# B4 Collagen

## Key Notes

**Function and diversity**

Collagen is the name given to a family of structurally related proteins that form strong insoluble fibers. Collagens consist of three polypeptide chains, the identity and distribution of which vary between collagen types. The different types of collagen are found in different locations in the body.

**Biosynthesis: overview**

The collagen polypeptides are post-translationally modified by hydroxylation and glycosylation on transport through the rough endoplasmic reticulum and Golgi. The three polypeptides form the triple-helical procollagen, which is secreted out of the cell. The extension peptides are removed to form tropocollagen, which then aggregates into a microfibril and is covalently cross-linked to form the mature collagen fiber.

**Composition and post-translational modifications**

One-third of the amino acid residues in collagen are Gly, while another quarter are Pro. The hydroxylated amino acids 4-hydroxyproline (Hyp) and 5-hydroxylysine (Hyl) are formed post-translationally by the action of proline hydroxylase and lysine hydroxylase. These $Fe^{2+}$-containing enzymes require ascorbic acid (vitamin C) for activity. In the vitamin C deficiency disease scurvy, collagen does not form correctly due to the inability to hydroxylate Pro and Lys. Hyl residues are often post-translationally modified with carbohydrate.

**Structure**

Collagen contains a repeating tripeptide sequence of Gly–X–Y, where X is often Pro and Y is often Hyp. Each polypeptide in collagen folds into a helix with 3.3 residues per turn. Three polypeptide chains then come together to form a triple-helical cable that is held together by hydrogen bonds between the chains. Every third residue passes through the center of the triple helix, which is so crowded that only Gly is small enough to fit.

**Secretion and aggregation**

The extension peptides on both the N and C termini of the polypeptide chains direct the formation of the triple-helical cable and prevent the premature aggregation of the procollagen molecules within the cell. Following secretion out of the cell, the extension peptides are cleaved off by peptidases, and the resulting tropocollagen molecules aggregate together in a staggered array.

**Cross-links**

Covalent cross-links both between and within the tropocollagen molecules confer strength and rigidity on the collagen fiber. These cross-links are formed between Lys and

| | |
|---|---|
| | its aldehyde derivative allysine. Allysine is derived from Lys by the action of the copper-containing lysyl oxidase, which requires pyridoxal phosphate for activity. |
| **Bone formation** | Hydroxyapatite (calcium phosphate) is deposited in nucleation sites between the ends of tropocollagen molecules as the first step in bone formation. |
| **Related topics** | (B1) Amino acid structure (B2 Protein structure and function) (H4) Protein targeting |

## Function and diversity

**Collagen**, which is present in all multicellular organisms, is not one protein but a family of structurally related proteins. It is the most abundant protein in mammals and is present in most organs of the body, where it serves to hold cells together in discrete units. It is also the major **fibrous element** of skin, bones, tendons, cartilage, blood vessels and teeth. The different collagen proteins have very diverse functions. The extremely hard structures of bones and teeth contain collagen and a calcium phosphate polymer. In tendons, collagen forms rope-like fibers of high tensile strength, while in the skin collagen forms loosely woven fibers that can expand in all directions. The different types of collagen are characterized by **different polypeptide compositions** (Table 1). Each collagen is composed of three polypeptide chains, which may be all identical (as in types II and III) or may be of two different chains (types I, IV and V). A single molecule of type I collagen has a molecular mass of 285 kDa, a width of 1.5 nm and a length of 300 nm.

**Table 1. Types of collagen**

| Type | Polypeptide composition | Distribution |
|---|---|---|
| I | $[\alpha1(I)]_2\ \alpha2(I)$ | Skin, bone, tendon, cornea, blood vessels |
| II | $[\alpha1(II)]_3$ | Cartilage, intervertebral disk |
| III | $[\alpha1(III)]_3$ | Fetal skin, blood vessels |
| IV | $[\alpha1(IV)]_2\ \alpha2(IV)$ | Basement membrane |
| V | $[\alpha1(V)]_2\ \alpha2(V)$ | Placenta, skin |

## Biosynthesis: overview

Like other secreted proteins, collagen polypeptides are synthesized by ribosomes on the rough endoplasmic reticulum (RER; Section H4). The polypeptide chain then passes through the RER and Golgi apparatus before being secreted. Along the way it is **post-translationally modified**: Pro and Lys residues are hydroxylated and carbohydrate is added (Figure 1). Before secretion, three polypeptide chains come together to form a triple-helical structure known as **procollagen**. The procollagen is then secreted into the extracellular spaces of the connective tissue where extensions of the polypeptide chains at both the N and C termini (**extension peptides**) are removed by peptidases to form **tropocollagen** (Figure 1). The tropocollagen molecules aggregate and are extensively **cross-linked** to produce the mature **collagen fiber** (Figure 1).

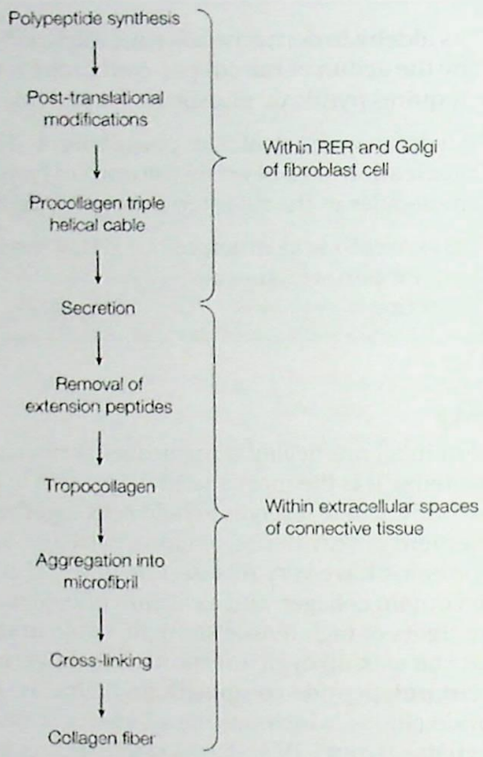


Figure 1. Overview of the biosynthesis of collagen.

## Composition and post-translational modifications

The amino acid composition of collagen is quite distinctive. Nearly *one-third* of its residues are **Gly**, while another *one-quarter* are **Pro**, significantly higher proportions than are found in other proteins. The hydroxylated amino acids **4-hydroxyproline** (**Hyp**) and **5-hydroxylysine** (**Hyl**) (Figure 2) are found exclusively in collagen. These hydroxylated amino acids are formed from the parent amino acid by the action of **proline hydroxylase** and **lysine hydroxylase**, respectively (Figure 2). These enzymes have an $Fe^{2+}$ ion at their active site and require **ascorbic acid** (**vitamin C**) for activity. The ascorbic acid acts as an antioxidant, keeping the $Fe^{2+}$ ion in its reduced state. Proline hydroxylase and lysine hydroxylase are dioxygenases, using a molecule of $O_2$. α-Ketoglutarate, the citric acid cycle intermediate (Section L1), is an obligatory substrate and is converted into succinate during the reaction (Figure 2). Both enzymes will hydroxylate only Pro and Lys residues that are incorporated in a polypeptide chain, and then only when the residue is on the N-terminal side of Gly. Hyp is important in stabilizing the structure of collagen through hydrogen bond formation (see below). In **vitamin C deficiency**, Hyp (and Hyl) are not synthesized, resulting in the weakening of the collagen fibers. This leads to the skin lesions, fragile blood vessels and poor wound healing that are characteristic of the disease **scurvy**.

The other post-translational modification that occurs to collagen is **glycosylation**. In this case, the sugar residues, usually only glucose, galactose and their disaccharides, are attached to the hydroxyl group in the newly formed Hyl residues, rather than to Asn or

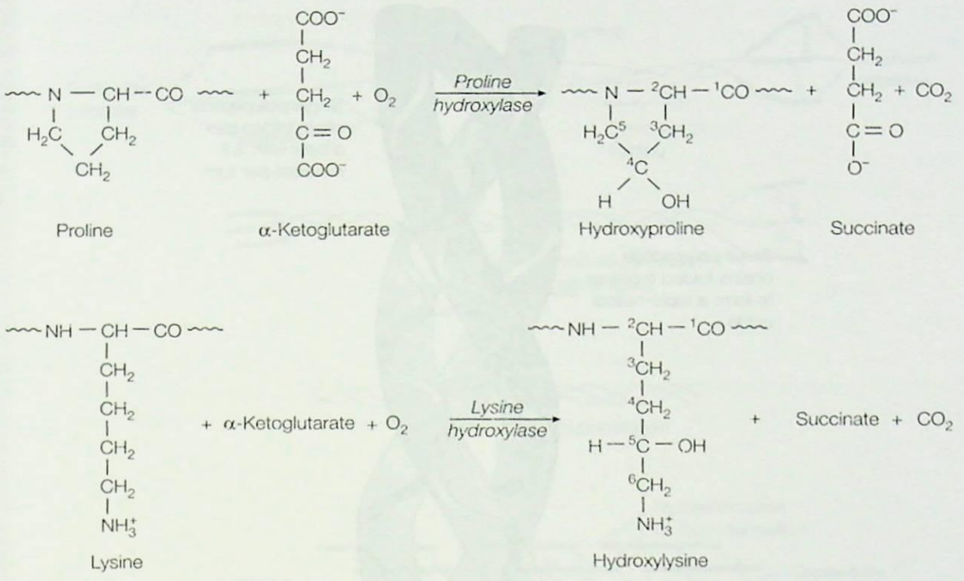


Figure 2. Formation of hydroxyproline and hydroxylysine.

Ser/Thr residues as occurs in the more widespread N- and O-linked glycosylation (Section H5). The amount of attached carbohydrate in collagen varies from 0.4% to 12% by weight depending on the tissue in which it is synthesized.

## Structure

The **primary structure** of each polypeptide in collagen is characterized by a repeating **tripeptide** sequence of **Gly–X–Y** where X is often, but not exclusively, Pro and Y is often Hyp. Each of the three polypeptide chains in collagen is some 1000 residues long and they each fold up into a **helix** that has only 3.3 residues per turn, rather than the 3.6 residues per turn of an α-helix (Section B2). This **secondary structure** is unique to collagen and is often called the **collagen helix**. The three polypeptide chains lie parallel and wind round one another with a slight right-handed, rope-like twist to form a **triple-helical cable** (Figure 3). Every third residue of each polypeptide passes through the center of the triple helix, which is so crowded that only the small side-chain of Gly can fit in. This explains the absolute requirement for Gly at every third residue. The residues in the X and Y positions are located on the outside of the triple-helical cable, where there is room for the bulky side-chains of Pro and other residues. The three polypeptide chains are also staggered so that the Gly residue in one chain is aligned with the X residue in the second and the Y residue in the third. The triple helix is held together by an extensive network of **hydrogen bonds** (Section B2), in particular between the primary amino group of Gly in one helix and the primary carboxyl group of Pro in position X of one of the other helices. In addition, the hydroxyl groups of Hyp residues participate in stabilizing the structure. The relatively inflexible Pro and Hyp also confer rigidity on the collagen structure.

The importance of Gly at every third residue is seen when a **mutation** in the DNA encoding Type I collagen leads to the incorporation of a different amino acid at just one position in

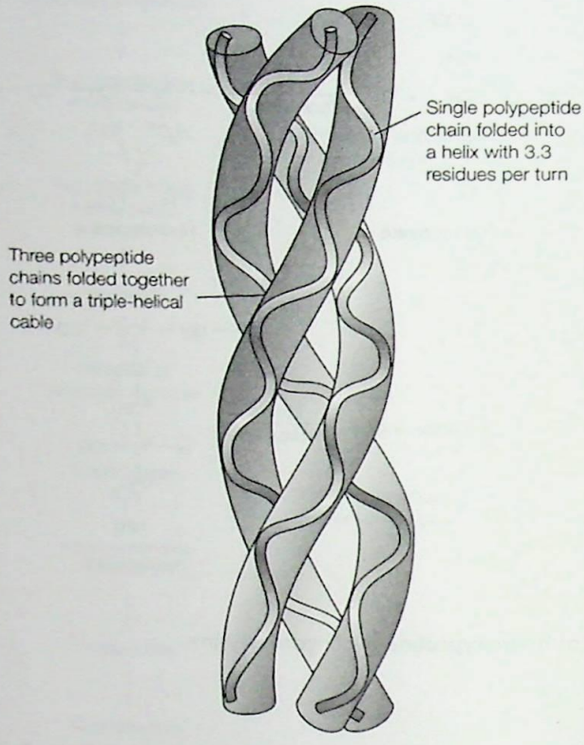


Figure 3. Arrangement of the three polypeptide chains in collagen.

the 1000 residue polypeptide chain. For example, if a mutation leads to the incorporat of Cys instead of Gly, the triple helix is disrupted as the -$CH_2$-SH side-chain of Cys is large to fit in the interior of the triple helix. This leads to a partly unfolded structure t is susceptible to excessive hydroxylation and glycosylation and is not efficiently secre by the fibroblast cells. This, in turn, results in a defective collagen structure that can g rise to **brittle bones** and **skeletal deformities**. A whole spectrum of such mutations known, which cause the production of defective collagen and result in **osteogen imperfecta** (brittle bones).

## Secretion and aggregation

When the collagen polypeptides are synthesized they have additional amino acid r dues (100–300) on both their N and C termini that are absent in the mature collagen fi (Figure 4). These **extension peptides** often contain Cys residues, which are usually abs from the remainder of the polypeptide chain. The extension peptides help to align c rectly the three polypeptides as they come together in the triple helix, a process that m be aided by the formation of disulfide bonds between extension peptides on neighb ing polypeptide chains. The extension peptides also prevent the premature aggregat of the procollagen triple helices within the cell. On **secretion** out of the fibroblast, extension peptides are removed by the action of extracellular **peptidases** (Figure 4). resulting tropocollagen molecules then **aggregate** together in a staggered head-to- arrangement in the collagen fiber (Figure 4).

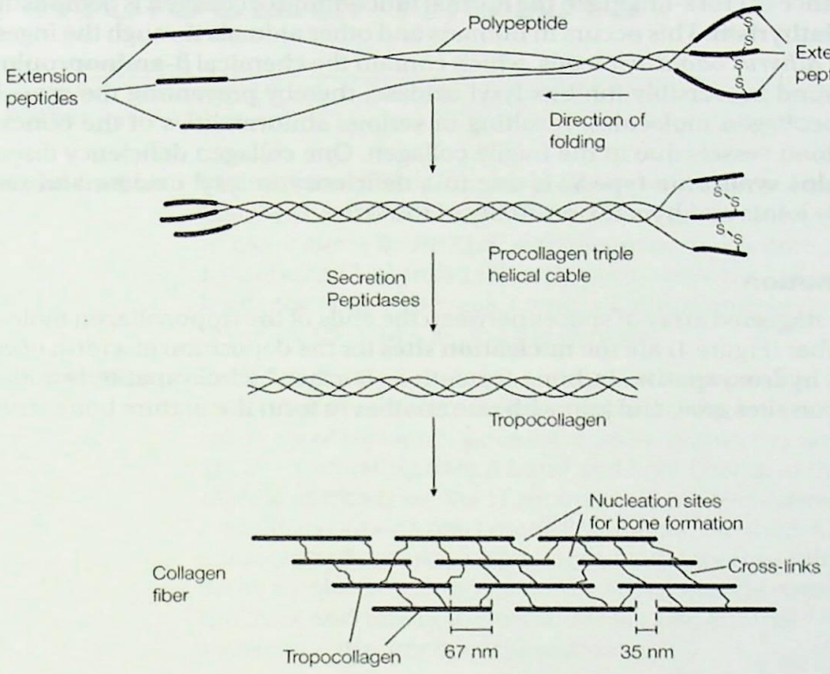


Figure 4. Role of the extension peptides in the folding and secretion of procollagen. Once secreted out of the cell, the extension peptides are removed and the resulting tropocollagen molecules aggregate and are cross-linked to form a fiber.

## Cross-links

The strength and rigidity of a collagen fiber is imparted by **covalent cross-links** both between and within the tropocollagen molecules. As there are few, if any, Cys residues in the final mature collagen, these covalent cross-links are not disulfide bonds as commonly found in proteins, but rather are unique cross-links formed between **Lys** and its aldehyde derivative **allysine**. Allysine residues are formed from Lys by the action of the monooxygenase **lysyl oxidase** (Figure 5). This **copper**-containing enzyme requires the coenzyme **pyridoxal phosphate**, derived from vitamin $B_6$ (Section M2), for activity. The aldehyde group on allysine then reacts spontaneously either with the side-chain amino group of Lys or with other allysine residues on other polypeptide chains to form covalent interchain bonds.

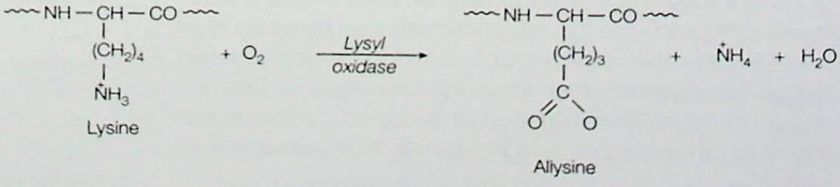


Figure 5. Conversion of lysine to allysine by lysyl oxidase.

The importance of cross-linking to the normal functioning of collagen is demonstrated by the disease **lathyrism**. This occurs in humans and other animals through the ingestion of sweet pea (*Lathyrus odoratus*) seeds, which contain the chemical **β-aminopropionitrile**. This compound irreversibly inhibits lysyl oxidase, thereby preventing the cross-linking of the tropocollagen molecules, resulting in serious abnormalities of the bones, joints and large blood vessels due to the fragile collagen. One collagen deficiency disease, the **Ehlers–Danlos syndrome type V**, is due to a deficiency in lysyl oxidase and results in hypermobile joints and hyperextensibility of the skin.

### Bone formation

The regular staggered array of spaces between the ends of the tropocollagen molecules in a collagen fiber (Figure 4) are the **nucleation sites** for the deposition of a form of **calcium phosphate**, **hydroxyapatite**, in bone formation. Further hydroxyapatite is added until the nucleation sites grow and join with one another to form the mature bone structure.