.

Mendelian Genetics

Some phenotypic characteristics, traits and diseases are due to allelic variations in single genes (**monogenic**) or a smallish number of genes (**oligogenic**), but most traits are caused by the interactions of many genes and are **polygenic**. **Have a close listen to this video on genetic variation** (<https://tinyurl.com/ULC-variation>, or QR code). It gives examples of traits and diseases affected by a single or few genes, or by many genes. **See if you can fill out the table below, giving nine examples from each category.**

Mono/oligogenic traits/disease	Polygenic traits/diseases
1.	1.
2.	2.
3.	3.
4.	4.
5.	5.
6.	6.
7.	7.
8.	8.
9.	9.

Most phenotypes of complex organisms, including plants animals and fungi, depend on the activities of many genes. However, as you can see from your list above, there are some cases where a single gene can have a major effect on a phenotype. **Genetics research was born in 1856 when Gregor Mendel began his work, identifying the patterns of heredity when phenotypes are controlled by powerful single-gene activities.**



Life of Mendel, by Hugo Iltis (CC BY-NC 4.0)

Gregor Mendel, who was born in 1822 and died in 1884, was a monk and scientist in the city of Brno (in what is now the Czech Republic). In 1856 he started a decade-long series of experiments on the inheritance patterns of traits in the garden pea *Pisum sativum*. By careful cross-fertilization and study of the traits of 30,000 plants, over multiple pea generations, he found that when crossing two parent plants with different traits for a particular characteristic, the progeny (the child) **was not** an average of the phenotype of the parents. Instead, Mendel found that the trait of one of the parent plants dominated, whereas the trait of the other parent receded (went away). So, he called these **dominant** and **recessive** traits. In other words, the progeny all looked like only one of the parent plants, not a mix of both. Geneticists refer to this first generation of progeny as the **F1 generation** (F stands for filial, which means son or daughter).

Significantly, when he took these progeny and bred them together, the recessive trait re-appeared in about one out of every four of *their* progeny (which are referred to as the **F2 generation**). Mendel's strange results were published in 1866 in an obscure journal. Around 1900, 16 years after he died, his research was re-discovered and recognized as being important.

Mendel's crosses and Punnett's squares

This table shows some of Mendel's results on the inheritance of pea traits

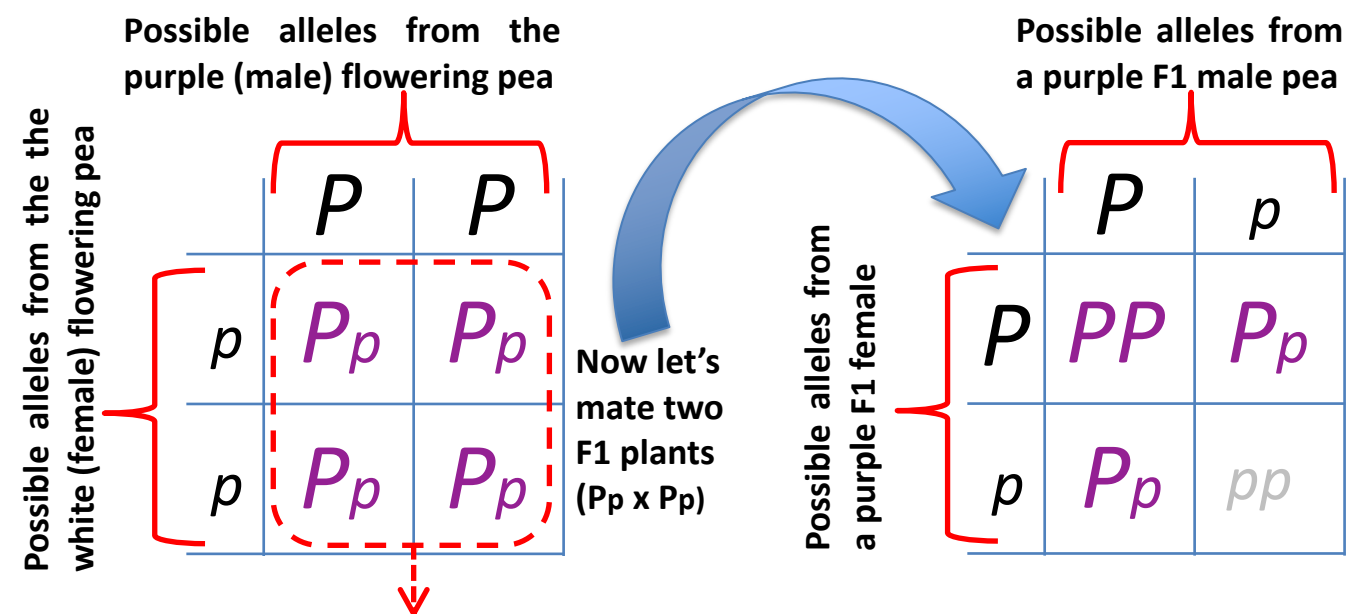
Characteristic	Parent traits	F1 generation traits	F2 generation traits	Ratio of F2 traits
Flower colour	Purple or white	All were purple	705 purple and 224 white flowered plants	3.15 to 1
Plant height	Tall or dwarf	All were tall	787 tall and 277 dwarf plants	2.84 to 1
Seed texture	Round or wrinkled	All were round	5474 round and 1850 wrinkled seeded plants	2.96 to 1
Seed colour	Yellow or green	All were yellow	6022 yellow and 2001 green seeded plants	3.01 to 1

As seen in the table, even though all F1 plants showed the dominant trait, when these were bred to produce the F2 generation a different pattern appeared. In fact, for every three F2 generation progeny with the dominant trait there would be one F2 plant with the recessive trait.

A useful way to understand this pattern is by using what is called a Punnett square, named after a British geneticist, Reginald Punnett, who invented the method in 1905. Let's use a Punnett square to look at the genetics of flower colour in peas. Because there is only one gene we need to think about, producing either purple or white flowers, our Punnett square is for what is called a monohybrid cross.

We know from Mendel's data on the F1 generation traits that the version of the gene (the allele) that gives purple flowers is dominant, so we represent that allele with a capital *P*. The allele that gives white flowers is represented by a lower case *p*. Remember how we have two copies of each chromosome (and the genes on them), one from each parent? So do peas.

The parent peas Mendel used were **homozygous** for the flower colour gene. That is, each parent had two identical alleles: so *PP* for the purple flowering parent and *pp* for the white flowering parent. Only one allele from each parent is randomly chosen to go into the sperm or egg (these are termed **gametes**) used in fertilization to produce an F1 progeny. The Punnett square is a way to show the possible allele combinations that could occur.



Notice how the F1 progeny must all be the same. They are **heterozygous** but all would produce purple flowers because they have one copy of the dominant allele

Notice that in the F2 generation, there is a 1:2:1 (*PP*:*Pp*:*pp*) **genotypic ratio** but a 3:1 **phenotypic ratio** of purple to white flowering plants

Punnett square practice and Mendelian predictions



Perhaps you don't like peas, but everyone loves puppies! Watch this Punnett squares tutorial, walking you through the inheritance pattern of [Labrador coat colour](https://tinyurl.com/ULC-labcoat). (<https://tinyurl.com/ULC-labcoat>)



Test your understanding by using the Punnett square below. Can you predict the genotypic and phenotypic ratios of the progeny (F1 plants) that would result when crossing a *heterozygous* tall female pea plant with a dwarf male? What letter will you choose to represent the alleles? By convention (general rule) the allele possibilities from the male parent are placed along the top row and those of the female parent are placed down the side.

Genotypic ratio:

Phenotypic ratio:

As you know from page 42, high levels of malaria in parts of Africa have resulted in the **HbS** allele (**s**) being common there. One copy of the *HbS* allele in heterozygous carriers (**Hb/HbS**, or **Ss**) of the sickle cell trait helps protect those people from malaria by interfering with the life cycle of the parasite transmitted by mosquitos. However, having two copies of the *HbS* allele causes sickle cell disease and, without access to good medical care, many children with the disease die before the age of five.



Mike Blyth, Wikimedia (CC-BY-3.0)

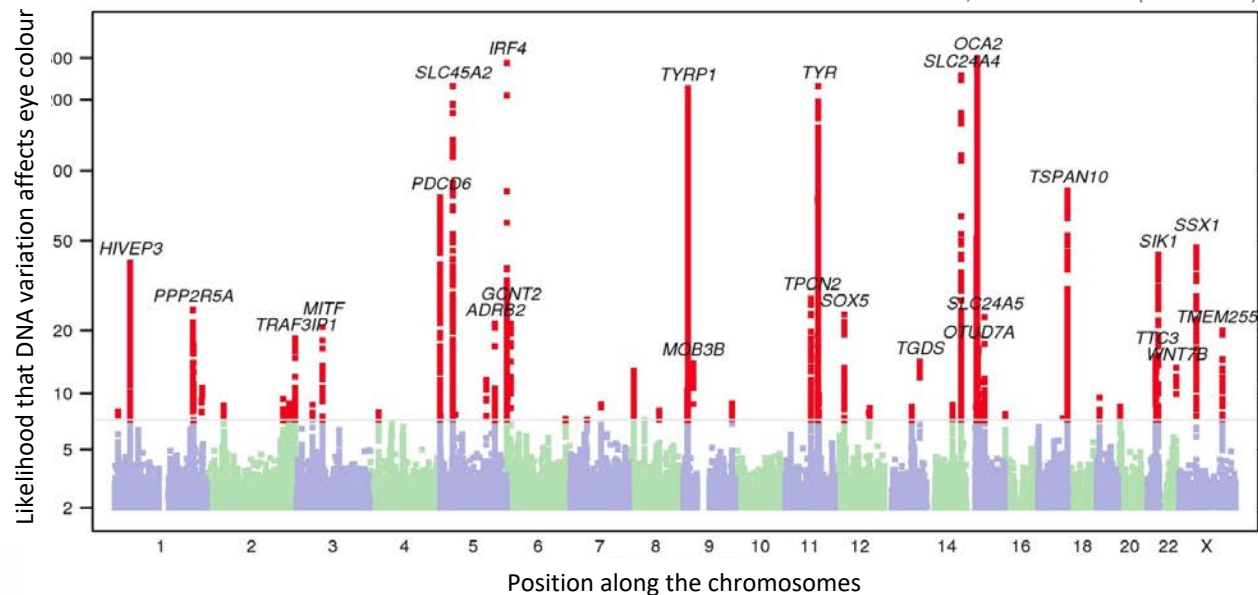
Sickle Cell Disease in Nigerian Children

Imagine you are a doctor working in Nigeria and want to figure out how many children with sickle cell disease the country's health care system needs to care for. In Nigeria, one quarter of the population are heterozygous carriers of the sickle cell trait. 7.6 million babies were born in Nigeria in 2020. Can you estimate how many of those births are to parents who both are *HbS* heterozygotes and, using the Punnett square, from that **calculate how many babies with a risk of sickle cell disease are born each year in Nigeria. Also, suggest steps the Nigerian government could take to reduce this disease burden.**

Mendelian mythbusting

Because you have a good understanding of the complexity of genomic DNA and how it codes for 20,000 or more proteins, we can let you in on a secret that now might not come as much of a surprise. In the teaching of Mendelian genetics, **school students have often been given examples of human traits incorrectly described as monogenic (determined by a single gene)**. Commonly taught examples are eye colour, earlobe attachment, tongue rolling and hitchhikers thumb (distal hyperextensibility of the thumb). But in actual fact, these traits are influenced by the activity of many genes.

Simcoe *et al.*, *Sci. Adv.* 2021 (CC-BY-4.0)



This graph, called a Manhattan Plot (**can you guess why?**), shows data from a recent genetics study looking at the eye colour of over 150,000 Europeans. The researchers examined how DNA sequence variations across the genome correlated with eye colour. This is called a Genome-Wide Association Study, or **GWAS**. The X-axis shows the position of the DNA sequence variants along the chromosomes. The Y-axis shows how strongly the data links DNA variation at that position to eye colour. 25 genes, scattered across the genome, seem particularly important.

One of the few human characteristics that seems to have a monogenic basis is the type of ear wax a person produces. Maybe it is not surprising that this is seldom explored in classroom experiments! Ear wax (the medical term is cerumen) is secreted from glands in the ear and helps to carry dead cells, bacteria and debris (such as dust) out of the ear. There are two different forms of ear wax in human populations: wet ear wax, which is yellow-brown and moist, and dry ear wax, which is crumbly and grayish.

The type of ear wax a person produces depends on a single DNA base-pair variation in the gene for the transporter protein ABCC11. Most humans have the dominant ABCC11 allele, but it is thought that 30,000 to 40,000 years ago, in what is now Siberia, a mutation occurred resulting in a Gly180Arg (G180R) change in the ABCC11 protein. This stops the protein folding properly and reduces the secretory function of the cerumen glands. Today, the recessive ABCC11 allele is found mainly in Asian populations and native peoples of the Americas (whose ancestors migrated from Asia, across the Bering land-bridge, sometime in the last 15,000 years).

Even in this simple example of a single gene affecting a human trait, it is important to understand that ABCC11 is **not** a gene specifically **for** ear wax type. [ABCC11](https://tinyurl.com/ULC-ABCC11) is also used by many other parts of the body, including in other secretory glands, such as those that produce armpit sweat and the antibody-rich breast milk (colostrum) mothers produce in the early post-natal period (the video can also be accessed via QR or <https://tinyurl.com/ULC-ABCC11>).

Polygenic traits and diseases

You have already seen that eye colour is influenced by at least 25 genes, although some (like OCA2) have a bigger impact than others. Other characteristics like height, body shape, skin colour, personality and intelligence, or the risk of developing heart disease, arthritis, type 2 diabetes (T2D), cancer, Alzheimer's disease, schizophrenia or autism may be affected by hundreds, or even thousands, of genes.

In these cases, each gene contributes only a tiny amount to a **polygenic trait or disease risk**. However, adding up the small influences of all these genes gives a **polygenic score** (<https://tinyurl.com/ULC-Polygenic>). Across a population, the polygenic scores will form a distribution, and most people will have close to an average height, intelligence or disease risk, but some individuals will have scores that put them nearer the higher or lower end of the typical range.

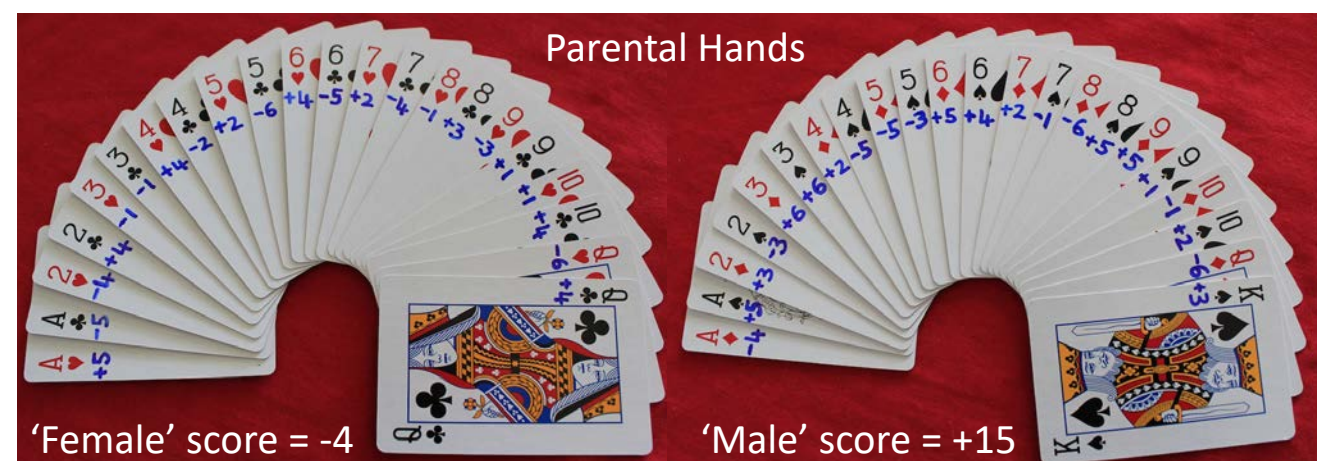
To give an insight on the basis of your own polygenic traits and disease risks, you are going to take part in a class activity, in which each pair of students uses a deck of cards, a dice, a coin and a felt-tip marker pen. **You will record your data on page 49 and graph the class data on page 50. If in lockdown, you can do this activity singly. In which case, find a deck of cards you are allowed to write on.**

STEP 1. From the deck of cards, remove any Jokers, all the Jacks, the Queen of Spades and the Kings of Hearts, Diamonds and Clubs. *You won't need any of these cards.*

STEP 2. Give the remaining Spades and Diamonds to one person of the pair, and the Hearts and Clubs to the other. The cards (hand) each person holds symbolizes their chromosomes (instead of the 23 pairs we really have, here

we will just use 11 pairs (2 cards each for Ace to 10, then **either** two Queens **or** a King and Queen, as the sex chromosomes). At this step, the red cards in each hand represent chromosomes inherited from the person's mother, black cards from their father. In this activity, the person with the Diamonds and Spades holds the male hand (they have a Queen and King as the sex chromosomes, symbolizing our X and Y chromosomes). The person with Hearts and Clubs has the female hand (two Queens, or XX).

STEP 3. For every card in the two hands throw a six sided dice. With the marker pen, write that number on the card. Then toss the coin. If it is a tail, put a negative sign next to the number on the card. If it is a head, put a positive sign beside the number. These numbers represent the contribution of that chromosome copy to the overall score for the polygenic trait or disease risk. In reality, the chromosome might carry hundreds of DNA variations (SNPs, see the genetic variation video on page 43) that sum up to positively or negatively affect the trait.



STEP 4. Using the numbers you wrote on the cards, add up the score for each hand. We will call these the parental polygenic scores.

Class activity: Polygenic inheritance of height

We are going to imagine that these polygenic scores are for height and for simplicity we will pretend the polygenic scores equal the number of centimetres a person is taller or shorter than the national adult male or female average. In New Zealand the average adult male height is 178cm and the average adult female height is 165cm. Therefore, the person holding the male hand shown on the previous page would have an adult height of 193cm (178 + 15) and the female would be 161cm in height (165 - 4). **What height(s) do you expect their children to reach when they become adults?** To answer this question, you and your partner are going to produce ten children.....

Like other cells of our body, the single celled human embryo (the **zygote**) contains 46 chromosomes – 22 *pairs* and either two X chromosomes or an XY pair. However, prior to fertilization, the egg and the sperm (the **gametes**) only have 23 chromosomes each. This is because sperm and egg cells are produced through a process of cell division known as **meiosis**. The main thing to know about meiosis is that in making a sperm or egg, one chromosome copy from each pair of chromosomes is chosen **at random**. This is known as Mendel's law of **independent segregation**. This contributes to variation in the offspring, in contrast to the asexual reproduction of many plants and fungi that use spores instead of gametes.

STEP 5. To mimic independent segregation, shuffle your hand then turn over cards one by one, sorting them into two gametes. If a card is for a chromosome you haven't already turned over, put it to the side, face up (= **Gamete 1 pile**). If you already have that chromosome number, place the card back on the bottom of the deck.

Put only the first picture card (King or Queen) into the Gamete 1 pile. When you are finished, Gamete 1 should have eleven cards. The eleven remaining cards are **Gamete 2**. Of course, if you had the parent male hand, your gamete hands represent sperm and if you had the female hand, your gamete hands represent eggs.

STEP 6. Now for **fertilisation!** combine your Gamete 1 cards with your partner's Gamete 1 hand, and your Gamete 2 cards with your partner's Gamete 2 hand (producing non-identical twins). As in the picture below, note how the children have inherited a mixture of chromosomes from all four grandparents - from their maternal grandmother (Hearts) and grandfather (Clubs) and from their paternal grandmother (Diamonds) and grandfather (Spades).

STEP 7. So now you have your first two children (of ten). As in Step 4, **add up the numbers you wrote on the cards to get a polygenic score for each child.** Using that score, predict the height of each child when they reach adulthood.



STEP 8. Separate the cards back out into the two parental piles (Hearts & Clubs, or Diamonds & Spades), shuffle them and with your partner repeat the processes of independent segregation and fertilization four more times (to generate eight more children). **Record your data on the next page.**

Polygenic data for height

You can record your data for your parental chromosome polygenic scores and for their children (F1 generation) using the table below. As an example, data from the two parental hands pictured on page 46 are shown here. Remember that to get the heights, apply the polygenic score to the national average of **165cm** for females and **178cm** for males.

	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	10	10	Q	Q/K	score	sex	height
Example Parent 1	+5	-5	-4	+4	-1	-1	+4	-2	+2	-6	+4	-5	+2	-4	-1	+3	-3	+1	+1	+4	-6	+4	-4	F	161cm
Example Parent 2	-4	+5	+3	-3	+6	+6	+2	-5	-5	-3	+5	+4	+2	-1	-6	+5	+5	+1	-1	+2	-6	+3	+15	M	193cm

	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	10	10	Q	Q/K	score	sex	height	
Your Parent 1																										
Your Parent 2																										
Child 1																										
Child 2																										
Child 3																										
Child 4																										
Child 5																										
Child 6																										
Child 7																										
Child 8																										
Child 9																										
Child 10																										

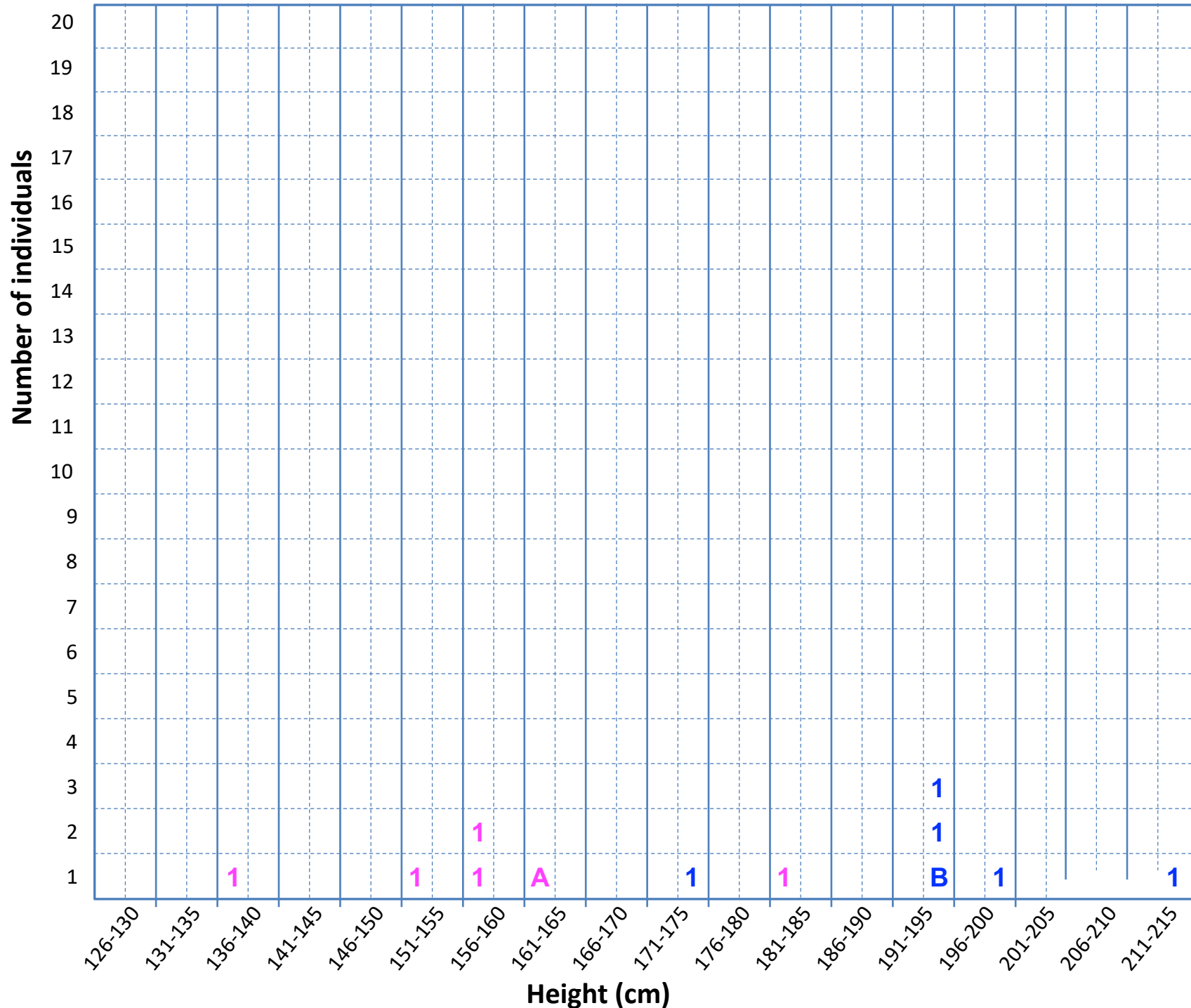
If your 'parent' hands could have an unlimited number of children, what would be the tallest and shortest adult height possible? To answer this for our example parents here, each would transmit to the **tallest** offspring their highest scoring of each chromosome. The remaining chromosomes give the **shortest** offspring possible. **How does your data compare?**

	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	10	10	Q	Q/K	score	sex	height
	+5	-5	-4	+4	-1	-1	+4	-2	+2	-6	+4	-5	+2	-4	-1	+3	-3	+1	+1	+4	-6	+4	+67	M	245cm
	-4	+5	+3	-3	+6	+6	+2	-5	-5	-3	+5	+4	+2	-1	-6	+5	+5	+1	-1	+2	-6	+3	-56	F	109cm

Distribution of heights from cards data

Send your parent and offspring sex and height data to your teacher, who will collate the classroom data and send it back for you. Using the graph below, make a histogram of parent and offspring heights in 5cm bins (meaning a size range: e.g. 126-130, 131-135...), graphing females (pink) and males (blue) separately. The parents (A & B) and ten offspring from our example are entered already. To record your data in the same way, enter your parents as C & D and put a 2 inside each square to indicate the second family entered. Enter eight more families from the class data in the same way (E & F, 3), (G & H, 4)....

Were there outliers too tall or short to graph?
What conclusions do you draw from the data?



How does this contrast with a Mendelian cross? What does this tell you about inheritance of polygenic disease?

The microbiome in health and disease

The interactions between our own bodies and the bacteria, protists, fungi and viruses that live on and inside us provide vital examples of gene-environment interactions and **pleiotropic** effects. The digestive tract, our gut, is essentially a specialized hollow tube that runs through us, so the inside of our gut tube (the lumen) is continuous with the outside world and can be considered part of the environment we inhabit. What lives in this environment? Trillions of microbes of several thousand species. Given a typical microbe might have 1,000 to 2,000 genes, the human gut is the playground of several million bacterial genes, vastly outnumbering our own 20,000 genes. We have a **symbiotic** relationship with our microbes, as both parties benefit. For example, we provide a stable nutrient-rich environment and gut bacteria produce vitamin B12, essential for our health.

Elite long-distance runners have increased Veillonella bacteria in their gut microbiome. These convert lactate to propionate, increasing the body's resilience to exercise stress

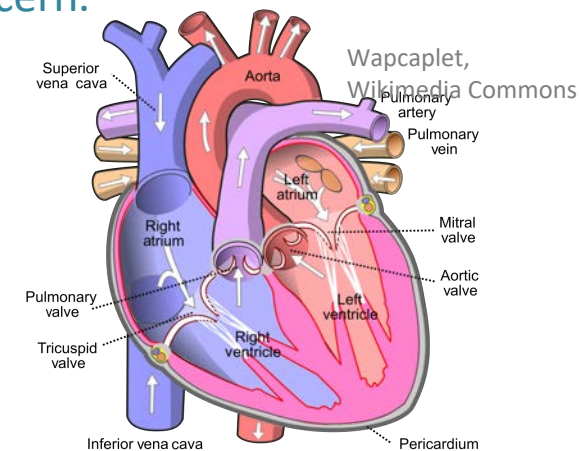


Virgin Money London Marathon

To learn more about the way in which the microbiome is established in young animals, from [termites](https://tinyurl.com/ULC-Termite) to [wasps](https://tinyurl.com/ULC-Wasps) to human [babies](https://tinyurl.com/ULC-Babies), watch the linked videos (or use the QR codes on page 56: <https://tinyurl.com/ULC-Termite>, <https://tinyurl.com/ULC-Wasps>, <https://tinyurl.com/ULC-Babies>).

Sometimes, bacterial species lead to disease. The bacteria *Streptococcus pyrogenes* (Group A Streptococcus, or GAS) is present in the nose and throat of 5-10% of healthy individuals. However, in some cases, GAS causes illness. In New Zealand, the connection between strep throat (caused by GAS), rheumatic fever and Rheumatic Heart Disease (RHD) is a major concern.

Watch this video about [RHD](https://tinyurl.com/ULC-RHD) in NZ. (<https://tinyurl.com/ULC-RHD>). Sore throats in young people (aged 4–19 years), especially Māori and Pacific children, should be checked by a doctor and, if strep throat, treated with antibiotics.



Streptococcus pyrogenes infection can trigger the body's immune system to produce antibodies against the 'M' protein of the bacteria. Unfortunately, **parts of the M protein resemble proteins used by the heart valves, so the body's immune response mistakenly attacks the heart too.** To see the [heart](https://tinyurl.com/ULC-Heart) in action, watch the computer simulation here (complex but cool!): <https://tinyurl.com/ULC-Heart>.

Gene-environment interactions are important in RHD. Infections with *S. pyrogenes* are more likely in damp and crowded households. The disease is therefore most often experienced by lower-income families. However, the risk of RHD is only partly due to social inequities. Some evidence suggests that genetic factors are also important. For example, a recent GWAS study found a link between DNA variation in an antibody gene and RHD susceptibility in indigenous peoples of the Pacific.

Epigenetics

On the previous page, we saw that initial differences in the infant microbiome or acute infection by *Streptococcus pyrogenes* can potentially have life-long consequences. **Many early-life events are known to affect health later in life.** These include when an infant experiences a stressful environment or, *in utero*, when a fetus is exposed prenatally to hormonal signals from a mother who is suffering from malnutrition, infection, stress or depression.

If an expectant mother is exposed to an extremely stressful event, this can result in alterations to the development and function of her offspring's **Hypothalamic-Pituitary-Adrenal (HPA) axis** (also see page 14). As adults, those offspring will have an increased risk of disorders such as obesity, diabetes and cardiovascular disease.



WWI French Soldiers
The National Library
of Scotland
CC BY-NC-ND 2.0

Bereavement is among the most stressful of human experiences. By examining the lifespan of orphans born to French mothers widowed during World War I, researchers found that in cases where fathers died *before*, but not after, the birth of a child, the lifespan of the orphans (as adults) was shorter, on average, by 2.4 years.

These and many other 'natural experiments' provide strong evidence for a developmental origin of health and disease. To understand how early experiences affect later health researchers have also studied these processes in animals.

Many experiments have focused on how maternal stress in mother rats affects care for their pups, and what the changes in care mean for the health and behaviour of those offspring when they reach adulthood (video also linked here: <https://tinyurl.com/ULC-Stress>). Some mother rats spend a lot of time licking and grooming their pups and those pups grow up to be less stressed and less aggressive than rats reared by mothers that seldom licked or groomed them. Female rat pups that became mothers themselves showed the same high or low maternal activity as their mothers. However, this was not due to genetics, because when female pups were switched at birth, they grew up to exhibit the same type of maternal behaviour as their foster mums. So how do stressful events early in life produce changes that last a lifetime? This occurs by **epigenetics** (meaning above, or on top of, genetics). As you learned on p.36, the DNA of certain regions can be marked by adding a methyl group (CH_3) onto some of the cytosines (**DNA methylation**) and histone tails are also modified to alter chromatin. These epigenetic alterations switch genes on or off. Most surprisingly, changes due to environmental experiences might even be handed down to the next generation, through the sperm or egg! (or use the link <https://tinyurl.com/ULC-Epigenetic> or QR code).



PikPng

CROSSWORD

2

1 2 3 4 5 6 7 8 9 10
11 12 13
14 15 16 17
18 19 20 21 22 23
24
25 26 27 28 29
30 31
32 33 34
35 36 37 38
39 40 41 42 43
44 45 46
47 48 49 50
51 52 53
54 55
56 57 58
59 60
61 62
63 64 65 66
67 68 69 70
71
72 73 74

See how much of this crossword you can complete from memory, then use the page references to fill in the rest.

Across

1. A term used - instead of mutants - to describe people with different alleles (7,8). p.19
8. A pleiotropic gene with roles in eye colour and kidney function. p.51
9. The Czech city that was home to Gregor Mendel. p.43
11. The critical amino acid coded by an ABCC11 allele, giving rise to dry ear wax. p.46
12. The number of phenotypes expected from a Mendelian cross of heterozygotes. p.44
13. The medical term for ear wax. p.46
14. An active person who might benefit from *Veillonella* bacteria in the gut. p.52
19. Rodent studied in the lab to investigate the link between stress and epigenetics. p.53
20. A common name given to a paternal relative (father or grandfather).
21. Little and white, the name for the fib told in classrooms about many traits. p.46
22. A sugar found in milk, that most mammals cannot digest when adult. p.42
24. The amino acid found at position 6 in the HbS sickle cell haemoglobin. p.42
25. Short name given to a primate cousin of humans, with reddish-orange hair.
28. A gene that plays a role in determining the colour of the eyes and skin. p.51
29. The amino acid coded by the RNA codon UUU (in DNA, TTT). p.13
31. The 'Father' of genetics. p.43
32. Name given to a square grid used to find the outcome of Mendelian crosses. p.44
33. A gamete that, in humans, always carries an X chromosome. p.48
34. A molecule produced in muscles as a result of significant exercise. p.52
35. A term used for the characteristic of skin or eye colour. p.51
37. The unfinished human genome sequence that was first published in 2001. p.37
39. A serious medical condition that can develop after *S. pyrogenes* infection (9,5,7). p.52
43. The organism Gregor Mendel focused his genetic studies on. p.43
46. A 'law' that must be followed, such as always pairing A with T and G with C. p.34
47. Short name for a gene encoding Tyrosinase, an enzyme required for the synthesis of melanin. p.51
48. Sometimes used to describe a very dark skin tone (and a wood used for piano keys).
50. How a person might describe the feeling of their throat, if ill. p.52
51. Cute canine companions carrying conspicuous coat colours. p.45
53. The opposite of you.
54. A gene expressed in melanocytes, coding for a protein required for the production of melanin. p.51
55. Dangerous medical complication in people who are HbS homozygotes. p.45
57. The single-cell embryo that arises from the union of egg and sperm. p.48
59. One side of a long running debate about factors determining behaviour/phenotype. p.51
61. The building blocks of organs, tissues and bodies.
62. A unit of genetic material that is either heterozygous or homozygous in most cells of the body. p.44
63. Shapes produced by clumping of haemoglobin in sickle red blood cells. p.42
64. Contrasting with clue 59, the environmental factors on the other side of the debate. p.51
67. Gametes produced by a male. p.44
70. A feline. Those with a c-kit mutation have white coats, blue eyes and may be deaf. p.51
72. Recessive phenotype Mendel observed when examining inheritance of Pea height. p.44
73. A rodent maternal care behaviour that may be culturally transmitted to offspring. p.53
74. Gene responsible for determining the type of cerumen produced by a person. p.46

Down

2. African country with world's highest number of people with sickle cell disease. p.45
3. Edible item useful for training pets (including deaf cats and labrador puppies).
4. Abbreviation for a pleiotropic *Drosophila* gene. p.51
5. A Mendelian law of the assignment of chromosome copies to the gametes (11,11). p.48
6. Word relating to the time (or place) of birth. p.46
7. Biological particle used in asexual reproduction and spread of an organism (e.g. fungi). p.48
8. Illness due to a genetic alteration in the shape and function of red blood cells (6,4,7). p.42
9. Tissue in which ABCC11 is expressed, regulating the production of colostrum, a specialized milk for neonates, rich in antibodies. p.46
10. Word for a trait or biological process regulated by just a small number of genes. p.43
15. Axis of several brain and glandular regions that work together in the stress response. p.14
16. Abbreviation used to signal amusement.
17. An important level of biological regulation that sits on top of, or above, genetics. p.53
18. Organism responsible for rheumatic fever (5,1,13). p.52
23. The expression of a genetically determined characteristic. p.42
26. A common name given to a grandmother.
27. The property of many genes that serve multiple different functions in an organism. p.51
28. A type of language someone might use to communicate with a deaf c-Kit mutant cat.
29. The observable characteristics or traits of an organism. p.41
30. The part of a pea plant most responsible for the difference in tall and dwarf phenotypes.
35. Prefix used with 6 down to indicate the period before birth. p.53
36. Haploid cells (carrying only one of each pair of chromosomes) created through meiosis. p.44
38. An amino acid specified by the DNA codon GCN (where N = A, C, G or T). p.10
40. The allele of a gene which expresses its phenotypic effect even when heterozygous. p.43
41. The common term for a symptom caused by *Streptococcus pyrogenes* infection (5,6). p.52
42. Structures in the heart that in RHD are damaged by the bodies own immune response. p.52
44. A pleiotropic gene that regulates aspects of eye colour, as well as the immune system. p.51
45. Term given for when the two copies of a gene are identical alleles. p.44
49. Tyrosylaspartic acid dipeptide (a combination of tyrosine and aspartic amino acids). p.10
52. Word for a trait or biological process regulated by a large number of genes. p.47
53. A pigment produced by melanocyte cells found in the skin and the iris of the eye. p.51
56. A disease transmitted by mosquitos which has had a strong influence on human genetics, especially involving genes important for red blood cell function. p.45
58. A large-scale genetic study looking for associations between DNA variations and traits or disease risks. Data from these studies are often presented as Manhattan plots. p.46
60. The gonad of a female fish, carrying a large number of gametes (eggs).
65. What a genetic score for a polygenic disease attempts to quantify. p.47
66. Aromatic molecular shape found in all four nucleotides, as well as some amino acids. p.7
68. A plant structure in which you might find two peas that look very similar.
69. Short version of 24 across. p.10
70. Sulfur-containing amino acid often used to link together different peptide chains. p.17
71. The immediate progeny of a genetic cross. p.43

QR codes for more video links

Lactase



Sickle



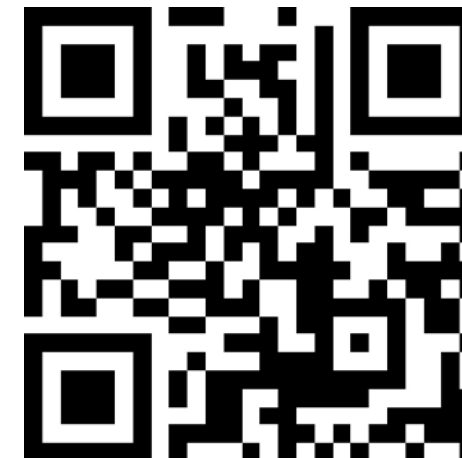
HbS



Variation



Labcoat



ABCC11



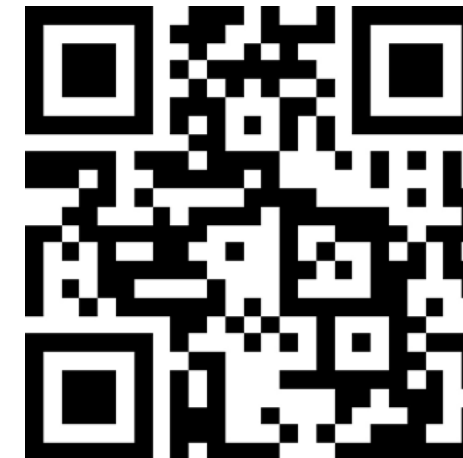
Polygenic



Skin



Termites



Wasps



Babies



RHD



Heart



Stress



Epigenetics



