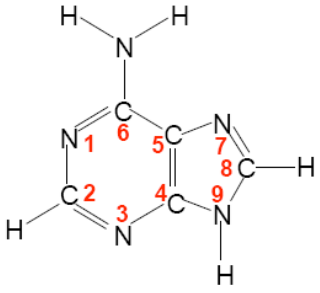
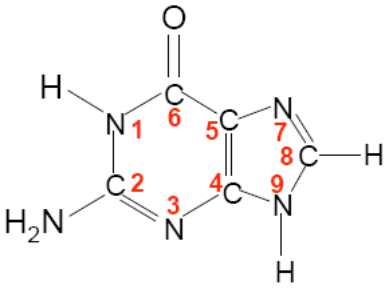
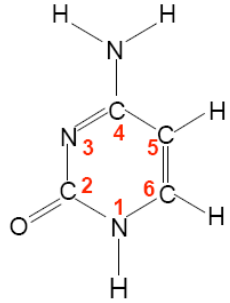
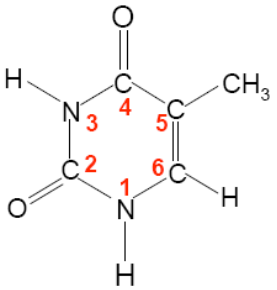
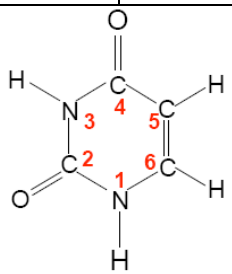
Nucleic Acids

Nucleotides

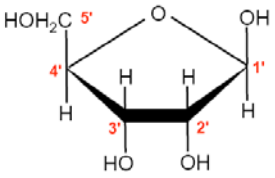
Nucleic acids are **polymers** made out of **nucleotides**. They are themselves made up of **three** building blocks:

1. They each contain a **base**. These are based on one of two **planar rings**

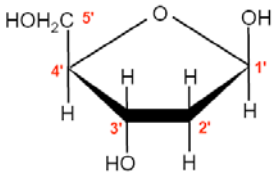
Based on **Purine**	
 **Adenine (A)**	 **Guanine (G)**

Based on **Pyrimidine**	
 **Cytosine (C)**	 **Thymine (T) – found in DNA**
 **Uracil (U) – found in RNA instead of thymine**	

2. The second building blocks are **pentose sugars** – **<u>ribose</u>** in **<u>RNA</u>** and **<u>deoxyribose</u>** in **<u>DNA</u>**.



β-D-ribose



β-D-2-deoxyribose

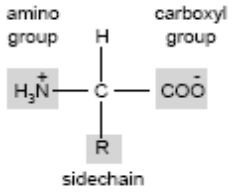
(Note that we **number** the **carbons** on the **sugars** with a **prime** to **distinguish** them from the **base** carbon atoms).

3. The final building block is simply **inorganic phosphate**.

Proteins

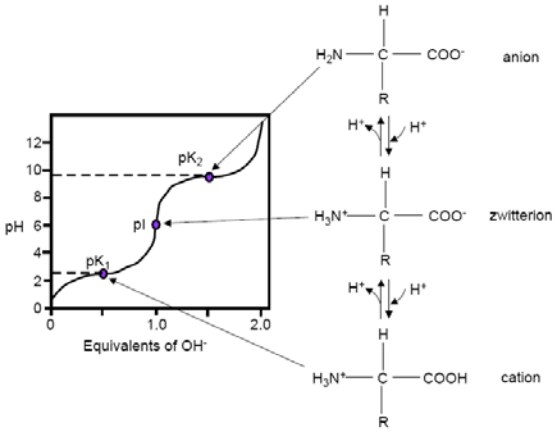
Amino Acids

The **monomer** buildings blocks of **proteins** are **amino acids**. They all have the **same** general structure in their **backbone**, but they **differ** in the "**R**"–group, which is called the **sidechain**. Amino acids are **ionised** at **neutral pH** – they are **zwitterions**



We call the **central carbon atom** then **α–carbon**.

As such, the amino acid can exist in **three states**, depending on which of its **side-groups** are **charged**. They therefore give rise to the following simple **titration curve**



A few technical details:

- If the **sidechain** (see below) also contains an **ionisable group**, the titration curve will have a **third pKa** value.
- The reason why the **carboxyl** group **loses** a **proton** *before* the **amino** group does (as OH^- concentration **increases**) is because the **carboxyl** group **ionises** more **easily**:
  - The **ionization constant** of the **amino group** in **amino acids** is **roughly 10^{-9} M**, and **pKa_2 = 9.9**.
  - The **ionization constant** of the **carboxyl group** is **roughly 10^{-2} M**, and **pKa_1 = 2.4**.

Reminder

⇒ The **ionisation constant**, $\mathbf{K_a}$, of an **acid** (or, in the case of $\mathbf{NH_3^+}$, a **conjugate acid**) is the **equilibrium constant** for its **ionisation**. For example, for the **carboxyl group**, the ionisation that occurs is

$$COOH \rightarrow COO^- + H^+$$

and so

$$K_a = \frac{\left[COO^-\right]\left[H^+\right]}{\left[COOH\right]}$$

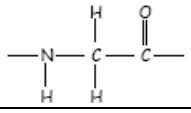
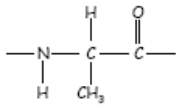
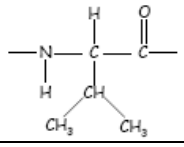
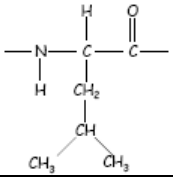
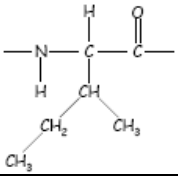
⇒ The **pKa** value of an **acid** is the **pH** at which it is **half dissociated**. It is **related** to the **ionisation constant** by

$$pK_a = \log\left(1/K_a\right)$$

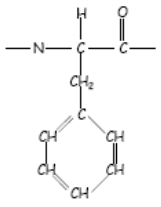
- The **pKa** of the **carboxyl group** in **acetic acid** is **4.8**. This means that amino acids **lose** their **protons** about **300 times more readily** than acetic acid. This is due to **neighbouring positively charged amino group** which **withdraw** electrons from the **carboxyl group**, **stabilising** the **deionised** form.

There are **twenty** different types of **amino acids**. They can be divided on the basis of their **size** and **chemical nature**, and it is their **sidechains** that confer **distinctive properties** on **individual proteins**. The **chemical diversity** of the **R-groups** provides proteins with **limitless structural** and **functional diversity**.

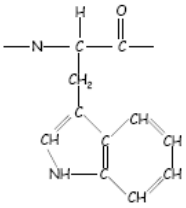
The amino acids are

Without a sidechain	
Glycine (Gly, G) 	
With **aliphatic hydrophobic** sidechains	
Alanine (Ala, A) 	Valine (Val, V) 
Leucine (Leu, L) 	Isoleucine (Ile, I) 
With **aromatic hydrophobic** sidechains	

Phenylalanine (Phe, F)

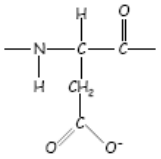


Tryptophan (Trp, W)

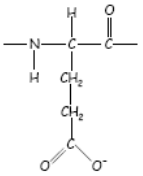


With **negatively charged** sidechains (at **neutral** pH)

Aspartic Acid (Asp, D)

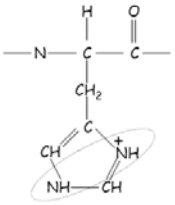


Glutamic Acid (Asp, E)



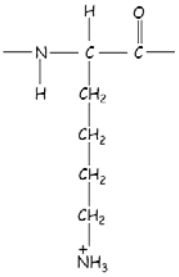
With **positively charged** sidechains (at **neutral** pH)

Histidine (His, H)

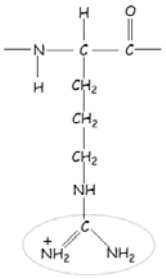


The **nitrogen atoms** indicated only have a **weak affinity** for H^+, and so are only **partially positive** at **neutral pH**. (This **nitrogen-containing** ring is also **aromatic**).

Lysine (Lys, K)



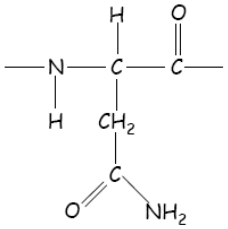
Arginine (Arg, R)



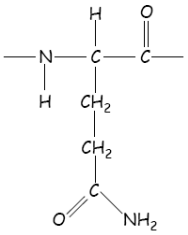
The group indicated is **very basic**, because the **positive charge** is **stabilized** by **delocalization**. It is therefore able to **accept protons more easily**.

With **polar** sidechains due to **NH_2 groups**

Asparagine
(Asn, N)

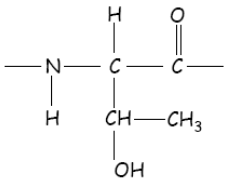


Glutamine
(Gln, Q)

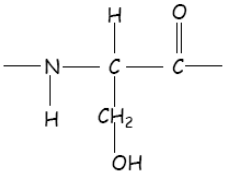


With **polar** sidechians due to **OH groups**

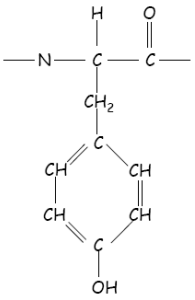
Threonine
(Thr, T)



Serine
(Ser, S)



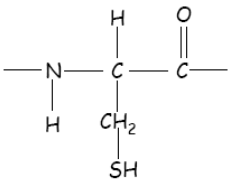
Tyrosine
(Tyr, Y)



(This amino acid is polar *and* aromatic)

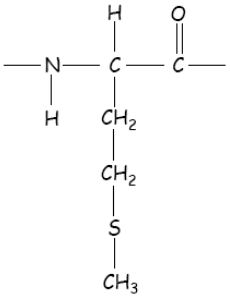
With sidechains containing **sulphur**

Cysteine
(Cys, C)



Cysteine has the ability to make
disulphide bonds

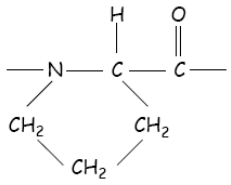
Methionine
(Met, M)



(Not very commonly found)

Immino Acids

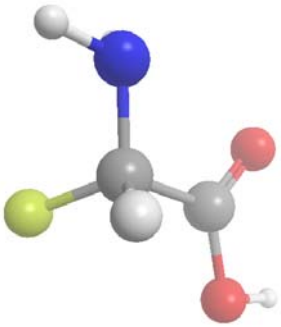
Proline
(Pro, P)



The fact that the **R-group** of **proline loops back** onto itself makes at an **immino acid** rather than an amino acid.

Of the **twenty** amino acids, **nineteen** of them have **four different** groups attached to the **α–carbon** and are therefore **chiral** (**glycine** is the exception).

An amino acids is classed as **L** or **D** depending on its **similarity** to **glyceraldehyde**. **Only L** amino acids are used in **natural proteins** (in **carbohydrates**, the **D**-form is used). A simple rule to determine the **chirality** of a given acid is the **CoRN** rule:



With the **hydrogen** atom on the **α–carbon** pointing **towards** us, one should be able to read the word **CoRN** (where **Co** is the **carboxyl group**, **R** is the **R-group** and **N** is the **amino group**) **clockwise**.

Protein Structure

We now deal with the **structure** of **proteins**. These are so **diverse** and **complicated** that they have been **broken down** into **four categories**, so that they can be **classified** more easily.

Proteins are **polypeptides**. They are made up of several amino acids **joined together** by **peptide bonds**. This **chain** of amino acids then **folds** into **complex structures**.

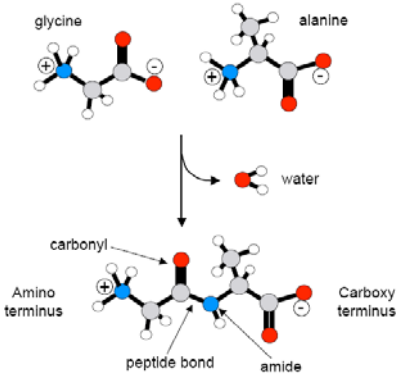
1. The **primary structure** of a protein is its **amino acid sequence**, which is **genetically determined**.
2. The **secondary structure** of a protein refers to the **regular structures** that are formed by the **backbone** of the protein.
3. The **tertiary structure** refers to the **organisation** of these **structures** in **three dimensions**.
4. The **quaternary structure** refers to the organisation of **subunits** in **oligomers**.

Before we begin looking at each of these structures, we look at the **peptide bond**, which hold amino acids **together** in proteins.

The Peptide Bond

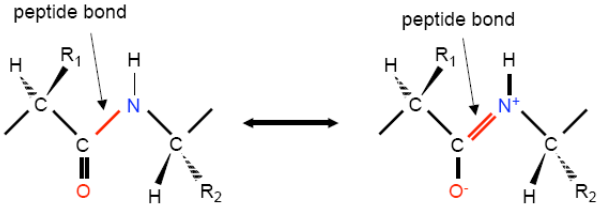
Peptide bonds are formed by **condensation reactions** between **amino acids**, which involve a **loss of water**. The **amino** and **carboxyl** groups **lose** their charges as a result of the reaction. *In vivo*, this is done by **ribosomes**.

For example, between **glycine** and **alanine**:



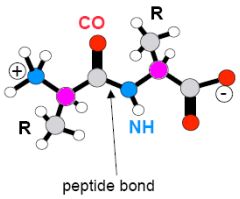
The resulting polypeptide has **directionality**. One end (called the **N-terminus**) has a **free amino group** and the other (call the **C-terminus**) has a **free carboxyl** group. By convention, polypeptide chains are **numbered from the N-terminus to the C-terminus** and **written** from **left to right**.

The **peptide bond** is **not** a **normal covalent bond**. It is a **resonance hybrid**, in **equilibrium** between **two different forms**.

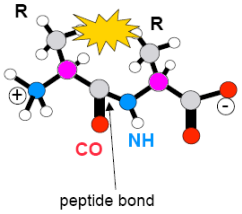


As such, the **peptide bond** is about **40% double bond** in **character**. A consequence of this is that there is **no freedom of rotation** around the peptide bond. It is **rigid** and **planar**. There is, however, some **plasticity**, due to the **partial single-bond character** of the peptide bond.

The **peptide bond** is **always trans** – ie: with the **α–carbons** on **opposite sides** of the bond: [note that the two **fuschia α–carbons** are pointing in **opposite directions** – one **into** the **plane** of the **paper**, and one **out of it**]



This is because the **cis** form is very **unstable**, due to the **clashing** of **R-groups** (represented by an artistic yellow star in this diagram)



Primary Structure

The **primary structure** of a **protein** is its **amino acid sequence**. It is **genetically defined**, and each protein has a **unique sequence encoded** in one or more **genes** that are **translated** during **protein synthesis**.

Short stretches of protein are still **sequences** from the **N-terminus** using **Edman degradation**, where each amino acid is **identified** and **removed** from the protein, starting at the N-terminus. This method, however is only **suitable** for **up to** about **15 amino acids**. A much more **efficient** method used nowadays is to **sequence** the **DNA** and use this to **read off** the **sequence** of the **protein**. (This method was, in part, discovered by Fred Sanger in 1953 – a discovery which won him the Nobel Prize).

The only **covalent bonds** linking **amino acids** are **peptide bonds**, but **disulphide bridge** (**disulphide bonds**) can also be formed between **cysteine residues**. These are **rare** in **intracellular** proteins.

Some proteins are **rich** in certain **amino acids** which are **suitable** for the **structural needs** of the protein. For example, proteins that **bind** to **DNA** (which has a **negatively charged phosphate backbone**) are like to be rich in the **positively charged** amino acids (**lysine** and **arginine**).

Secondary Structure

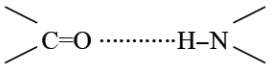
The **secondary structure** of a protein refers to the **regular structures** that are formed by the **backbone** of the **protein**.

The primary structure of proteins contains a number of **hydrogen bond donors** and **hydrogen bond acceptors**. These **bond with each other** in proteins to form **extensive**

hydrogen bonded networks. This is achieved by forming **regular extended arrays** called **secondary structures**.

Stable secondary structures must fulfil the following criteria

- The **peptide bonds** must be **planar** and have **favourable** bond **lengths** and **angles**. (The **Ramachandran plot** gives us information as to what these favourable angles might be – see the next section).
- Every **carbonyl oxygen** and **amide nitrogen** must be involved in **H-bonding**, like so:

  This is because these two atoms are the most reactive ones in the molecule and need to be stabilized.[1]
- The H-bonded atoms are in a **straight line**. This is the configuration at which the H-bond is **strongest** and **most stable**.
- The **spatial orientation** of every residue **with respect to the next** must be the same (ie: the **operation** of going from **one residue** to **another** [eg: going up one step in a helix] is the **same** for **every residue**).
- The **sidechains** projection **out** form the **structure**, to ensure **minimal steric interference** in the structure.

There are **two main types** of **secondary structures** in proteins that **meet** these **criteria** – the α**-helix** and the β**-sheet**.

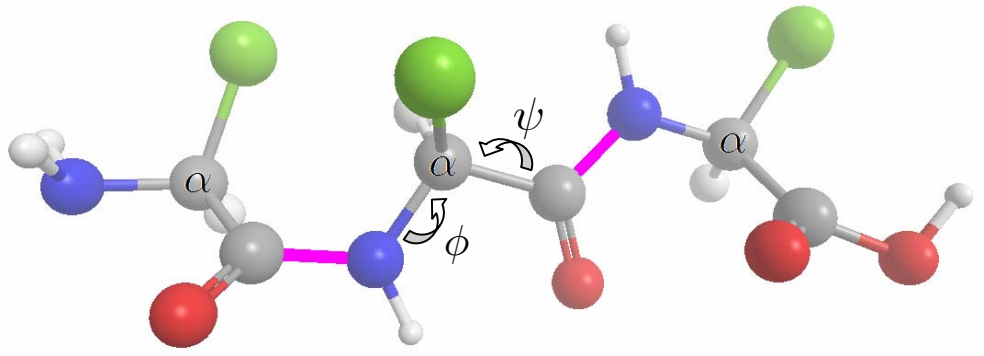
Before we look at these structures, however, we examine the **first criterion above** (relating to **bond angles**) in more detail.

Bond angles in polypeptides

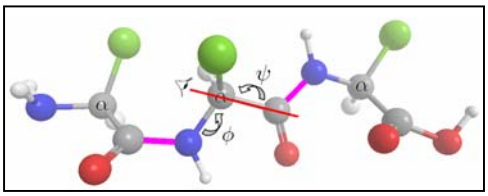
As was mentioned above, **no rotation** is possible around the **peptide bond**. However, rotation *is* possible around the **amino–α carbon** bond and the **carboxyl–α carbon** bond. These two angles are **denoted** by ϕ and ψ **respectively**. Since these are the **only** bonds that can rotate, the **conformation** of an **amino acid** in a **protein** can be **described solely** by the ϕ and ψ angles.

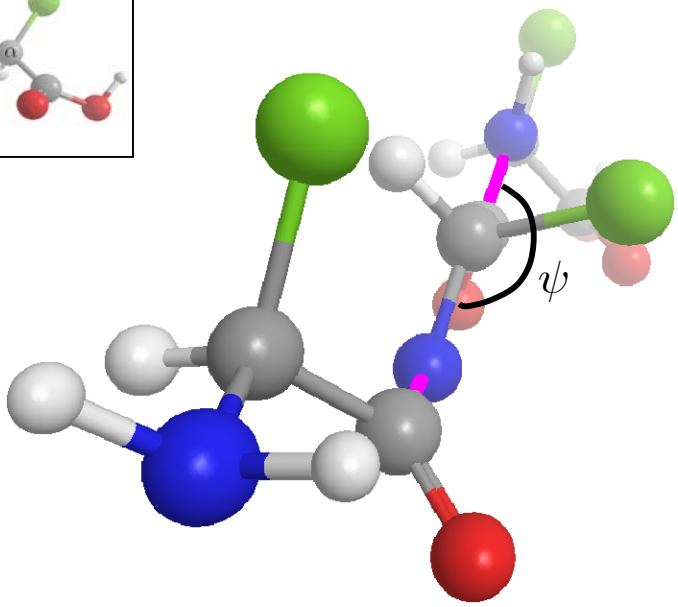
[1] True?

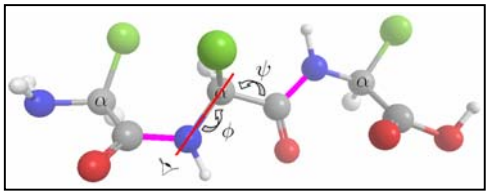
In this diagram, the fuchsia bonds are the peptide bonds whose rotation is constrained.

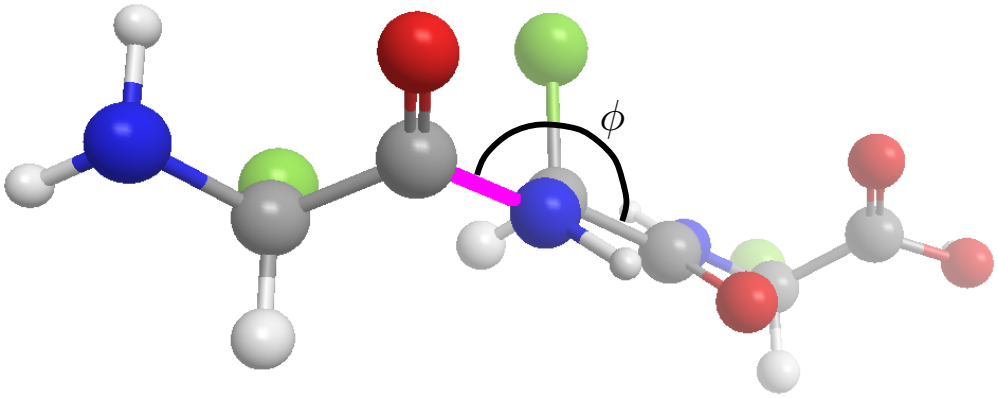


The best way to **visualise** these bonds is to use **Newman Projections** – these involve looking down the molecule in a certain direction to make finding the angle simpler:

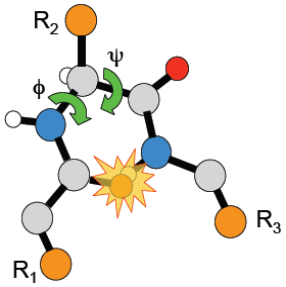




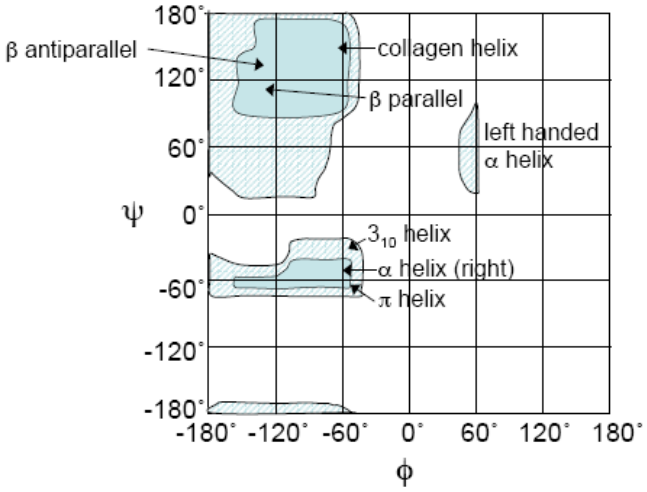




Now, it turns out that **not** *all* **angles** give rise to a **stable** polypeptide. For **example**, if **both angles** are **0**, then **steric clashing** makes the molecule very **unstable**



In fact, **most** combination of **angles** gives rise to **unstable** molecules. Those that are **viable**, however, can be **calculated** and displayed in a **Ramachadran Plot**



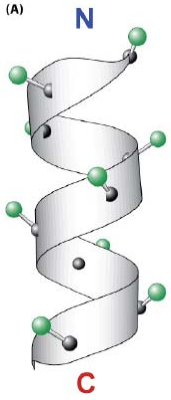
A few notes on the plot:

- The areas in **white** on the plot indicate **unstable** angle combinations. The **dark regions** indicate the **most stable** combinations.
- **Most** angles in **proteins** fall within the **permitted regions** of the plot. **Two** proteins are **atypical**:
  - **Proline** is **restricted** to a ϕ of **–60° to –77°**. This is because the **N-α carbon bonds** is part of the **pyrrolidine ring** by by the **sidechain**. This **restricts motion** around the bond.
  - **Glycine** has only a **hydrogen aotm** for a **sidechian**. This means it has **more streric freedom** than most amino acids, and has **more combinations** of angles allowed.
- The plot is only valid to **L-amino acids**. If proteins were built from **D-amino acids**, the possible combinations would be **different**.

We now explore two of these "permitted configurations" – the **α–helix** and the **β–sheet**.

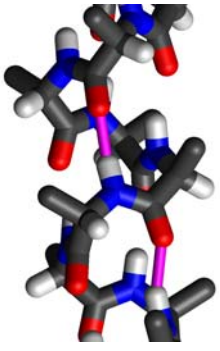
The α-helix

The **α-helix** is a **spring like structure** – the **inner part** of the spring is formed by the **main chain** of the protein [*in black in the diagram*], and the **sidechain R-groups** [*in green in the diagram*] radiate **outwards** from the center.



A few points regarding the α-helix

- The **main chain** of the protein, consisting of **amino acids** linked by **peptide bonds** forms the **backbone** of the helix.
- The helix is **stabilized** by **H-bonds** between the **C=O** groups and **N–H** groups in the main chain.
  - These point **inwards**, towards the center of the helix.
  - The **C=O** group of each amino acid is bonded to the **N–H** of the amino acid located **four residues** ahead of it in the chain.
  - This means that an amino acids **i** is H-bonded to amino acid **i+4**. We therefore call this an **(i, i+4)** helix.
  - Other helical forms are possible, like **(i, i+3)**, and **(i, i+5)**, but the **(i, i+4)** helix is the most stable.

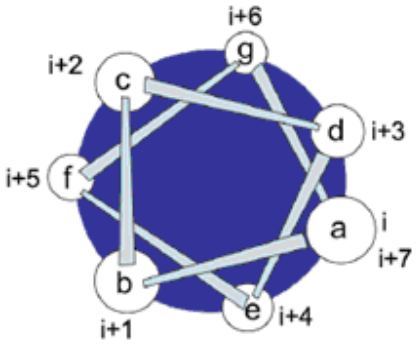


- As a result of these bonds (and of the fact that there is no "space" inside the helix), the **inside** of the **helix** is **hydrophobic**. This **strengthens** the **helix** (see next section).
- The helix has **directionality** – one end of the molecule has a COOH group (the **C end**) and one end has an NH_2 group (the **N end**). Every **–NH** group in the

helix points towards the **N terminal** and every **–COOH** group groups towards the **C terminal**.

- Some technical details
  - The helix is about 5Å high.
  - Each amino acids is related to its neighbour by
    - A rotation of 100° along the helix axis.
    - A rise of 1.5 Å.
  - One full turn is 5.4 Å high.
  - One full turn contains 3.6 residues.
  - The α-helix is a **right-handed** helix. When looking down the **N end**, the helix curls **clockwise**. [This is for **L** amino acids. For D amino acids (if they existed in natural proteins), the helix would curl anticlockwise]. *This is in contrast with the DNA double-helix which is* ***left-handed****.*

Helices are commonly drawn as **helical wheels** where **7 amino acids residues** are represented by the letters A–G.



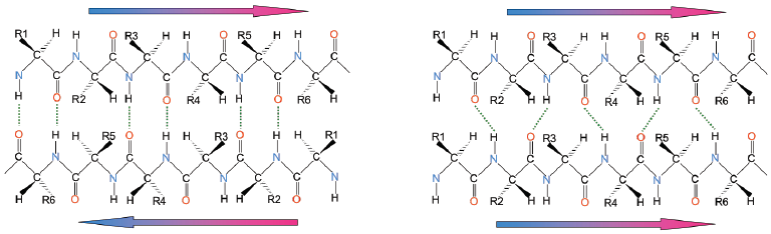
(This is drawn looking down the N end).

This is a simple way to see **which amino acids** are **close to each other** in the helix, and which **bond to each other**.

The β-sheet

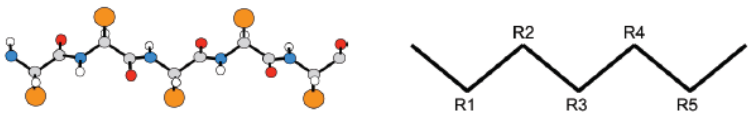
In a β-sheet, the polypeptide backbone is extended into **strands**, which are connected by **ladders** of **hydrogen bonds** to **other strands**.

The strands can run in **opposite directions** in **antiparallel sheets** or in the **same directions**, in **parallel sheets**.



The **parallel** form is **less stable**, because the hydrogen bonds involved aren't straight.

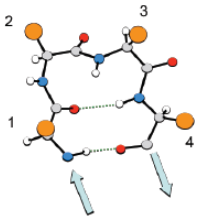
In both cases, the **sidechains** poke **alternately above** and **below** each strand. Thus, the carbons **i**, **i+2**, **i+4**… are on the same side of the sheet.



The **distance** between each R group on one side of the molecule is 7 Å.

Putting secondary structural elements together

To put **secondary elements** together, we need **turns** between them. The **reverse β-turn** is a very abundant type of turn, which enable **polypeptide chains** to **revers direction**



The **β-turn** must contain **at least four residues**.

Tertiary Structure

We now examine the different ways these secondary structures organise each other and bind with each other to form the final shape of the protein.

Forces

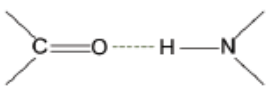
We first look at the range of sources that exist which can hold different amino acids together in secondary structures.

1. **Disulphide bonds (Covalent bonds – 350–450 kJ mol^{-1})**

   CH_2— S—S —CH_2

   Disulphide bonds occur between **cystine** residues and hold the protein together. Typically, however, there are no more than two or three per molecule, so other forces must be involved.

2. **Hydrogen bonds (2–20 kJ mol^{-1})**

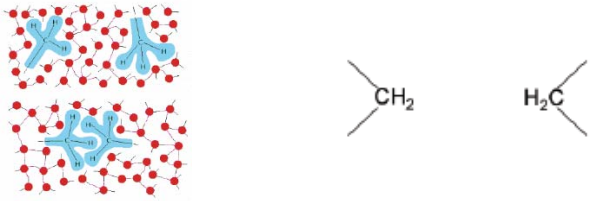


Several amino acid side chains have hydrogen bond donor and acceptor atoms and so have the ability to form hydrogen bonds with other amino acids.

3. **Electrostatic interactions (40–200 kJ mol^{-1})**

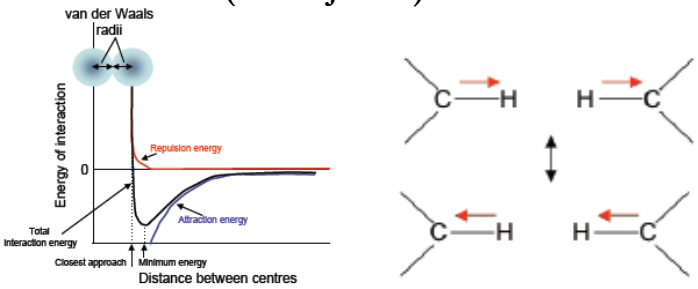
- Form between **amino acids** that have **positive** (**Arg**, **Lys**) and **negative** (**Glu**, **Asp**) charge at **physiological pH**.
- The **magnitude** of the force is **hard to quantify** for proteins in an **aqueous environment**.

4. **Hydrophobic interactions (3–10 kJ mol^{-1})**



- Due to the **preference** of **non-polar sidechains** to **associate with each other** rather than with **water**.
- This is an **entropic effect** – water molecules form **structured**, **H-bonded** structures around **hydrophobic groups**. These do not form if the hydrophobic groups are **H-bonded** with **each other**.
- In other words, water molecules are **more ordered** round the **separate units** than round the **unit bound together**.

5. **van der Waals Interactions (0.4–4 kJ mol^{-1})**



- Can occur between **uncharged atoms**.
- When the **electron distribution** around one atom **fluctuates**, it **induces** similar fluctuations in **neighbouring atoms**. This leads to **attraction** between the atoms.
- The forces operates only between atoms that are **close together** and falls off rapidly with the **sixth** of the **distance** between them.
- When the atoms are **very close**, **repulsive forces** between **electron clouds** come into play (**inverse twelfth power** of **distance**) and **counter** the **attractive forces**.
- The **closest** two atoms can approach each other is the **sum** of the **van der Waals radii**.

- In **reality**, the atoms will be **slightly further apart**, at a position such that their energy is as low as possible.

A few noteworthy points regarding these forces:

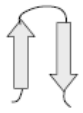
- **Individually**, these forces are rather **weak**. There are, however, **very many** of these (**non-covalent interactions**) in a **protein**, and **collectively**, they can be **relatively strong**.
- They can also form **bonds** between **different macromolecules** or between **macromolecules** and **various ligands**.
- Many of these interactions occur **only** when **water** is the solvent.

Supersecondary structures

Teritary structures can be dissected into various commonly found **combinations** of **α-helices** and **β-strands**. These are called **supersecondary structures** (or **motifs**).

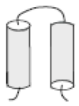
Several of the most common (though there are many others) are:

❖ The **β hairpin**



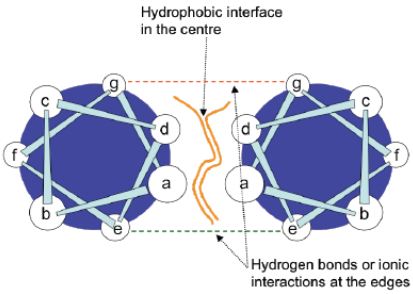
This simply consists of **two antiparallel β–strands** connected by a **β–turn** and held together by **hydrogen bonds**.

❖ The **α hairpin**



This is a similar structure, involving **two α-helices**. A few points:

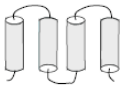
- The two helices **cannot** bond to each other using **hydrogen bonding** because the **backbone CO** and **NH** groups are **already involved** in hydrogen bonds **within** the helix.
- The helices, therefore, tend to be **amphipathic** – **one side** of each helix has amino acids with **hydrophobic sidechains** and the other has amino acids with **hydrophilic sidechains**. [This is illustrated in the next diagram, where **d** and **a are hydrophobic** and **c, f** and **b** are **hydrophilic**].
- The helices therefore **arrange themselves** so that both **hydrophobic** sides **face each other**:



This configuration is reinforced by two phenomena

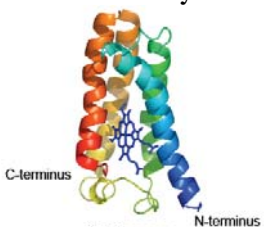
- **Hydrophobic** and **van Der Waals** forces act along the interface between the two **hydrophobic faces**.
- At the **edge** of the **hydrophobic interface**, **hydrogen bonds** or **ionic interactions** strengthen the **attraction** at the interface.

❖ The **four helix bundle**



This is a simple extension of the **α-hairpin**.

It is commonly found in proteins that **bind** to the electron and oxygen carrier **haem**. For example, in the electron carrier **cytochrome c**:



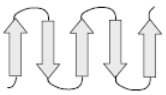
The reason why the four helix bundle is so **well adapted** to **carrying** the haem group is because it is only held together by **hydrophobic interactions**. Since the group **doesn't unduly disturb** these interactions (because more can be made with the group), the resulting structure is stable. [In a **β-hairpin**, on the other hand, extensive **disruption** of hydrogen bonds would be required to introduce the group].

❖ The **β–α–β motif**

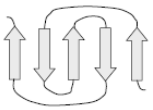
This is **often found** in proteins with **parallel β-sheets** since they **can't** be **connected** by a **simple β-turn**. This is, in effect, the way we **stabilise parallel β–sheets**.

Proteins that **bind** to **nucleotides** are often made up of **β–α–β motifs** arranged into a **rossman fold**.

❖ The **β–meander**

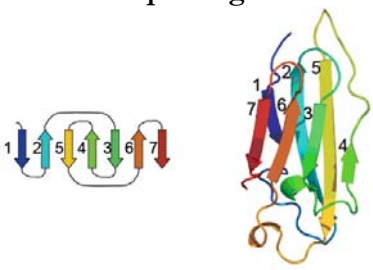


- The **Greek key**



This motif is a common β–structure.

The **β–sandwich** structure uses a **Greek key** motif, and is the basis of the **immunoglobin fold**, which is the **repeating unit** of **antibodies**.



It is held together by **hydrophobic interactions** between sidechains (like the **α–hairpin**)

Methods used to determine 3D structure

Various methods are used to determine the 3D structure of proteins. The most common are

- **X-ray diffraction**
  - **Sample** must be in the form of an **ordered structure** (**crystalline** or **fibrous**).
  - **X–rays** are **fired** at the sample and **scattered**, mainly by the **electrons** in the molecule.
  - The **scattered waves recombine** or **cancel** (depending on their **phase difference**).
  - The resulting **diffraction pattern** is a **3D array** of **spots**.
  - The **structure** is related to the **diffraction pattern** by a **Fourier transform**.
  - If the **resolution** of the **electron density map** is good enough, it can be interpreted to give a **3D model** of the protein.
- **NMR spectroscopy**
  - Applicable only to **small** proteins (**molecular mass** < **40,000**) in **solution**.
  - The **solution** is placed in a **strong magnetic field** and **irradiated** at **radio frequencies**.
  - The resulting **absorption spectrum** can be used to determine **inter-proton distances** in the protein.
  - If a sufficient number of these is determined, a **3D model** that **uniquely satisfies them** can be **computer**.

- **Cryo-electron microscopy**
  - This is a **microscopy technique**.
  - The sample is **frozen** in a **very cold liquid refrigerant** to **preserve** and **protect it** during **observation**.
- **Atomic force microscopy**
  - An **atomically sharp tip** is **scanned** over a **surface**.
  - **Feedback mechanics** enable **piezo-electric** scanner to maintain the tip at a **constant force** (measuring height) or **constant height** (measuring force) **above** the **sample surface**.

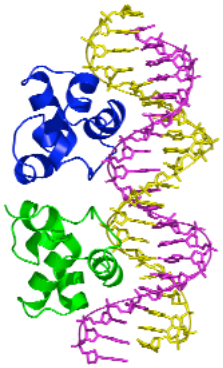
Quaternary Structure

Not all proteins are monomers. Many are in fact **oligomers**, with **dimers** (2 molecules) and **tetramers** (4 molecules) being particularly common.

Example I – CRO

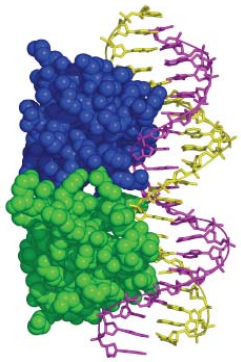
DNA binding proteins are often **dimers**, because they bind to pieces of DNA with **repeating** sequence.

For example, the protein **CRO**, from **bacteriophage λ** (a **virus** that infects **bacteria**) is a **dimer**, each **subunit** having two **α–helices** that **bind** to **adjacent major grooves** on **DNA**.



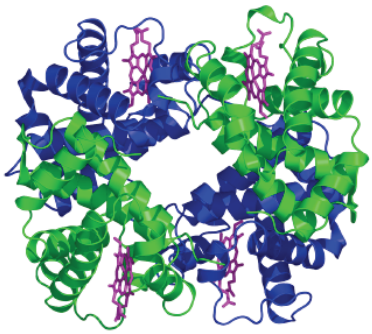
It turns out that each helix is *just* the right size to bind to the major groove.

Note, also, that despite the fact the dimers look far apart on the diagram above, they are in fact very close to each other in a space-filling representation



Example II – Haemoglobin

Haemoglobin is made up of **four myoglobin-like** chains, assembled into a **tetramer**. To form its **quaternary structure**, there has to be a **precise molecular fit** between the different chains.



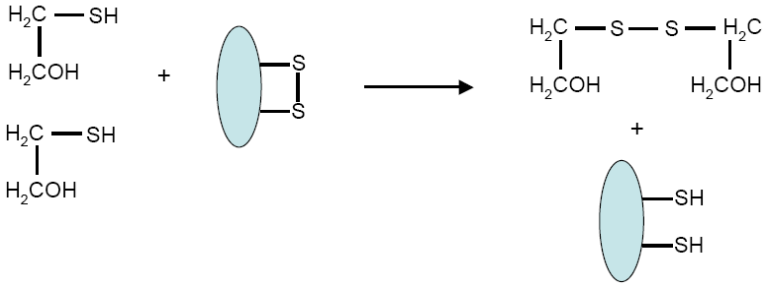
In sufferers of **sickle cell anaemia**, there is a **mutation** on the **surface** of **two** of the **monomers** that **replaces** an **amino acid** with a **charged sidechain** to one with a **hydrophobic sidechain** (glutamic acid to valine). This causes haemoglobin molecule in **interact** with **each other** (via hydrophobic interactions) and to **polymerise** in the **red blood cells**, causing the characteristic **sickle shape**.

Protein Folding

How do proteins "know" what shape to take? Early experiments with **ribonuclease** (an **enzyme** that **degrades RNA**) shed some light on the question.

The enzyme was first treated with two chemicals

- **β–mercaptoethanol** – this reduces any **disulphide bonds** in the protein



- **8M Urea** – Urea **readily** forms **hydrogen bonds**. At **high concentrations**, it **disrupts** hydrogen bonding between **water molecules**. This disrupts the **hydrophobic effect**.

As a result of this treatment, the enzyme is **denatured** (unfolded). However, when the **reducing agent** and the **denaturant** are **removed**, the protein can **spontaneously refold** to the **catalytically active** form!

From this, we can conclude that

- The **amino acid sequence** is sufficient to **fully specify** the **3D structure** of the protein.
- The **native form** of the **protein** is the **most thermodynamically stable** structure.

For more **complicated** proteins, **spontaneous folding** of this kind is **unlikely**, because there can be various **stable structures** (even though they **might not** be the **most stable**).

In vivo, other proteins called **chaperones** may **help** proteins to **fold correctly**. This is particularly the case for proteins with **multiple subunits**, which are likely to be **more difficult** to fold.

An **example** of such a **chaperone** is a "**doughnut**-shaped" molecule which "**threads**" the amino acid chain through. This allows **some parts** of the protein to fold without **hindrance** from **other parts**.

In any case, the chaperone **does not provide** any more **information**. The **amino acid sequence** still **fully specifies** the **3D structure**.

There is also evidence that proteins **fold** via **defined pathways**. Originally, some secondary structure elements are **transiently** formed (often **α–helices**, because they require no **3D folding** of the protein to form – as opposed to β–sheets which do). These may then **fold together** into a **sub-domain**, which can stabilise **further secondary structural elements**. In this way, the structure is **built up** in **stages**, until the final, **native structure** is formed.

Protein Misfolding

Protein misfolding is thought to be important in some **diseases**. For example

- In **Alzheimer's disease**, **multiple copies** of a protein called **Aβ assembles** into **amyloid firbrils**, which are the main components of the **plaques** observed in the brains of Alzheimer's sufferers. These are composed of **many copies** of the **amyloid protein**, arranged into a **single**, **repetitive β–sheet**.
- The **prion protein** responsible for **BSE** ("mad cow disease") also makes **amyloids**. It is thought that a protein is the **infectious material**, which acts as a **centre** for the **formation** of **amyloid firbrils** by the **native prion proteins**.

Enzymes

Proteins as Enzymes

The Basics

- Most **reactions** in the cell do not **spontaneously** take place **fast enough** to be **useful**.

- Enzymes are **catalysts** – they do not change the **equilibrium** of reactions but **speed up** the **rate at which equilibrium** is reached, with varying levels of efficiency
  - **OMP decarboxylase** increases rate from **2.8×10^{-16} s^{-1}** to **39 s^{-1}** – a **1.4×10^{17}**-fold increase.
  - **Triose phosphate isomerase** increases rate from **4.3×10^{-6} s^{-1}** to **4300 s^{-1}** – a **1.0×10^{9}**-fold increase.
  - **Carbonic anhydrase** increases rate from **1.3×10^{-1} s^{-1}** to **1×10^{6} s^{-1}** – a **7.7×10^{6}**-fold increase. This is one of the **fastest** known enzymes.

- Enzymes are almost always **proteins**, which make good enzymes thanks to the **large variety** of **sidechains** available that can provide a **variety** of different **environments**.

- Enzymes are **highly specific** both in the **reactions** that they catalyze and in their **choice of substrates**. This is due to the **precise interaction** of the **substrate** with the **enzyme**, which results from the **intricate 3D structure** of the **enzyme protein**. They usually catalyze a **single reaction** or a **set of closely related reactions**. **Side reactions**, leading to the wasteful product of **by-products**, are **rare**.

  The **specificity** of enzymes **varies**, though. We can take **<u>proteolytic enzymes</u>** (**which hydrolyse** a **peptide bond**) as an example
  - Most will also catalyse a **related reaction** – the **hydrolysis** of an **ester bond** (which, incidentally, is **more easily monitored** and so therefore useful in the **experimental investigation** of such enzymes).
  - Difference **proteolytic** enzymes vary in their degree of **substrate specificity**
    - **Subtilisin** (found in **certain bacteria**) will cleave **any peptide bond**.
    - **Trypsin** (a **digestive enzyme**) will only cleave bonds on the **carbozyl side** of **lysine** and **arginine** residues.
    - **Thrombin** (participates in **blood clotting**) catalyses the hydrolysis of **Arg-Gly** in **particular peptide sequences <u>only</u>**.
    - [**DNA Polymerase I** (a **template-directed enzyme**) **adds nucleotides** to a DNA strand being synthesised **based** on **another**

strand that serves as a **template**. It is **remarkably precise** and inserts the wrong nucleotide only **1 in 1000000** times].

- Many enzymes require **cofactors** for activity – we use the following terminology

  Apoenzyme + Cofactor = Holoenzyme

  Cofactors can be split into two groups
  - **Metals** – eg: **carbonic anhydrase** needs **Zn^{2+}**.
  - **Small organic molecules** called **coenzymes** – eg: **glycogen phosphorylase**, which **mobilises glycogen** for energy, needs **pyridoxal phosphate (PLP)**.

    Coenzymes are often **derived** from **vitamins**, and can interact with enzymes in two ways:
    - **Prosthetic groups** are **tightly bound** to the enzyme.
    - **Cosubstrates** are **loosely bound** to the enzyme, because they **bind to** and are **released from** the enzyme just as normal substrates are. Enzymes that use the **same cosubstrate** are often **mechanistically similar**.

- Enzymes are also able to **transform energy** from **one form** into **another**. Very often, during a **reaction**, the **energy of the reactants** is converted with **high efficiency** into another form, and enzymes **catalyse** these reactions. An example is **photosynthesis**.

- Enzymes are **classified** into **six major groups**, based on the **type of reactions** that they catalyse
  - **Oxidoreductases** catalyse **oxidation** and **reduction** (the **transfer** of **electrons**) – for example
    - **Lactate dehydrogenase**
    - Any **dehydrogenase**, **reductase**, **oxidase** and **catalase**.
  - **Transferases** catalyse the **transfer** of **functional groups** – for example
    - **Nucleoside monophosphate kinase**
    - Any **acetyltransferase**, **methylase**, **protein kinase** and **polymerase**.
  - **Hydrolases** catalyse **hydrolysis reactions** where a molecule is **split** by the **addition of water** – for example
    - **Chymotrypsin**
    - Any **protease**, **nuclease**, **phosphatase**.
  - **Lyases** catalyse the **cleavage** of **C-C**, **C-O**, **C-S** and **C-N** bonds but do **not add** anything – they form **double bonds** – for example
    - **Fumarase**
    - Any **decarboxylase**, **adolase**
  - **Isomerases** catalyse **atomic rearrangements** within molecules (**intramollecular group transfer**) – for example
    - **Triose phosphate isomerase**
    - Any **racemase**, **mutase**

- **Ligases** catalyse the **ligation** of **two subtrates** at the **expense** of **ATP hydrolysis** – for example
  - **Aminoacyl-tRNA synthase**
  - Any **DNA ligase**, **peptide synthase**, **fatty acid synthase**

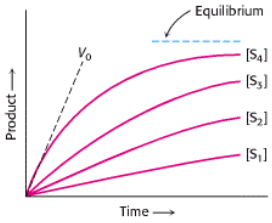
Concepts of Catalysis

- Enzymes have a specific **active site** that recognises and binds to the **substrate** of the reaction.
- Enzymes are **highly specific** (usually **sterospecific**) as a result of the **precise configuration** of **amino acids** at the **active site**.
- The enzyme forms a **complex** with the **substrate** (the **enzyme-substrate complex**). This can then break down to a **free enzyme and product** or back to a **free enzyme and reactant**. The enzyme can also **reverse** the reaction:

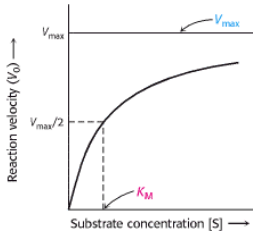
$$\mathbf{E + S \rightleftharpoons ES \rightleftharpoons E + P}$$

Enzyme Kinetics

- The **rate** of an enzyme reaction depends on the **substrate concentration** and on the **enzyme concentration**. We usually assay enzymes in situations where [**Substrate**] >> [**Enzyme**], because the enzyme is **recycled at each step**.
- Under these conditions, we can assume **steady state** – every molecule of **enzyme always** has a molecule of **substrate boud to it**.
- The **rate** of the reaction is the **amount of product formed** per **unit time**. The **initial rate** (V_0) is the **gradient** of the **linear part** of the graph, when **substrate** has **not yet started to run out**:



  Eventually, the **substrate** starts to **run out** and the **rate decreases** – the graph **flattens out**.
- We can then plot these **initial rates** against the **original substrate concentration**



We find that **increasing [S] increases V_0** until the enzyme becomes **saturated with substrate**.

This graph is called the **Michaelis-Menten** plot and is represented by the following equation:

$$V_0 = V_{\max}\frac{[S]}{K_m + [S]}$$

Where K_m is the **Michaelis-Menten constant** and V_{maz} is the **maximum velocity** catalyzed by **this amount of enzyme**:

- At **[S]** $= K_m$, we have that $V_0 = ½ V_{max}$.
- When **[S]** is low (much **smaller** than K_m), $V_0 \approx (V_{\max} / K_m)[S]$ – the reaction is **first order** with respect to **[S]**.
- When **[S]** is very large, $V_0 = V_{max}$ and the reaction is **zero order**.

- K_m is an **intrinsic property of an enzyme working on a particular substrate**. It is defined as follows:

$$K_M = \frac{k_{-1} + k_2}{k_1}$$

Where the *k*s are as follows:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P$$

(The second reaction is considered to be **irreversible** because **[P]** is **so small** in the **initial stages** of the reaction, when V_0 is measured).

- If k_2 is **small** compared to k_1 (which is true in many case), then $K_M = \frac{k_{-1}}{k_1} = K_d$, the **dissociation constant** of the enzyme.
- Thus, if k_2 is **small** compared to k_1, K_m measures the **affinity of the enzyme for its substrate** (the **strength** of the **[ES] complex**). Otherwise, K_m is also a measure of how well the enzyme **catalyses** the **formation of product**.

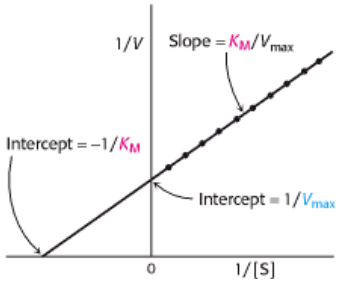
- k_2 is also known as k_{cat} – the **catalytic constant**. Under initial conditions [when there is no reverse reaction], this is also the **turnover number** – the **number of catalytic events per enzyme molecule per second**.
- At V_{max}, all enzyme molecules are [ES] complexes, and so

  $$V_{max} = k_{cat}[\mathrm{E}]$$

  This, however, only applies to the **maximum velocity**. For other states, the constant k_{cat}/K_m is the one that characterises the kinetics of the enzyme, because:

  $$V_0 = \frac{k_{cat}}{K_m}[\mathrm{E}][\mathrm{S}]$$

- The values of K_m and V_{max} can be found from **experimental data** using the **Lineweaver-Burke plot**:



Control of Enzyme Activity

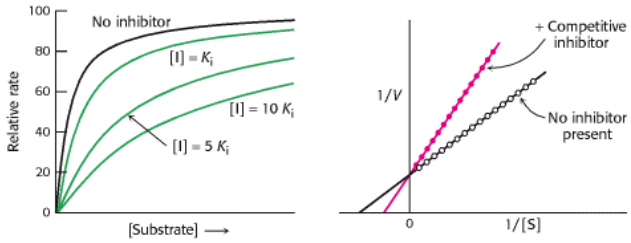
Enzymes can be **activated** or **inhibited** by **interaction** with **molecules** *other than their substrate*. This can be **reversible** (where the inhibitor molecule can freely dissociate) or **irreversible** (where the inhibitor molecule is **tightly bound** to the enzyme). **Rational design** of **inhibitors** is a **major research area** for the **design of new drugs**.

There are three **main types** of **inhibitors**

- **[Reversible] Competitive inhibitors** are **similar in shape** to the **substrate** and can **bind to the active site of enzymes**, preventing the **substrate** from doing so. They **decrease** the **proportion** of **enzyme molecules bound** to a **substrate**.

At any given point, therefore, the effects of a competitive inhibitor can be **relieved** by adding **more substrate**.

Michaelis-Menten kinetics reveal that in the presence of a **competitive inhibitor**, V_{max} is **unchanged**, but K_m <u>increases</u>. This **increase** in **apparent** K_m is consistent with the fact that the effects can be **relieved** by **adding more substrate**, and also makes **intuitive sense** – the fact that **[ES]** is smaller makes it **appear** like the enzyme has **less of an affinity** for the **substrate**.

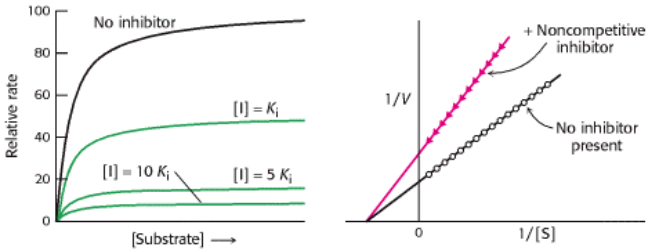


Examples of **competitive inhibitors** include **ibuprofen** and **aspirin**, which bind to **enzyme** that participate in **signalling** during an **inflammatory reponse** and **AZT**, which competes with the **cosubstrate** for **thymine synthesis** (**tetrahydrofolate**) and is used as a drug against **HIV reverse transcriptase**.

- In **noncompetitive inhibition**, the **inhibitor** and **substrate** can bind **simultaneously** to an **enzyme molecule** at **different binding sites**. The **conformational change** in the protein results in a **decrease in the turnover number**.

  This type of inhibition **cannot** be **relieved** in any way.

Michaelis-Menten kinetics reveal that in the presence of a **noncompetitive inhibitor**, K_m is **uncahged**, but V_{max} **decreases**. This is consistent with the idea that such an inhibitor **does not affect the affinity of the enzyme for the substrate**, but only its **turnover number**.



Examples include **deoxycycline**, and **antibiotic** that functions at **low concentration** as an inhibitor of a **proteolytic enzyme**. Some of the **toxic effects** of **lead** might also be caused by its **interaction** with **crucial sulfhydryl groups** in these enzymes [hence, it acts as a **noncompetitive inhibitor**].

- In **uncompetitive inhibition**, the **inhibitor** binds *only* to the **enzyme-substrate complex**. Again, it cannot be relieved.

  Here V_{max} is also **reduced** (because no product is ever formed), but K_m is also **reduced**, because the inhibitor effectively keeps several **ES complexes** "**stuck**" in that state. This means that a **lower concentration** of **S** is needed to get **half** of the **S** in the **ES** state.

  An example is the **herbicide glycophosphate**, which inhibits an **enzyme** in the **biosynthetic pathway** of **aromatic amino acids**.

Irreversible inhibition

When the inhibition is **irreversible**, the **inhibitor reacts** with the enzyme to form a **covalent adduct**. They come in various styles:

- ***Group specific reagents*** react with **specific side-chains** of **reactive amino acids**. Examples include **iodoacetic acid** interacts with **thiol groups** in a **cysteine residue** of **glyceraldehyde 3-phosphate dehydrogenase**.

Similarly, the **nerve agent sarin** interacts with the **active site serine sidechain** in **acetylcholinesterase** (which breaks down **neurotransmitter**).

These inhibitors can often be used to determine whether a **specific amino acid** plays a part in the **catalytic strategy** of the **enzyme**.

- ***Affinity labels*** or ***reactive substrate analogues*** are **structurally similar to the substrate** and **bind to the active site**.
- ***Suicide inhibitors*** are first **slight processed** by the enzyme before they **covalently modify it**.

Some enzymes are also produced as **zymogens**, whereby they are produced in an **inactive form** and then **switched on** when **needed** in **special circumstances**. Examples includes enzymes involved in **blood clotting** and **digestive enzymes**. The zymogens cannot easily be **switched off**, though, unless the enzyme is destroyed.

Genetic control

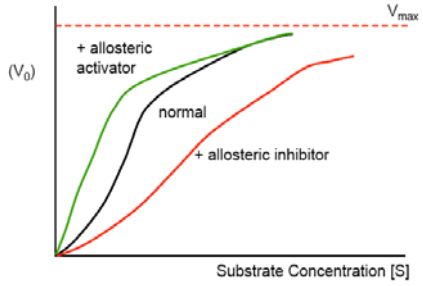
In **higher organisms**, this is a **longer term control mechanism**, but in **bacteria**, it is user for **shorter term mechanisms**. For example, **excess tryptophan** in bacteria **represses** the **production** of all the enzymes in the **tryptophan biosynthesis pathway**.

Compartmentation

Key reactions in **biosynthetic** and **degradative pathways** are usually **distinct** and catalysed by **different enzymes**, which allows a **greater level of control**. In **eukaryotic cells**, **different compartments** allow the maintenance of **competing forwards** and **reverse** reactions in **different comparments**.

Allosteric inhibition

- **Allosteric enzymes** are an important class of enzymes that **do not obey Michaeli-Menten kinetics.** Many of them are **associated** with **regulation of metabolic activity**.
- **They are** oligomeric enzymes (often **dimers** or **timers**) that that contain **distinct regulatory sites** and **multiple functional sites**.
- They **bind** their **substrate cooperatively** – the binding of a substrate at **one site** affects the **activity** of the **others**.
- Other **specific molecules** can also affect the activity of the enzyme.
- The **affinity** of the enzyme for the **substrate** depends on whether the **substrate** is **already bound**. The **Michaeli-Menten** curve for these enzymes is therefore **sigmodial** instead of **hyperbolic**:



- An example is **haemoglobin**, though it's not an enzyme!
- An example is **aspartate transcarbamoylase (ATCase)**. It is the **first step** in the **production of CTP** which is used to make **nucleic acids**. It is **inhibited by CTP** (**feedback inhibition**) and **activated by ATP** [lots of ATP = lots of energy = we want to grow!]
    - It consists of **distinct** and **separable catalytic subunits** and **regulatory subunits**.
    - The enzyme exists in an **R state** (active) and a **T state** (less active). The binding of **CTP favours the T state**, and the binding of **ATP** and the **binding of substrate favours the R state**.
    - The effects are **mediated** by **large changes in quaternary structure**.
- There are **two models** of **allosteric inhibition**:

- ***The concerted model*** – a single equilibrium exists between low and high affinity states, and the binding of substrates favours one or the other side of this equilibrium [the more allosteric activator is bound, the more the relevant side is activated].
- ***The sequential model*** – the binding of a ligand to **one site** in an assembly **increases the binding affinity of neighbouring sites**. **Mixed forms can exist**.