



# Extracellular matrix - Collagen

|             |  |
|-------------|--|
| Course      | <a href="#">Essential Protein Structure and Function</a> |
| Confidence  | Somewhat Confident                                       |
| Next Review | @April 25, 2024  |
| Last Edited | @May 1, 2024 3:41 PM                                     |

## Introduction - Fibrous proteins

### Types of proteins

- Fibrous
  - Collagen, elastin, myosin
- Membrane-bound globular
  - Integral membrnae proteins, transmembrane channels, receptors
- Water-soluble globular
  - Soluble enzymes, oxygen-carriers (Mb, Hb), electron carriers

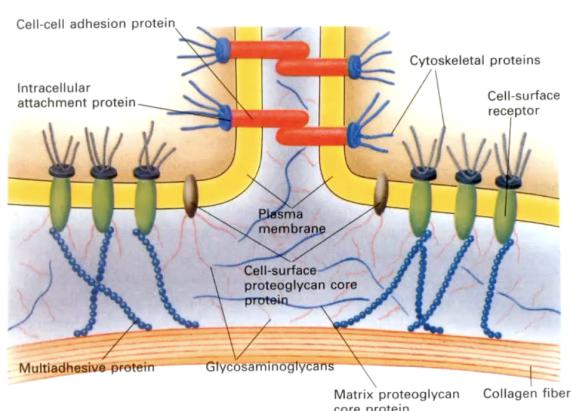
### Differences between globular and fibrous proteins

| Globular   | Fibrous   |
|--|---|
| Compact protein structure  | Extended protein structure  |
| Soluble in water (or lipid bilayer) - thus mobile in instinct                                | <u>Insoluble</u> in water (or lipid bilayer) - immobile                                     |
| Secondary structure is complex with a <u>mixture</u> of a-helix, b-sheet and loop structures | Secondary structure is simple based on <u>one type only</u>                                 |
| Quaternary structure is held together by <u>non-covalent forces</u>                          | Quaternary structure is usually held together by <u>covalent bridges</u>                    |
| Functions in all aspects of metabolism (enzymes, transport, immune protection,               | Functions in <u>structure</u> of the body or cell (tendons, bones, muscle, ligaments, hair, |

| Globular        | Fibrous |
|-----------------|---------|
| hormones, etc.) | skin)   |

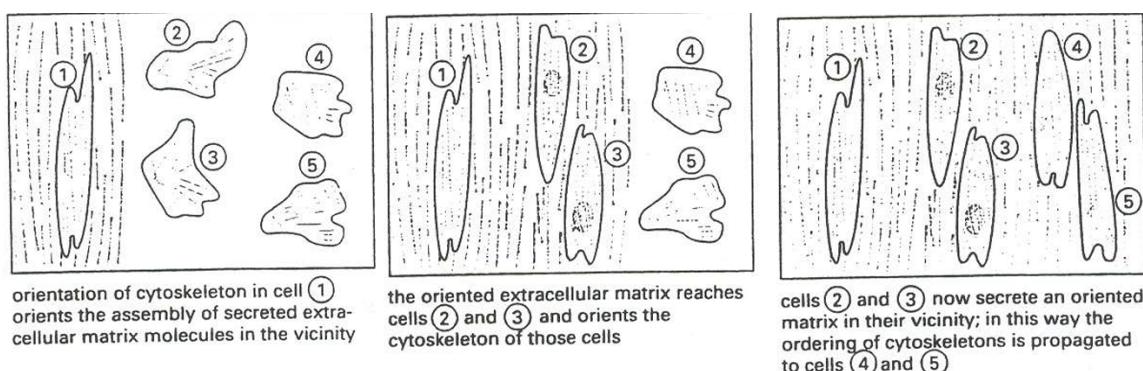
- Fibrous and globular proteins make up about half and half the total protein content in the body
- Functions of fibrous proteins: responsible for the structure of the body
  - Collagen: in connective tissues - relates to tensile strength (tendons, bones)
    - Most abundant protein in the body (skin, bone)
  - Elastin: in connective tissues - relates to elasticity and stretchability (ligaments, lung walls, aorta)
    - Similar to collagen but more highly crosslinked and elastic
  - alpha-Keratin: external protection - toughness (hair, nails, outer skin)
  - Membrane-bound fibrous proteins: spectrin
    - In the cytoskeleton of the RBC membrane
- Function of mixed- fibrous and globular proteins
  - Fibrinogen is a protective protein that seals bleeding wounds
    - Globular part - clot, fibrous part - seal the wound
  - Myosin is responsible for the contractile properties of muscle
    - Globular (myosin head) - enables contraction; fibrous - undergo contraction

## The Extracellular Matrix



- Note the collagen at the bottom: Cells have to link up with the collagen - via multi-adhesive protein
- Contains collagens, elastins, proteoglycans, etc.
  - Depending on the function of the tissue

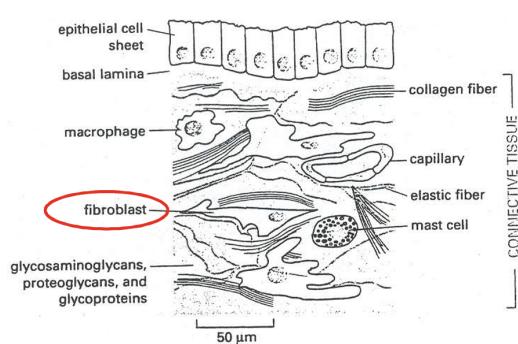
- Enable the cell to associate together property and build up a complete tissue.
- A complex structural entity surrounding and supporting cells that are found within mammalian tissues.
  - The ECM is often referred to as the connective tissue.
  - Composed of 3 major classes of biomolecules
    - Structural proteins: collagen and elastin
    - Specialised proteins: e.g. fibrillin, fibronectin, laminin
    - Proteoglycans: composed of a protein core to which is attached long chains of repeating disaccharide units termed of glycosaminoglycans (GAGs) forming extremely complex high molecular weight components of the ECM
- ECM is not an inert system. It regulates the dynamic behavior of cells by effecting the follow:
  - Growth
  - Migration
  - Proliferation
  - Shape
  - Metabolism
- ECM could propagate order from cell to cell within a tissue



- For simplicity, one cell influences the orientation of its neighboring cells. It is more likely, however, that the cells would naturally affect one

another's orientation.

- ECM consists of fibrous proteins in a hydrated polysaccharide gel. Collectively the connective tissue consists of the ECM plus cells.
- The cell types found in the matrix are:
  - **Fibroblasts** that secrete (generate) the ECM
  - **Chondroblasts** give rise to cartilage
  - **Osteoblasts** and **osteoclasts** are responsible for bone growth



- The connective tissue underlying an epithelial cell sheet.
  - It consists largely of extracellular matrix that is secreted by the fibroblasts

## Collagen structure

- Collagens are the most abundant proteins found in the animal kingdom. It is the major protein comprising the ECM.
  - 1/4 of entire protein content in the body
- There are at least 15 types of collagen.
  - Types I, II, and III are the most abundant and form fibrils of similar structure. Type IV collagen forms a two-dimensional reticulum and is a major component of the basal lamina.
- Collagens are predominantly synthesised by **fibroblasts** but **epithelial cells** also synthesise these proteins.

A single collagen chain

- 2/3 of the residues are proline
  - Can't form an alpha helix
  - Forms structure like the polyproline helix called a **collagen helix**
- Every third residue is glycine

- One surface is relatively flat, without bulk
- This surface twists gently about the axis
- Produces a stable, intertwined triple helix
- Substantial post-translational processing of sidechains leads to strong crosslinks

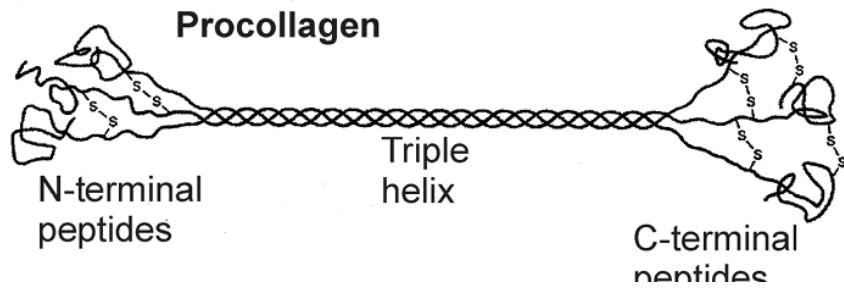
## Collagen assembly

### Summary

1. Synthesised as procollagen which is secreted from the cell
2. Cleaved to tropocollagen by procollagen peptidase
3. Assembly of tropocollagen leads to the collagen fibre
4. Chemical crosslinking of tropocollagen strengthens the fibre

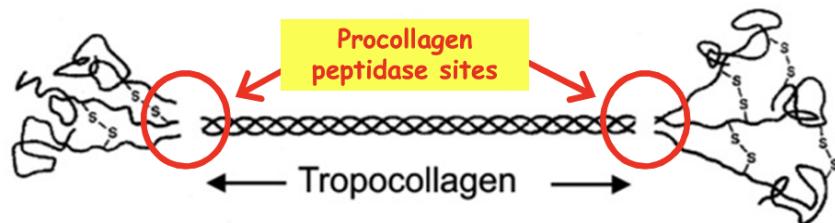
### First stage - Collagen synthesis: formation of procollagen

- Three separate pro- $\alpha$  chains are synthesised inside the cell
- Post-translational modification: Selected Pro and Lys residues are hydroxylated to hydroxyproline (Hyp) and hydroxylysine (Hyl), and secreted hydroxylysine residues are glycosylated
- Three pro- $\alpha$  chains assemble at the **C-terminus end** starting with disulphide bridge formation between the 3 chains (which lines up the three strands).
  - The 3 chains then zip up to form procollagen
  - The N-terminus end is also stabilised by disulphide bridges
- A typical single collagen (Type I) sequence has about 900 residues, with an essential Gly residue at every third position and Pro and Hyp residues are common, hence the sequence is approximately (Pro.Hyp.Gly)<sub>300</sub>.
- Note that collagen is the only protein that fold up backwards from the C terminus
  - Vs. other protein, N terminus starts to fold as being produced by ribosome (co-translational folding)
  - This is necessary for the triple helix to form?



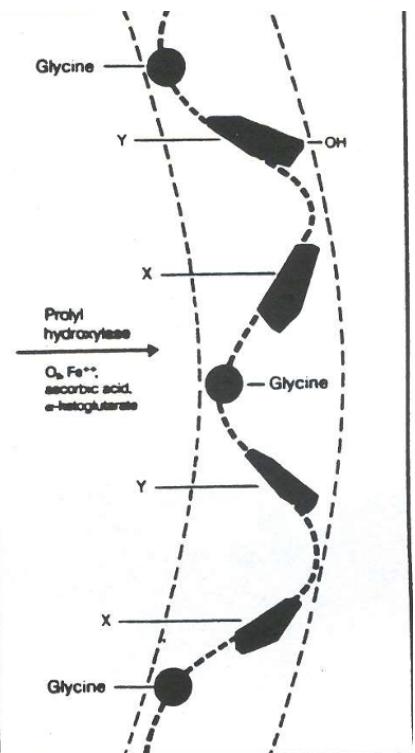
## Second stage - Cleavage of procollagen of tropocollagen

- The N-terminal and C-terminal peptides of procollagen are cleaved by **procollagen peptidase**
- **Tropocollagen** is one of the most elongated proteins known, with a length of 300 nm
- Amino acid composition is 35% Gly, 21% Pro/Hyp, 12% Ala, and 32% for the rest
- Secondary structure of tropocollagen is a triple helix



## Secondary structure - Triple helix

- Collagen is formed from tropocollagen subunits
  - The triple helix in tropocollagen is highly extended and strong



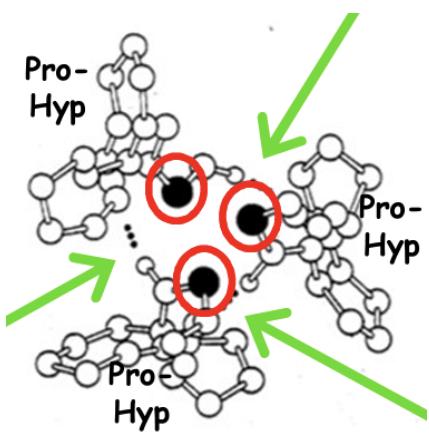
Note this is one of the three chains.

- Features

1. Three separate polypeptide chains arranged as a **left-handed helix** (note that an  $\alpha$ -helix is right-handed)
2. **3.3** residues per turn
3. Each chain forms hydrogen bonds with the other two
  - Thus the whole thing strengthens itself

Sequence is Gly-X-Y

- Note that the Gly is on the left as the helix spirals
- X and Y are often Pro and Hyp
- When spiral the whole structure, glycine is always on the inside



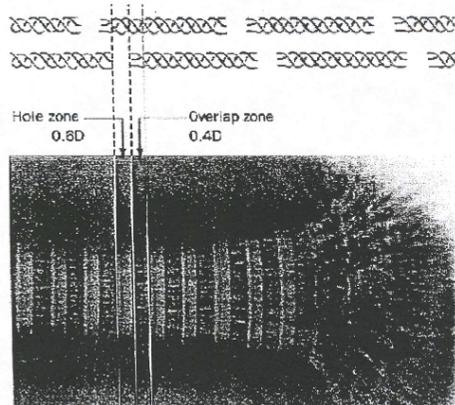
4. Gly residues are essential at every third position in the sequence. The sidechain points towards the centre of the triple helix
  - a. The Gly R group is only -H which enables the three polypeptide chains to be packed together
  - b. The Gly mainchain NH proton forms the H-bond to the CO oxygen of the adjacent polypeptide
    - i. Proline has no NH protons, thus cannot form H bonds.

1. Pro and Hyp residues are "imino acids". The sidechain forms a covalent link with the mainchain N atom, and this ring structure stiffens the triple helix
2. Note that Gly residues are buried while the other two amino acids are outside the triple helix on its surface. The outer ones form inter-triple helix

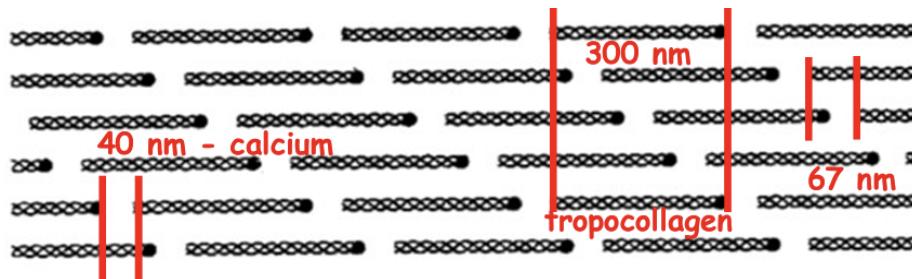
contacts to stabilise the collagen fibre.

### Third stage - Assembly of tropocollagen

#### Electron Micrograph of Fiber

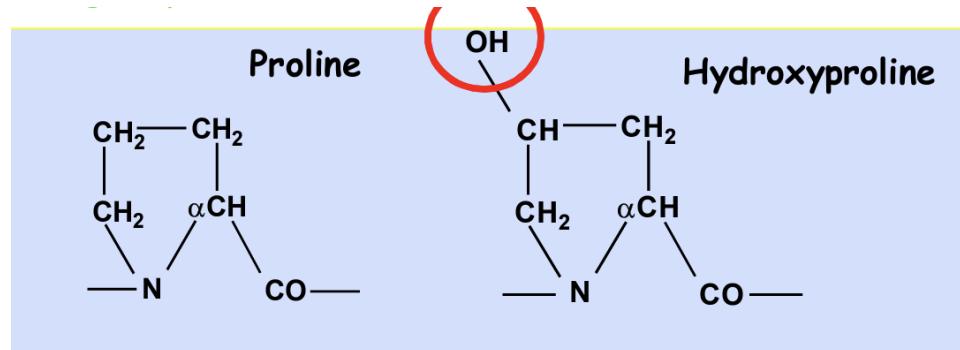


- Regular packing of the collagen molecule
- Banding pattern:
  - Light band, where there is a gap between two successive collagen molecule
  - Dark band, where all overlap



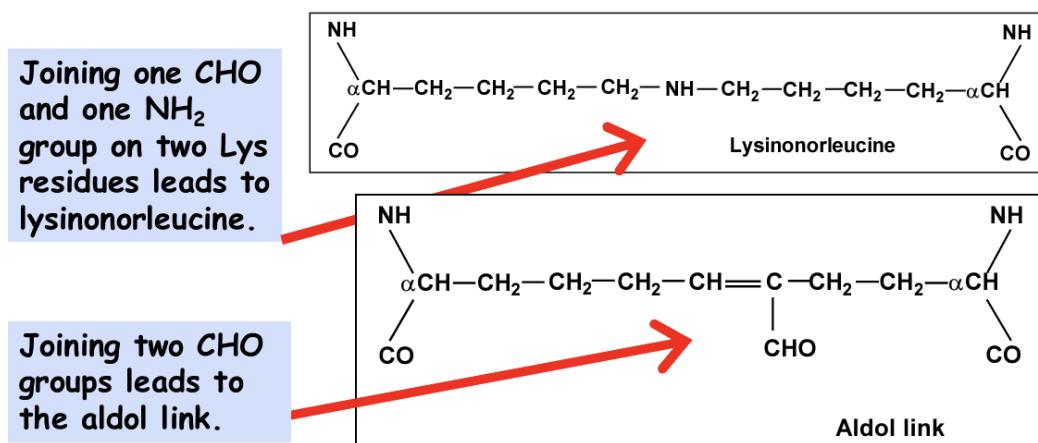
- Electron micrographs of a collagen fibre show periodic cross-striations with a repeat about 67 nm
  - This is accounted for by the formation of a one-quarter staggered array of tropocollagen molecules (i.e., each of which is displaced lengthwise by a quarter of its length)
- There are 40 nm gaps between the successive tropocollagen molecules which is where calcium phosphate is deposited in bone formation
- This assembly forms spontaneously by means of noncovalent hydrogen bond interactions involving the OH group of Hyp
  - Experiment
    - A synthetic polymer made up of Pro.Pro.Gly (i.e., no Hyp) repeats forms a triple helix with a melting (unfolding) temperature of 24C, which is below body temperature of 37C

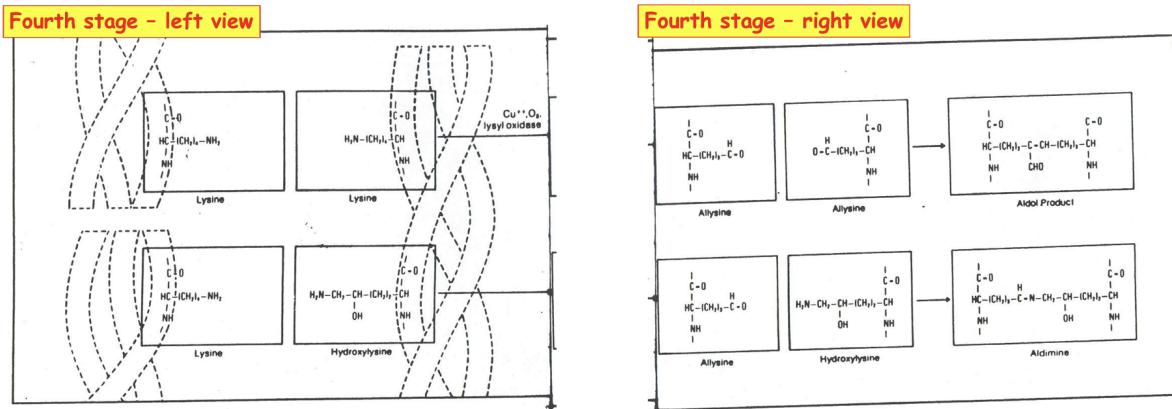
- Synthetic polymer made up of Pro.Hyp.Gly repeats has a melting temperature of 58C
- This OH group is crucial for assembly that it stabilises the triple collagen molecule together. Its absence leads to scurvy.



#### Fourth stage - Crosslinking of collagen fibre

- The assembly of tropocollagen subunits is **strengthened by covalent crosslinks** that are formed between **lysine residues**
  - Instead of cysteine used in globular proteins
- **Lysyl oxidase** generates a reactive aldehyde form of lysine which makes cross-links





- Lysine + Lysine = aldol product
- Lysine + Hydroxylysine = aldimine

## Different types of collagen

- Different collagen types have different alpha chain, either a homotrimer or a heterotrimer
  - Type I, IV, V are heterotrimer
  - Type II, III are homotrimer

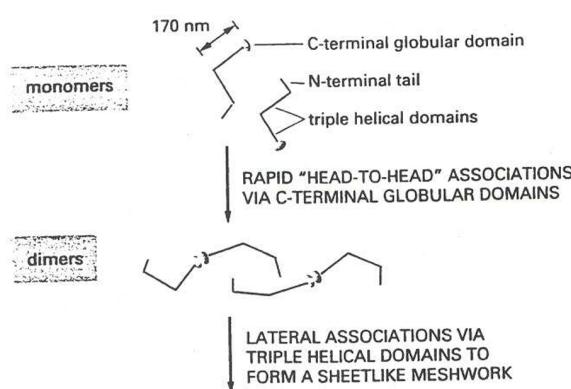
| Type | Polypeptide composition | Distribution                              |
|------|-------------------------|---|
| I    | [a1(I)]2, a2(I)         | Skin, bone, tendon, cornea, blood vessels |
| II   | [a1(II)]3               | Cartilage, intervertebral disk            |
| III  | [a1(III)]3              | Fetal skin, blood vessels                 |
| IV   | [a1(IV)]2, a2(IV)       | Basement membrane                         |
| V    | [a1(V)]2, a2(V)         | Placenta, skin                            |

| Collagen | Anchor      | Proteoglycan                      | Cell-surface receptor | Cells        |
|----------|-------------|-----------------------------------|-----------------------|--------------|
| I        | Fibronectin | Chondroitin and dermatan sulfates | Integrin              | Fibroblasts  |
| II       | Fibronectin | Chondroitin sulfate               | Integrin              | Chondrocytes |

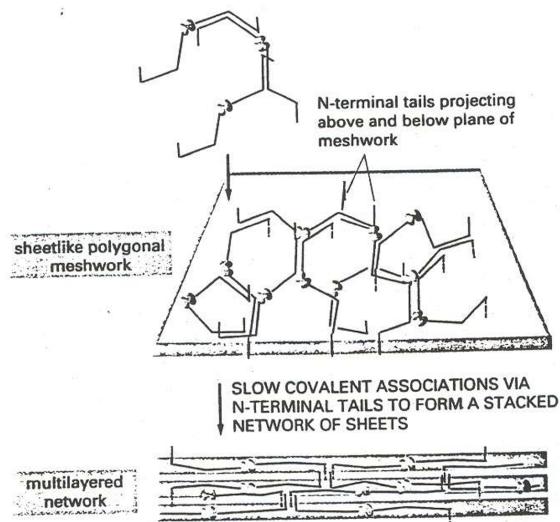
| Collagen | Anchor      | Proteoglycan                | Cell-surface receptor | Cells   |
|----------|-------------|-----------------------------|-----------------------|---|
| III      | Fibronectin | Heparan sulfate and heparin | Integrin              | Quiescent hepatocytes, epithelial; associate fibroblasts          |
| IV       | Laminin     | Heparan sulfate and heparin | Laminin receptors     | All epithelial cells, endothelial cells, regenerating hepatocytes |
| V        | Fibronectin | Heparan sulfate and heparin | Integrin              | Quiescent fibroblasts   |
| VI       | Fibronectin | Heparan sulfate             | Integrin              | Quiescent fibroblasts   |

## Type IV collagen

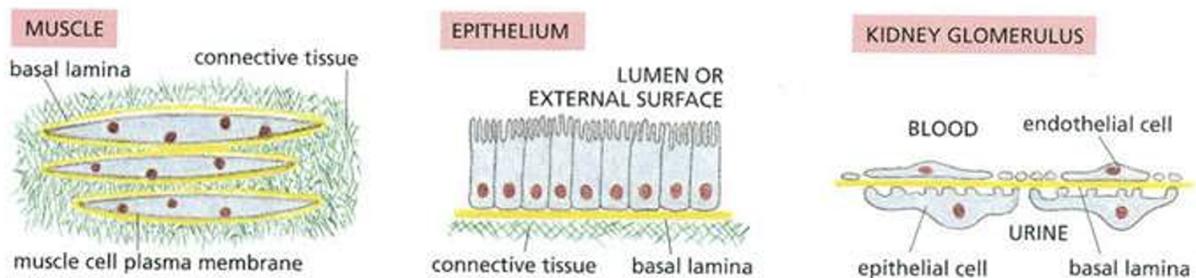
### Assembly of Type IV collagen molecules into a multilayered network



- Rapid “head-to-head” associations via C-terminal globular domains from monomers to form dimers
- Lateral associations of dimers via triple helical domains to form a sheetlike meshwork
  - N-terminal tails projecting above and below plane of meshwork
- Slow covalent associations via N-terminal tails to form a stacked network of sheets

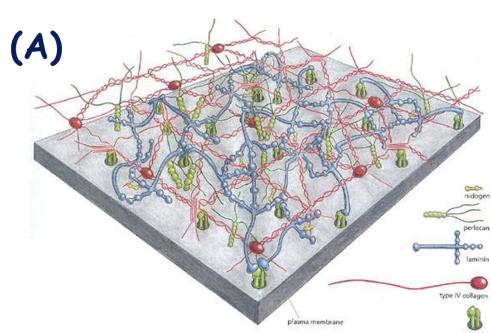


Where does the basal lamina (or basement membrane) occur?

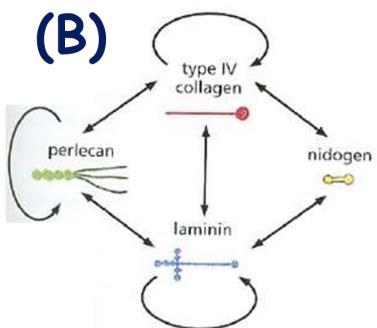


- This is a thin, tough, flexible sheet.
- Surrounds cells (e.g. skeletal muscle) or underlie the epithelium, or are between two cell sheets in the kidney

## Current model of a basal lamina



- The basal lamina (A) is formed by specific interactions between Type IV collagen, laminin, and entactin and the proteoglycan perlecan (B).
- Laminin: Make contact with the Type IV collagen and hold the network with the rest of the basal lamina
  - Make contact with the integrin
- Arrows in B connect molecules that can directly bind to each other.



## Collagen diseases

### Inherited diseases caused by mutant collagen genes

#### Brittle-bone disease (osteogenesis imperfecta)

- Caused by a mutation in one or the other of the two genes whose products are used to make Type I collagen.
- Like all the inherited collagen diseases, this one is inherited as a dominant trait.
  - This is because even though one collagen allele is normal, the assembly of the normal gene product with the mutant product produces defective collagen fibres.
- Results from a mutant collagen gene - a buried glycine residue is mutated to cysteine
  - The triple helix is partially unfolded at the N-terminal end
  - The tropocollagen subunits cannot associate in a regular packing, and collagen fibre formation becomes weaker
- The clinical symptom is “brittle bones” and patients suffer from skeletal deformities

#### Some forms of dwarfism

- Caused by a mutation in the Type II collagen gene.

#### Ehlers-Danlos Syndrome

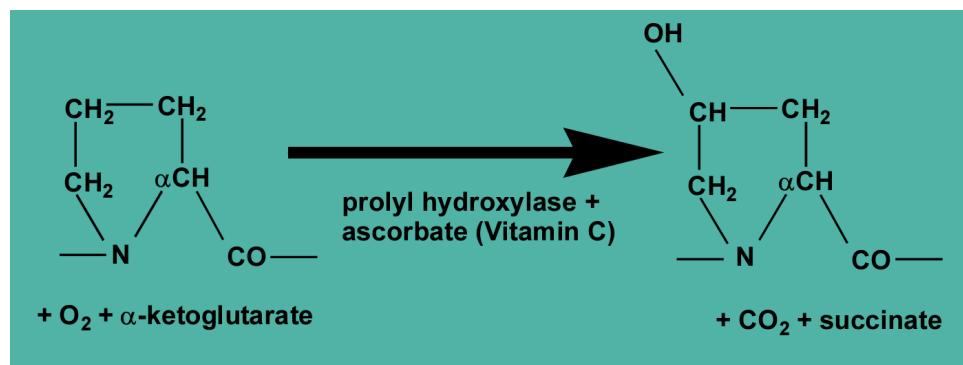
- Rubber-man syndrome
  - Cause by a mutation in a Type I collagen gene. The subject has hyperextensive joints, tendons, and skin

- This inherited disorder represents one type of Ehlers-Danlos Syndrome (EDS)
- Another type of EDS
  - Caused by mutations in the gene for Type III collagen
  - Patients are at risk of rupture of major arteries or the intestine
- Ehlers-Danlos syndrome (Type VII)
  - Results from reduced levels of procollagen peptidase, so procollagen is not fully converted into tropocollagen
  - A high level of procollagen is found in the skin and tendons of patients
  - The 40 nm gaps between tropocollagen molecules become blocked by uncleaved peptides which **prevent lysyl oxidase** from acting on tropocollagen to create the crosslinks
  - Patients suffer from stretchable skin, hypermobile joints and short stature

## Collagen disorders caused by diet

### Scurvy

- Famous disease of medieval Europe - and the Navy
- Caused by a deficiency of vitamin C
  - The sufferer is unable to add hydroxyl (OH) groups to proline to convert it into hydroxyproline

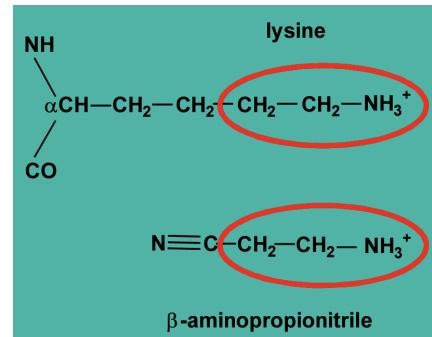


- Hydroxyproline is important for the correct assembly of collagen fibres.
  - It is only formed from proline after procollagen has formed

- This reaction of Pro with **prolyl hydroxylase** requires ascorbate (vitamin C) as a cofactor, which is found in fresh fruit.
- The lack of vitamin C in diet leads to poor collagen fibril formation, so skin lesions develop and blood vessel walls are fragile

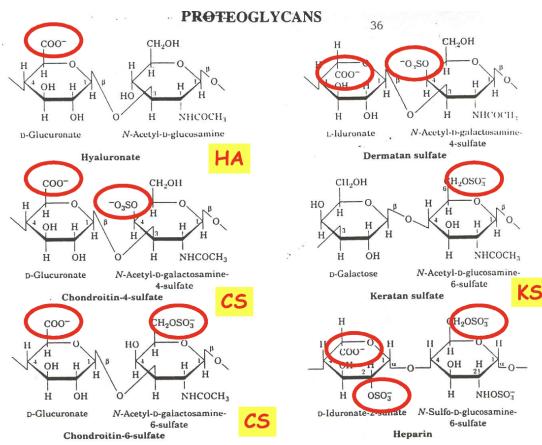
## Lathyrism

- A typical animal disease in the UK caused either by the ingestion of sweet pea seeds with beta-aminopropionitrile (toxic plants) or from a copper deficiency
- Alternatively it results from too low level of **lysyl oxidase** (another form of EDS)
- Beta-aminopropionitrile prevents the **conversion of lysine to the aldehyde form** by irreversibly inhibiting lysyl oxidase.
  - The compound has a very similar structure to the sidechain of lysine residue
- Lysyl oxidase requires copper for full activity, which is why a balanced diet is important for cattle and sheep
- The lack of cross-link formation in collagen brings about lathyrism



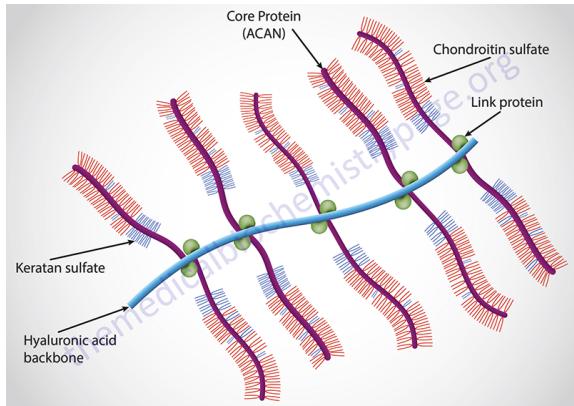
## Proteoglycans

- Responsible for the viscoelastic properties of joints

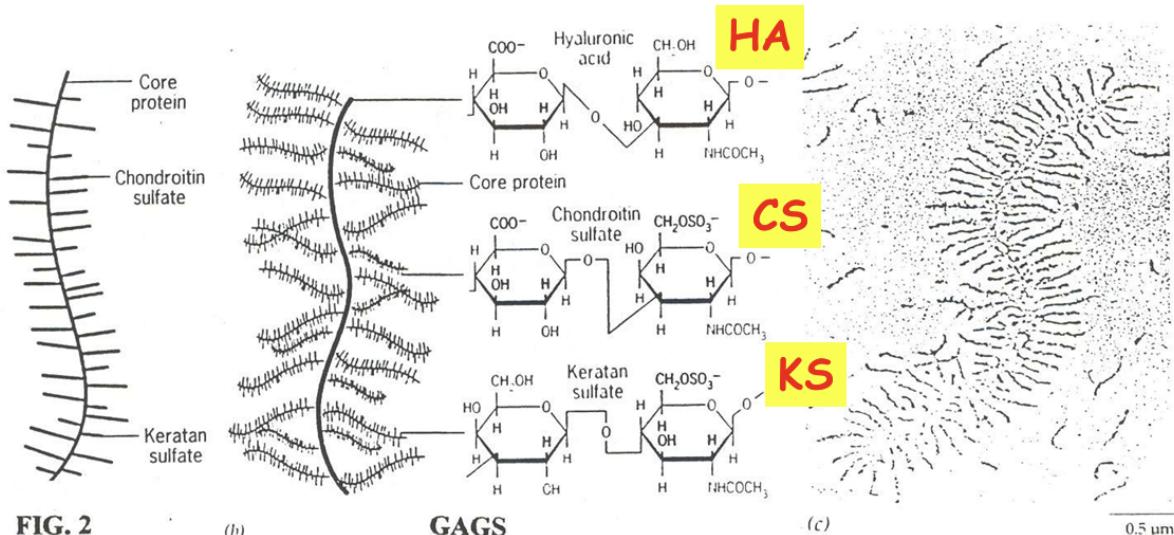


- Glycosaminoglycans (GAGs)
  - Disaccharide repeating units
  - Derivative of an amino sugar
    - -NH.CO.CH<sub>3</sub> or -NH.SO<sub>3</sub>
  - All negatively charged CO<sub>2</sub>- or SO<sub>3</sub>-

- GAGs are incorporated into proteoglycan aggregates
  - Long hyaluronate chain (HA)
  - Bound to a long core protein to which CS and KS are bound
  - Interaction stabilised by link protein
  - These large polyanions bind water and cations - hence achieving its functional viscoelastic properties



## Summary - structure of a proteoglycan aggregate



Schematic representation of a core protein

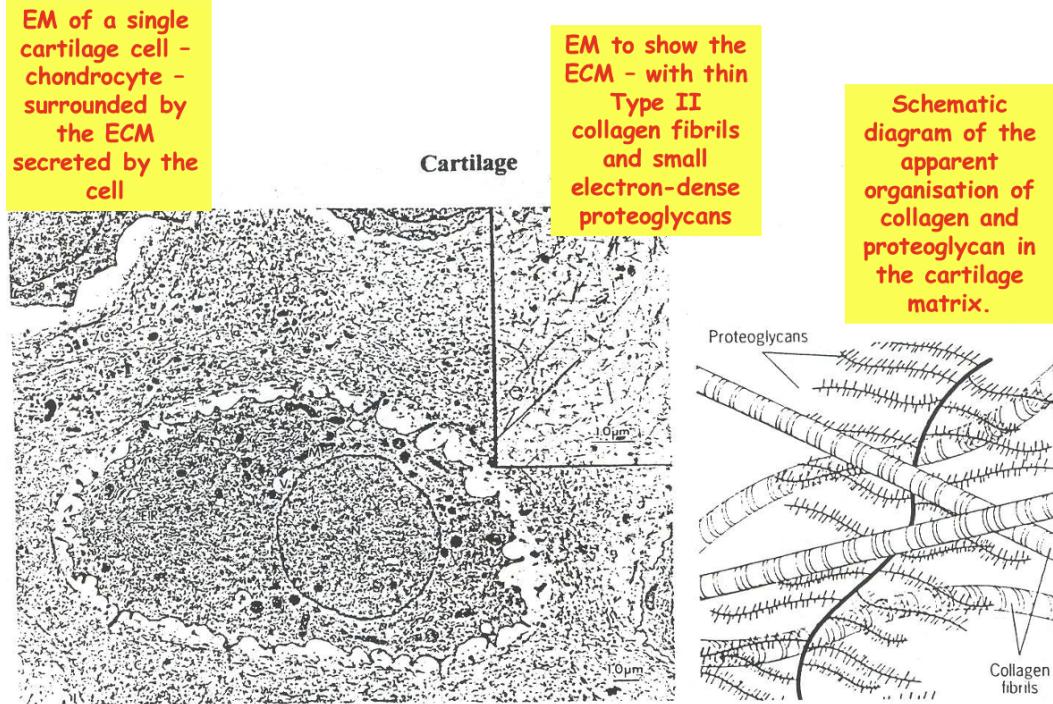
Schematic representation of a single proteoglycan

Note that HA has no sulphates while CS and KS have them

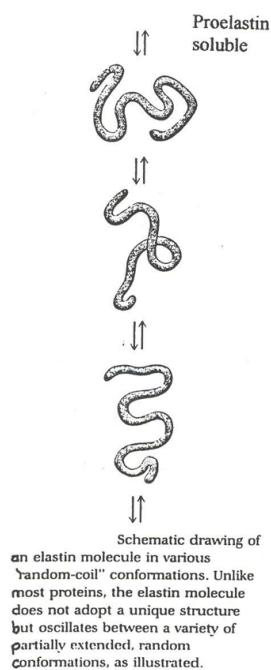
The final assembly looks like a "bottle-brush" in 3-D

- Important in cartilage, providing the smooth surface

## Structure of cartilage



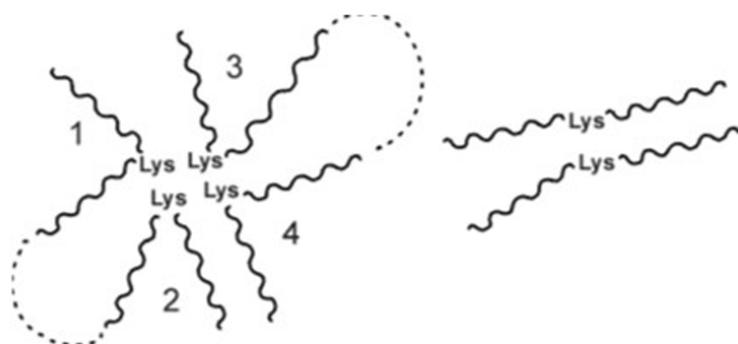
# Elastins

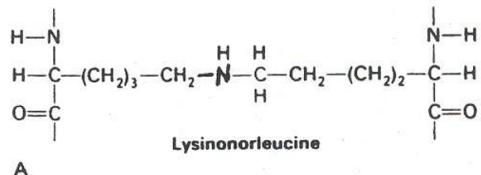


- Major component of elastic fibres
  - Rubber like
  - Found in blood vessels - expand and contract
  - i.e. aorta and ligaments
- Biochemical composition
  - Pro-Gly-Val-Gly-Val-Pro
  - A third of the residues are glycine
  - Rich in proline
  - Rich in small non polar residues

## Formation of elastin

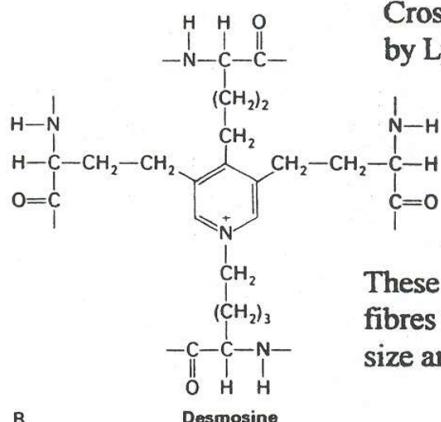
- Synthesised as proelastin
- Converted to tropoelastin (Mr 72000)
- Crosslinking of tropoelastin via lysine residues gives elastin
- Crosslinks: either desmosine (4 Lys) linking 2, 3, or 4 molecules of tropoelastin, or a lysinoleucine (2 Lys) or aldol link (2 Lys) linking 2 tropoelastin molecules





#### Lysinonorleucine

A



Desmosine

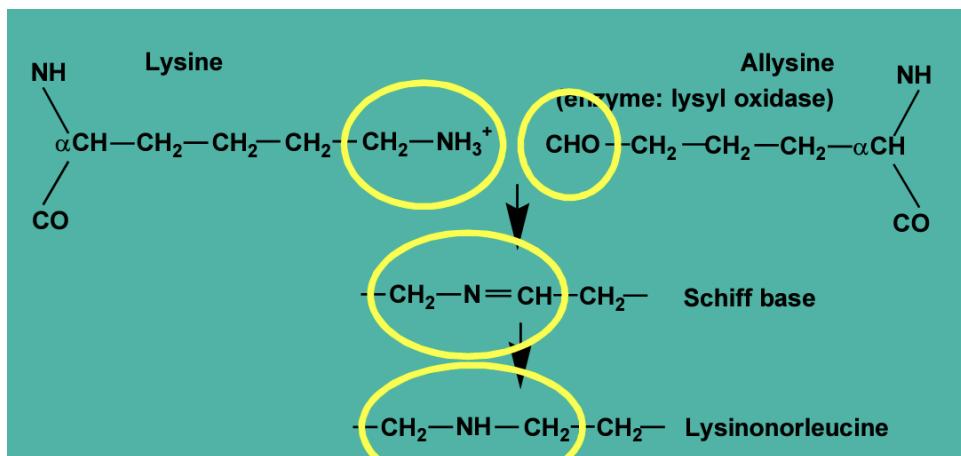
Cross linking performed by Lysyloxidase.

Elastin contains (A) lysinonorleucine and (B) desmosine cross-links. Desmosine is formed from four lysine residues.

These cross-links allow elastin fibres to return to their original size and shape after stretching

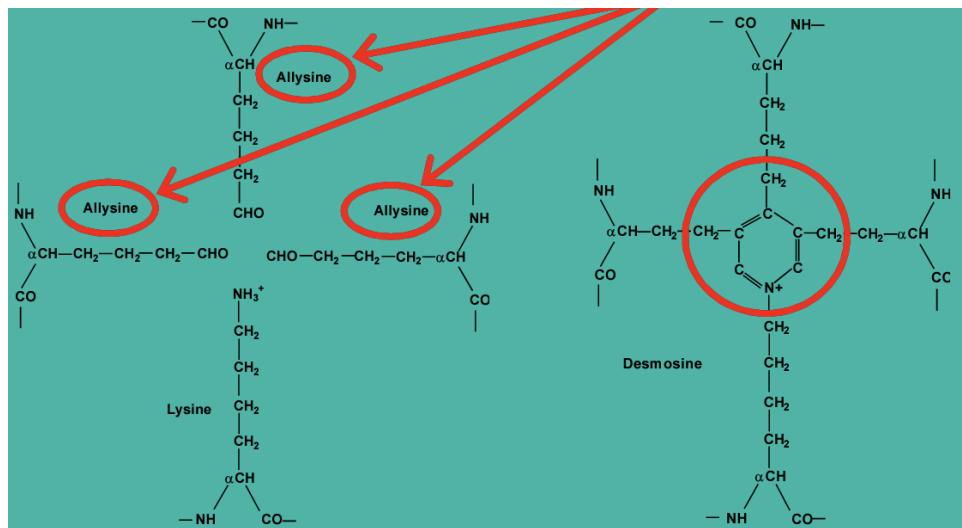
## Lysinonorleucine and the Aldol link

- Two lysines are crosslinked by oxidation (lysyl oxidase)



## Desmosine

- Formed by 4 lysines, 3 of which are oxidised (allysine)

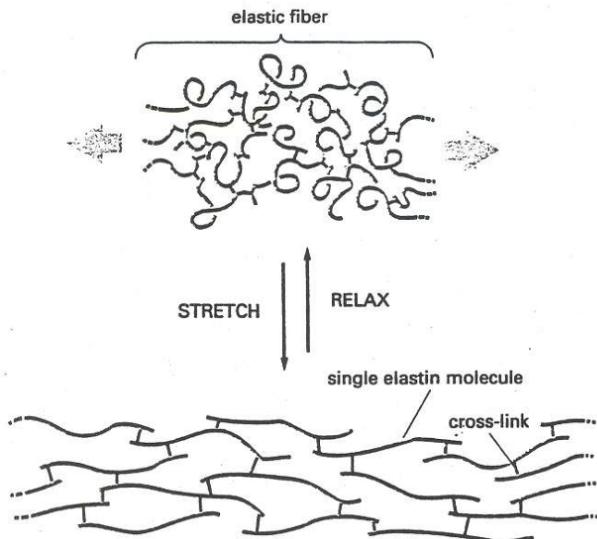


### Amino acid composition of elastin

- 33% Gly
- 10% Pro and Hyp
- 23% Ala
- 13% Val
- Hence 79% if the residues come from 4 amino acids
- There are large **hydrophobic peptides** rich in Ala, Val, Ile and Leu
- As these side chains do not interact with each other by hydrogen bonds, their role appears to allow the elastin network to deform

### Secondary structure of elastin - beta-spiral

- A different type of helix structure from those in the  $\alpha$ -helix or the collagen triple helix is present
- This is able to stretch and relax like a coiled spring
- This is constructed from helices of beta-turns based on the sequence Val.Pro.Gly.Val, and is called the **beta-spiral**



**Stretching a network of elastin molecules.** The molecules are joined together by covalent bonds (indicated in red) to generate a cross-linked network. In the model shown each elastin molecule in the network can expand and contract as a random coil, so that the entire assembly can stretch and recoil like a rubber band.

Polypeptide structure consists of a helical array of  $\beta$ -turns.i.e. a  $\beta$ -spiral.