2. Structure and Composition of Wool

Leslie Jones and George Rogers

Learning objectives

On completion of this topic you should have:
- An understanding of the range of proteins that can be extracted from wool
- A knowledge of how the proteins are obtained and their amino acid composition sequences
- A brief view of the organization of the gene families in human chromosomes
- Knowledge of where the proteins are located in the fibre
- An appreciation of the three-dimensional structure of intermediate filament proteins in general and the several higher orders of organization into keratin intermediate filaments (IFs).
- Understanding of the relationships of keratin IFs and the matrix proteins
- Knowledge of the disulphide and isopeptide bonds, some of their properties and where they are found in the fibre and follicle
- ***An overview of the properties of wool in relation to its protein structure***

Key terms and concepts

- Wool proteins can be solubilized by vigorous chemical procedures such as reduction or oxidation in denaturing conditions
- The three main families of keratin proteins, low-sulphur (intermediate filament, IF) high-sulphur and high glycine/tyrosine proteins (matrix proteins) can be roughly fractionated by traditional solution methods involving salts (eg. zinc acetate) and pH
- The wool proteins can be analysed by gel electrophoresis and amino acid composition
- The genes for virtually all of the human hair keratin proteins and some wool proteins have also been characterised and mapped on chromosomes. For humans, the intermediate filament keratin genes Type I are located with high-sulphur genes on chromosome 17q21.2 and the Type II genes on chromosome 12q13.13
- Protein sequences of human and wool keratin proteins (IF and matrix) have been obtained indirectly by DNA sequencing
- The basic secondary structure in IFs is the a-helix
- Keratin IFs require a Type I and a Type II chain to form a polar dimer, a left-handed coiled coil
- Two dimers are organised into tetramers for which there are three main arrangements of the dimers and the structure is called a protofilament
- Four dimers form a right-handed super-helix or protofibril and four of these are organised into the IF giving a total of 32 protein chains in an IF
- These arrangements can explain the 22nm and 46nm repeats observed by X-ray diffraction
- Keratin IFs (especially those from epidermis rather than hair keratin) can be dissociated into solution and accurately reconstituted into filaments
- The covalent bonds that hold wool proteins together are the disulphide bonds. Isopeptide bonds are also found in fibres but only when a medulla is present because this structure is constituted of a matrix-type protein called trichohyalin that is a major component of the inner root sheath of the follicle
- The location of disulphide bonds with and between amino acid sequences of IFs and matrix proteins in wool is mostly unknown; a factor in elastic behaviour of wool and a research task for the future

Introduction to the topic

Wool is sheephair and belongs to the group of hard mammalian structures that include nails, claws and hooves. These structures are tissues and their cells are differentiated from epithelial cells and contain proteins that are called keratins. Mammalian keratins are different from keratins found in non-mammalian species such as the skin of reptiles, avian claw, beak and feather keratins. The evolutionary relationship of mammalian and non-mammalian keratins is as yet unclear.

Our knowledge of the chemical make-up of mammalian keratin originally came from extensive research carried out on wool in several countries but mainly in the UK, USA, Germany and Australia from about the late 1930's to the end of the 1980's. By the 1980's research publications began to decrease as wool lost it's dominant position in the textile fibre market and basic research gradually diminished. Currently fundamental research is still going on but with a marked orientation to commercial outcomes aimed at maintaining wool in the fibre market. An important change that has occurred in the last 20 years is the application of wool knowledge to research on the structure and mechanisms of growth of human hair in order to understand and treat diseases of hair as well as develop new products for the cosmetic/social role that hair plays.

Wool keratin was once regarded as a highly variable protein product that was difficult to characterise as a protein, although X-ray diffraction studies in the 1940's by biophysicists such as Astbury in Leeds, UK had indicated that there was regularity in the structure. Until the late 1950's wool keratin was too complex to separate into discrete proteins and there were protein chemists who could not accept that wool keratin was a regular product of genes! It is a difficult protein to study because of its insolubility and fibrous nature. It is unlike globular proteins (haemoglobin for example) that could be crystallized and characterised as a single molecule. As a result relatively few protein chemists were interested in research on keratins and wool keratin in particular. The situation changed quite rapidly with the advent of more precise protein chemical methodology and a multidisciplinary approach to investigating wool, from X-ray diffraction and electron microscopy to methods for dissolving the proteins, fractionation by chromatography and electrophoresis and the determination of amino acid composition and sequences.

2.1 The proteins of wool

Isolation and identification of the proteins

The keratin protein chains of the cortex and cuticle components of wool are stabilised mainly by the covalent disulphide bonds. In both the helical and non-helical domains of the chains there are also secondary forces comprising secondary bonds, the ionic bonds between the side chain carboxyl (-COO^-) and amino (–NH_3^+) groups ("Coulombic" forces) and hydrogen bonds between peptide bonds (-C=O - - - - H-N-) that all act cooperatively to reinforce stability. Goddard and Michaelis (1935) were the first to attempt the fractionation of hair keratin proteins and since then modern methods have led to a better understanding of the protein complexity. The first step in solubilizing the proteins is to break the disulphide bonds and then to raise the pH to remove protons from -NH_3^+ groups and to break the hydrogen (H) bonds by the same mechanism because H bonds are formed by an ionic or polar state. The reaction to break the disulphide bond can be either by reduction using compounds such as thioglycollic acid or b-mercaptoethanol or by oxidation using reagents such as peracetic acid. Reduction is the most commonly used and the sulphydryl (-SH) groups formed have to be blocked so that they are prevented from reoxidising and this is usually accomplished by iodoacetic acid that reacts with the –SH groups adding on carboxymethyl groups (Figure 2.1).

Figure 2.1 Reactions of the disulphide bond for solubilizing keratin proteins (K) from wool. Source: Jones and Rogers (2006).

I. The reaction is reversible but with excess of the reducing agent such as thioglycollic acid where R = CH$_2$COOH the reaction is displaced to the right.
II. The reduced keratin proteins are reacted with iodoacetic acid at alkaline pH to give carboxymethyl groups on the sulphur groups.
III. The keratin proteins are oxidised with an oxidising agent, for example peracetic acid CH$_3$COOOH. The disulphide bond is cleaved and sulphonic acid groups are formed.
I. K-S-S-K + 2R-SH → R-S-S-R + 2K-SH
II. K-SH + I-CH$_2$COO$^-$ → K-S-CH$_2$COO$^-$ + HI
III K-S-S-K + 6[0] → 2K-SO$_3^-$

The weak negative charges in alkaline conditions of those groups cause the separation of the protein chains. The alternative procedure of oxidation converts the disulphide groups to sulphonic acid (-SO$_3^-$) groups that are highly acidic and again in alkaline conditions the negative charges result in repulsive separation of the protein chains and do not require derivatisation for stability. The solubilization is in fact a denaturation, changing the proteins from their native configuration into something else.   The solubilizing process by reduction requires the presence of a protein denaturant such as 8M urea in the mixture. About 90% of the wool can be dissolved leaving behind membranous residues and the cell membrane complex (CMC) from both cortex and cuticle. The CMC are insoluble because their proteins are not cross-linked by disulphide bonds but by isopeptide bonds (Rice et al. 1994).

The keratin proteins in the extracted protein mixture can be studied by gel electrophoresis and it is seen (Figure 2.2) that they fall into the three groups that were originally defined by their amino acid compositions namely, low sulphur, high sulphur and high glycine/tyrosine (Table 2.1). They can be fractionated by classical methods of precipitation with salts and pH (Figure 2.3) (Gillespie, 1991).

Further partial purification of these different protein classes was achieved by column chromatography on DEAE-cellulose. It is now known that the low sulphur proteins are derived from the filaments (intermediate filaments, IF) of the cortex and the high sulphur and high glycine/tyrosine proteins from the interfilament matrix (Figures 2.4 A & B) and to a minor extent from the cuticle.

Figure 2.2 2D gel electrophoresis of wool keratin proteins labelled with C[14] iodoacetic acid. Source: Gillespie (1991).
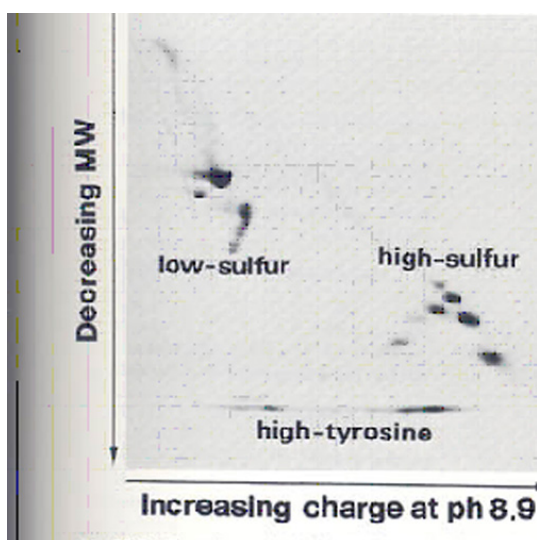
Table 2.1 Amino acid composition of the three classes of wool proteins*.
Source: Jones and Rogers (2006).

Amino acid	Low-sulphur SCMKA major fraction	High sulphur Total	High glycine tyrosine Total
Lysine	4.1	0.6	0.4
Histidine	0.6	0.8	1.1
Arginine	7.9	5.9	5.4
Cysteine (as SCMC*)	6.0	18.9	6.0
Aspartic acid	9.6	3.0	3.3
Threonine	4.8	10.3	3.3
Serine	8.1	12.7	11.9
Glutamic acid	16.9	8.4	0.6
Proline	3.3	12.5	5.3
Glycine	5.2	6.9	27.9
Alanine	7.7	2.9	1.5
Valine	6.4	5.6	2.1
Methionine	0.6	0.0	0.0
Isoleucine	3.8	3.6	0.2
Leucine	10.2	3.9	5.5
Tyrosine	2.7	2.1	15.1
Phenylalanine	2.0	1.9	10.4

* Adapted from Gillespie (1991). Expressed as moles per 100 moles

Figure 2.3 A schema of the protocol used for the initial fractionation of the main keratin protein groups from solubilized wool. Source: Gillespie (1991).

Fractionation of SCMK
To the SCMK, add zinc acetate to D.02M.[2] adjust pH to 5.8-6.0

Filtrate	**Precipitate**
High sulphur proteins (SCMKB)	**Dissolve in 0.02 M sodium citrate. dialyze. centrifuge**

	Supernatant	**Precipitate**
	Adjust to 1% protein. Add 2 vol acetone. 0.3 vol saturated $(NH_4I_2SO_1$	**Type I high-tyrosine proteins**

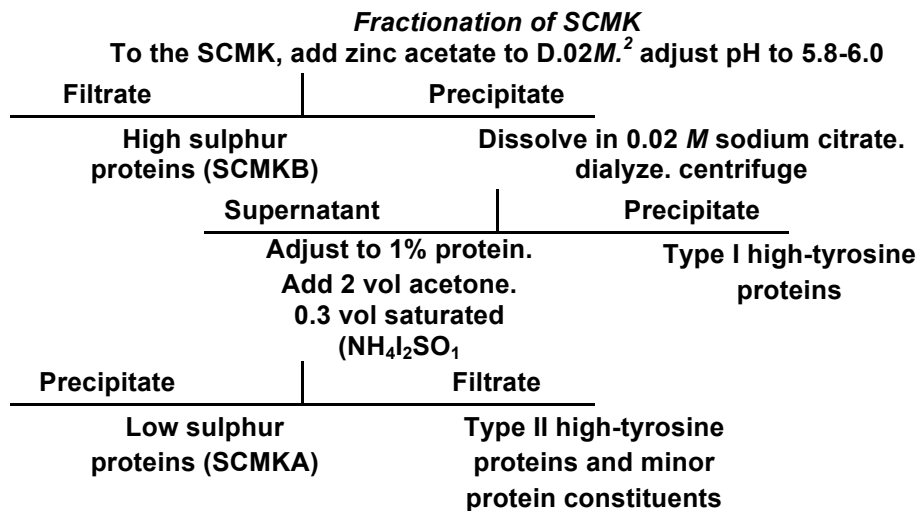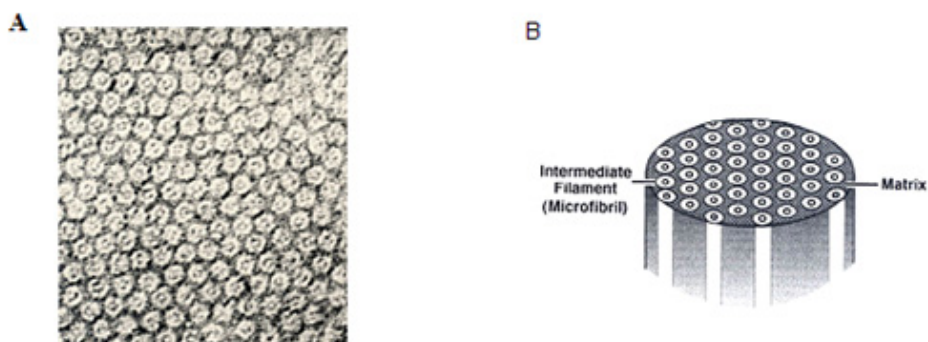| **Precipitate**	**Filtrate**
Low sulphur proteins (SCMKA)	**Type II high-tyrosine proteins and minor protein constituents**

Figure 2.4 A and B   A. TEM of a transverse section of wool cortex at high magnification. The filaments about 8-10 nm in diameter have a central core and are separated by a darker matrix . Cartoon B. is a representation of the TEM. Source: Jones and Rogers (2006).

The wool keratin families in terms of proteins and genes

It was apparent from gel electrophoresis and column chromatography that the three protein classes were not only impure but themselves consisted of many proteins of similar size and amino acid composition. Thus up to the 1990's it was known that the filaments, the IFs, were comprised of four chains that could be divided into two types Type I and Type II on the basis that they were sufficiently different in composition, molecular size, ~55kDa (kilodaltons) and ~59kDa and acidic and basic/neutral charge, respectively. The matrix proteins that "cement" the IFs together in aggregates called macrofibrils are collectively known as keratin-associated proteins or KAPs. The high sulphur (HS) proteins were divided into two groups one of which had proteins with cysteine contents (measured as S-carboxymethyl cysteine or SCMC) >30mole% and were called ultrahigh sulphur or UHS proteins. Each of these was shown by column chromatography to consist of many related chains with the total being around 50 in the family. In the high glycine/tyrosine (HGT) group two classes were distinguished on the basis of their glycine and tyrosine contents, Type I being smaller in size (<10kD) and having higher contents of both glycine and tyrosine than Type II. Electrophoresis of the glycine/tyrosine proteins indicated that there were at least 30 related protein chains. Protein analyses of Merino wool indicate that the cortex consists of about 20% HS+UHS KAPs, 12% HGT KAPs and 68% IFs (Gillespie 1991).

The more definitive evidence for large families of proteins in all three classes came from the isolation and characterisation of the wool keratin genes. Following those studies more extensive studies of the equivalent human genes have been published.

Single genes representative of the three classes of keratin proteins have been cloned, and sequenced and the amino acid sequences derived from the open reading frames (Powell and Rogers, 1997). It is interesting that the first amino acid sequence of a wool IF protein was obtained in 1982 after some 30 years research whereas by DNA cloning methods the first wool keratin protein sequences derived from gene sequences began to appear in the same year (Ward et al. 1982) after the first mRNA was isolated (Ward and Kasmarik, 1980). For several years protein sequencing of sheep keratin and DNA sequencing of the genes proceeded conjointly but protein sequencing ceased by 1989 and the gene counterpart continued for a further 10 years. What appears to be a complete documentation of all the hair keratin gene sequences has been achieved through the studies of a Heidelberg (Germany) group led by Jurgen Schweitzer.

Examples of the amino acid sequences of three wool proteins obtained from DNA sequencing are shown in Figure 2.5A, B, C.

Figure 2.5A The single letter code amino acid sequence of a Type II wool keratin IF. The 480 residues with a mean residue mass of 110 give a molecular mass of 53,800 kDa that accords with measurement from mobility on a gel. Note the higher number of cysteine residues at the N and C terminal ends of the molecule. Source: http://www.ncbi.nlm.nih. gov/entrez/viewer.fcgi?db=nucleotide&val=313243

```
  1   MTCGSACFSS   RSGVRNSSSS   VTTQRYCPGR   TFSCGSAGGS   RPSRCCVTAA
PYRGVSCYRG
 61   LTGGFSSRSV   CGGSRAGSSS   RSFGYRSGSP   SPPCITSVSV   NQSLLKPLNL
EIDHNAQRVK
121   NQEKEQIKSL   NSKFAAFIDK   VRFLEQQNKL   LETKWQLYQN   QRSCESNLGP
LFNGYIETLR
181   REAERVEADT   GRLASELNHV   QEVLEGYKKK   YEEEIALKTT   AENEFVKIKQ
EVNHVYVSKG
241 DLEANAHSLV EEVGFLKTLY EEELRVIQAH ISDTSVIVKM DNSRYLNMDS IVAEIKAHYD
301 EIASRSRAEA ESWYRSKYEE IKATVNRHGE TLRRTKEEIN ELNRLIQRLT AEIENAKSQN
361   SKLEAAVTQA   EQQGEAALND   ARGKLAGLEK   ALQKAKQDMA   SLVKQYQEVM
SSKLGLDIEI
421   ATYRRLLEGE   EQRLCEGISA   VNVSVSSSRG   GVVCGDLSGT   RTCGISSYGV
GACGSSYKKC
```

Figure 2.5B The single letter code sequence of a wool high sulphur protein that is one of the component proteins of the interfilament matrix. The molecule has 183 amino acid residues and a mass of 20,100 kDa. Note the concentration of cysteine and serine residues throughout the molecule as well as a repeated pentapeptide sequence of CCVPV/A (highlighted). Source:http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db= protein&val=57546909

```
  1 MGCSGCSGGC GSSCGGCGSR CGGCSSSCCV PVCCCKPVCC CVPACSCSSC
  GKGGCGSSCG
 61 GSKGGCGSCG GSKGGCGSCG GCGSSCCKPV CCCVPACSCS SCGKGGCGSC
GGSKGGCGSC
121 GGSKGGCGSC GGCGSGCGSS CCVPVCCCVP ACSCSSCGKG GCGSCGCSQS
SCCVPVCCQR
181 KI
```

Figure 2.5C. The single letter code sequence of one of the wool glycine/tyrosine rich protein present in the interfilamentous matrix. The 83 residue molecule has a mass of 9100 kDA and has three tyrosine residues at the C terminal end. Source: http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=protein&val=547810

```
1   MCGYYGNYYG   GLGCGSYSYG   GLGCGYGSCY   GSGFRRLGCG   YGCGYGYGSR
SLCGSGYGYG
 61 SRSLCGSGYG CGSGYGSGFG YYY
```

The genes of the keratin IFs have 6 introns in Type I and 8 introns in Type II. Interestingly there are no introns in the genes encoding the HS and the HGT proteins.

Over recent years it has been shown (Hesse et al. 2004) that in the human genome the Type I and Type II genes for both hair keratin and epidermal keratin are arranged in two clusters on separate chromosomes respectively 17q21 and 12q13 (Figure 2.6A and B).

Figure 2.6 Source: Simplified from Hesse et al.(2004)
A. The total complement of Type I keratin IF genes are clustered (green) in a domain of 977kb on human chromosome 17q21.2. The genes for the high sulphur KAPs are interposed (red).

11 IF   29 KAP genes   16 IF
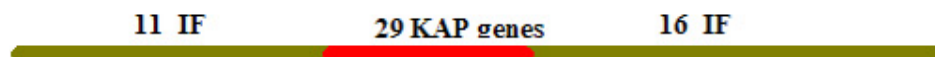
B. The total complement of Type II keratin IF genes are clustered (green) in a domain of 783kb on human chromosome 12q13.13.

27 IF genes

The Type I and II keratin genes are mainly oriented in one direction

There are 27 functional Type I genes (and 4 pseudogenes*) and 27 functional Type II genes (and 5 pseudogenes*) organised on the chromosomes with most of them oriented in the same direction, in accord with them having been produced by gene duplication. The number of genes places them among the 100 largest gene families so far discovered. In human hair keratin there are 11 Type I and a similar number of Type II genes. The genes encoding the high sulphur (HS) keratin associated proteins (KAPs) are interposed as a group of 29 genes divisible into 7 individual families within the Type I clusters on chromosome 17q21 (Hesse et al., 2004) (Figure 6A). The high glycine/tyrosine proteins are encoded by 17 genes that are found on a different human chromosome 21q22.1. The large number of genes for IF proteins and KAP proteins is believed to have arisen by the process of gene duplication with some mutations that produced variations in amino acid sequences. Why there should be so many genes and proteins for a structural purpose is not known but the population of these genes might allow a rapid expression of the proteins over the short distance of keratinisation in the follicle.

The number of wool keratin genes and their chromosomal arrangements in sheep would be expected to be similar to the human because the genes of mouse and rat are syntenic with human. Human hair keratin studies will continue to aid our understanding of the complexity of wool at the gene and protein level.

The molecular structure of wool proteins: the filaments and matrix

In the 1980's it was recognised that in animal cells there are three types of structures forming the so-called cytoskeleton – actin, microtubules and intermediate filaments. They were called intermediate because their diameter averages 10 nm in diameter compared with actin filaments at 8 nm and microtubules at 25 nm. The definitive visualisation in the electron microscope of hair keratin filaments and their organisation in hair was in fact first obtained for wool (Filshie and Rogers, 1961; Rogers, 1959a & b). At that time it was not realised that the keratin filaments belonged to a sub-group of a super family of IFs found in all tissues of the body where the different families were constructed from different proteins but all with a similar secondary and tertiary structure. In a sense the cortical cells of wool and hair fibres contain an elaborate cytoskeleton composed solely of over-expressed keratin IFs. All of the IFs of the superfamily are characterised by their shared tertiary structure of coiled-coils of highly a-helical chains despite amino acid sequence differences. The a-helix is a consequence of the primary structure (amino acid sequence) of a protein chain in which the N-H group of amino acid "n" can establish a hydrogen bond with the C=O group of amino acid n+4 along the sequence (Figure 2.7).

Figure 2.7. In the right-handed a-helix there are 3.5 residues per turn (the pitch) and H-bonds (dashed lines) are between CO and NH groups of amino acid residues n+4 along the pitch. Stability of the helix is contributed to by the H-bonds acting cooperatively. Source: http://www.answers.com/topic/alpha-helix (public domain).
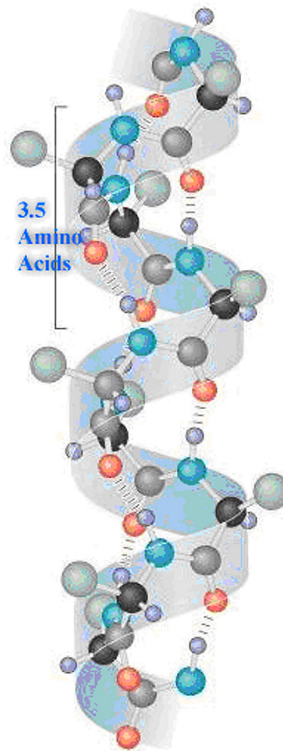
By comparing several chemical properties it was demonstrated that the proteins of filaments from epidermal cells (often referred to as cytokeratins) belonged to the superfamily and using antibodies to the epidermal proteins of the skin it was shown that hair keratins were a sub-group of the family. The IF superfamily is summarised in Table 2.2.

Table 2.2 Classification of IF Sequences
Note that there are two sequence types for the keratins whereas other IF groups have only one type of sequence and single genes. Source: http://hopkinsmedicine.org/ CoulombeLabPage/intermediate.htm (modified table).

Type	Protein	No genes	Mol Weight	Cell type
Type I Type II	Keratins Keratins	27 27	40-57 kDa 53-70 kDa	Epithelia Epithelia
Type III	Vimentin Desmin GFAP Peripherin	1 1 1 1	57 kDa 54 kDa 50 kDa 57 kDa	Mesenchymal cells Muscle Glia, astrocytes PNS neurons
Type IV	NF-L NF-M NF-H [alpha]-internexin	1 1 1 1	62 kDa 102 kDa 110 kDa 66 kDa	CNS neurons CNS neurons CNS neurons CNS neurons
Type V	Yamin A / C Lamins B_1, B_2	1 1 ea	70/63 kDa 67/72 kDa	Nucleus - Mature cells Nucleus - Develop cells
* Lens	Phakinin/CP49 Filensin	1 1	49 kDa 95 kDa	Lens Lens

The unravelling of the substructure of keratin IFs was aided by chemical studies of wool keratin in which it was shown that a helical tetramer could be isolated from wool and this led to the idea, later supported from other studies, of four chains as a basic structural feature that is discussed below. Our understanding of the molecular organisation of keratin IFs in wool has come mainly from the investigation of epidermal keratins or cytokeratins in the first place but the organisation is similar for hair and wool keratins with modifications from the sequences known from gene sequencing.

A central feature that distinguishes keratins from other IFs of the superfamily is that they are composed of two protein chains that have minimal homology in their amino acid sequences. That is, they are heteropolymers compared to vimentin and desmin that are homopolymeric that have the same basic coiled-coil structure but the chains are identical (neurofilaments can also be heteropolymers from combinations of the three types of chains L, M and H). The basic coiled-coil or dimer unit of IFs is about 500 amino acids long (molecular size is ~55kD) and takes up that configuration because the surfaces of a-helices are "knobbly" from the projecting side chains that fit together in a knob/hole fashion. The dimer is especially stabilised when every fourth amino acid residue out of seven in the a-helices is non-polar, a primary structural feature of the keratins. The parallel chain coiled-coil (Figure 2.8) is tripartite with non-helical amino (head) and carboxyl (tail) domains flanking a central a-helical domain about 300 amino acid residues in length. The domain is interrupted by non-helical linker (L) sequences giving the four helical domains 1A, 1B, 2A, 2B.

Figure 2.8 Two keratin chains, one Type I and one Type II form the coiled coil *and are oriented in parallel* and hence this is a polar structure. It should be noted that the coiling is left-handed (to see this, let your forefinger traverse along the chains – the thumb rotates anti-clockwise). Source: Parry and Steinert, (1995).
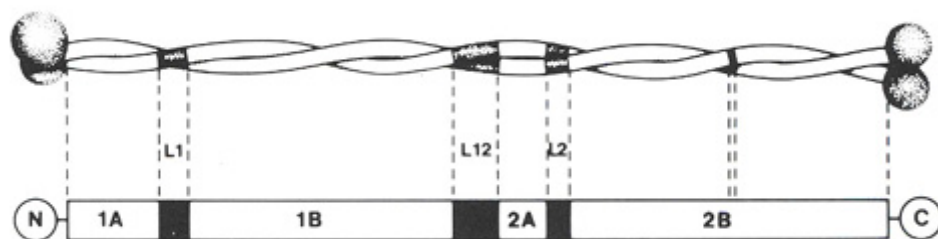
The next level of complexity is the aggregation of the dimers into tetramers (the isolation of a tetramer unit from wool was mentioned earlier) and then further to an intermediate filament. The four chain structures are sometimes referred to as protofilaments although they may not be clearly distinct structures in the intact IF. There are several possible ways in which this can occur and these are shown in Figure 9. Of importance to note is that the dimers are oriented in opposite directions in the tetramer so that polarity is cancelled out and the IF is a *non-polar* structure (cf. actin filaments and microfibrils of the cytoskeleton are *polar* structures).

Figure 2.9 Diagrammatic representation of the ways in which dimers of cytokeratin chains (single dimer, far left) can align in several ways to form into tetrameric units, the protofilaments. The units are aligned out of register in A_{11} and A_{22} , in register in A_{12} and end to end in A_{CN}.   The bars indicate lysine side chains that are close enough to be artificially cross-linked. Source: Parry and Steinert, (1995).
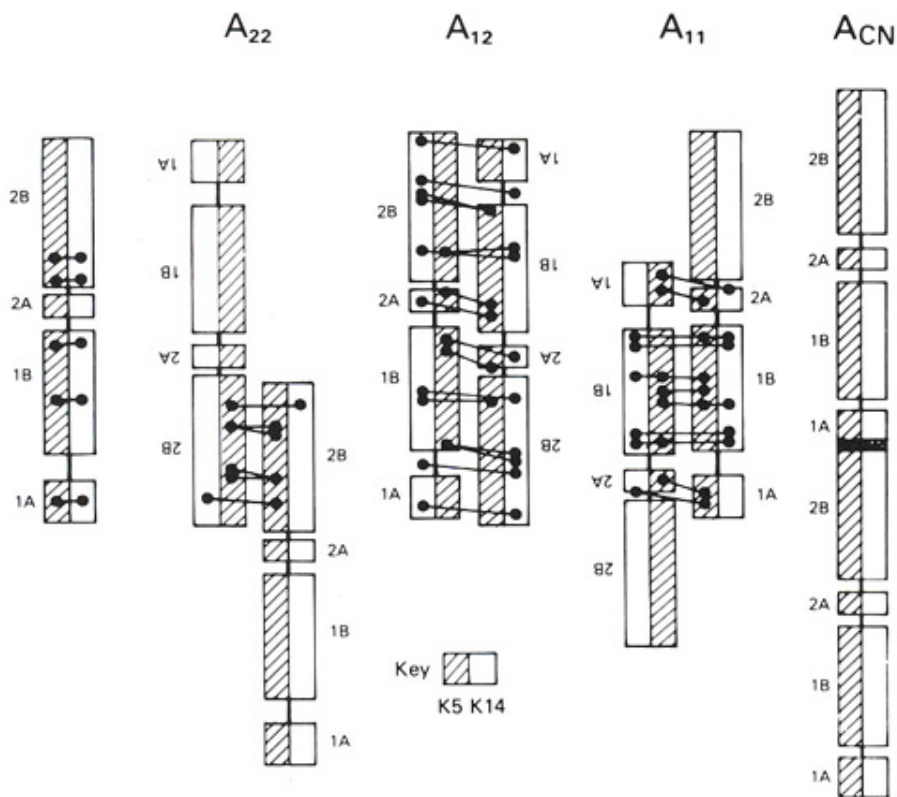
If the four chain arrangement in A_{12} is considered it can be seen that when 16 dimers are aligned into a unit of 32 chains the arrangement (Figure 2.10) has structural repeats that are detectable by electron microscopy and X-ray diffraction.

The current overall view of the keratin intermediate filament is shown in Figure 2.11.   The protofibrils are oriented in a slow left-handed helical configuration in the intermediate filament shown on the right of the Figure.

Figure 2.10 A 2D display of 16 dimers (each dimer represented as a double line) in a 32 chain intermediate filament. From the length of the dimer molecule the alignment gives repeating regions of 22nm and 47nm.
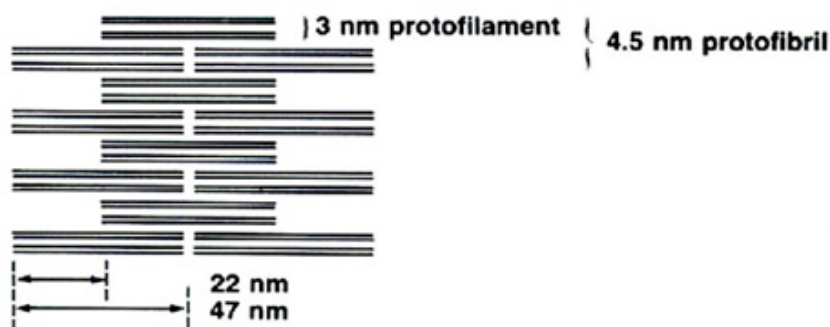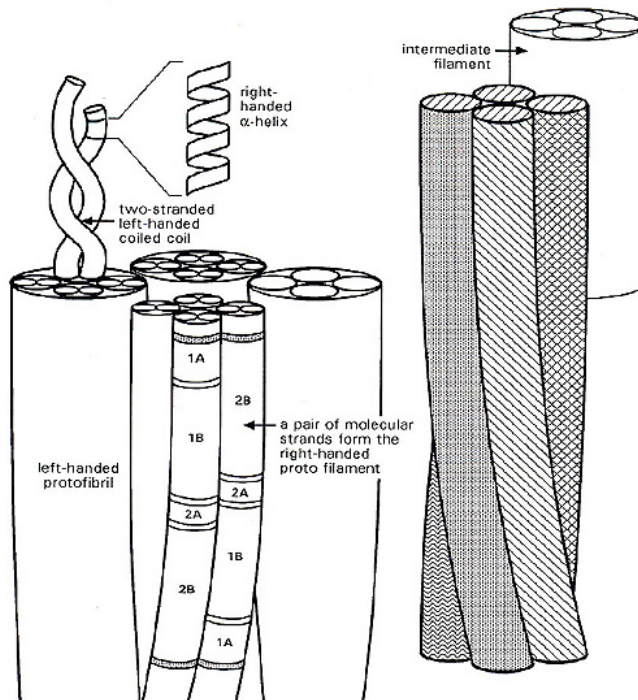**Source: Steinert and Freedberg, (1991).**

**Figure 2.11 The general features of intermediate filament structure. The disposition of 32 chains inside the IF is illustrated in the left hand diagram.
Source: Parry and Steinert, (1995).**

A model for the arrangement of the dimers can be visualised if it is imagined that the cylindrical IF is wrapped in a piece of paper and the paper is then unfolded. It reveals the dimers (Figure 2.12) organised into what is termed a surface lattice in which there are equivalent structural points, i.e. 1A and 2B segments joined by the diagonal lines in the Figure. This arrangement of alternating A_{11} and A_{22} tetramers (Figure 2.9) was proposed by Steinert at al. (1993) from studies of cytokeratins of the epidermis in which lysine residues that were close enough in the IF could be chemically cross-linked and the dimers that were adjacent were identified by sequencing peptides.

**Figure 2.12 The postulated arrangement of 32 keratin chains as dimers in a cytokeratin filament. The 16 dimers are arrayed antiparallel and the helical form is shown by the sloping lines. The 22.6 nm in the margin is the length of the pitch of the helical surface and is half of the 45.2nm repeat along the length of end-to-end dimers (protofilament).
Source: Parry and Steinert, 1995; Steinert et al. (1993).**
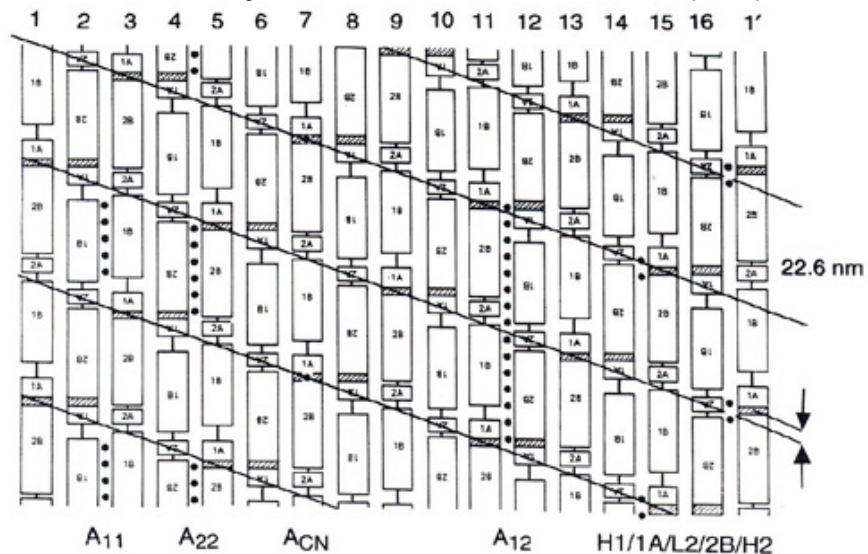
A confirmation of the presence of 32 chains in the cross-sections of hair keratin filaments has been concluded from using a technique called scanning transmission electron microscopy STEM). Hair keratin filaments were isolated from hair follicles by manual dissection were measured for mass-per-unit length and from that information it was calculated that on average, 16 dimers span the width of a single 10 nm wide filament Jones, 1997 #1430}. The presence of a central core seen in the cortex in transverse section (Figure 4A) was also confirmed. Several of the IF family including keratin IFs can be solubilized and under appropriate salt and pH conditions the filaments can be reconstituted in vitro. In the case of the keratins, hair keratin IFs are difficult to reconstitute whereas it can be achieved when the cytokeratins of the epidermis are used (Figure 13). This is probably because elaborate disulphide cross-linking requires vigorous conditions to disassemble the hair structure.

Figure 2.13 Cytokeratin intermediate filaments 10nm in diameter and of indeterminate length reconstituted in vitro from a purified solubilized epidermal cytokeratin using conditions of low ionic strength (<1mM) buffer at pH 7.
Source: http://www.hopkinsmedicine.org/CoulombeLabPage/intermediate.htm.
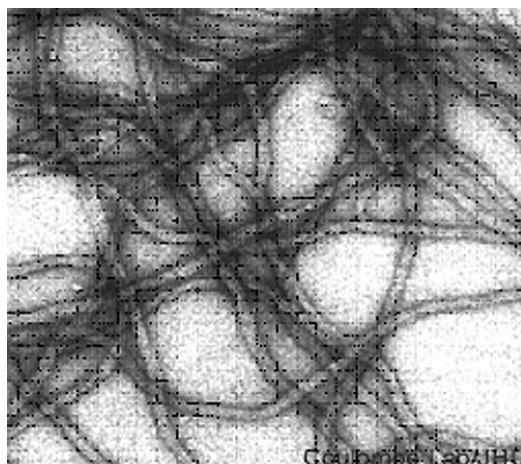
Cross-Linking in the Cortex, Cuticle, Medulla and Cell Membrane Complex of Wool

The covalent bond that stabilises the structure of wool and hair keratin molecules is the disulphide bond. There are many reactions that the bond can undergo but the main ones for composition studies of wool are reduction and oxidation, already discussed (2B.1). The stereochemical nature of the disulphide bond makes it impossible for the bond to form within an a-helix and with difficulty between adjacent a-helices so in keratin dimers the bonds could exist in the non-helical linker regions or the terminal N & C domains. The distribution of disulphide bonds in these regions is shown in Figure 2.14. Furthermore in the filament – matrix structure of wool the HS protein of the matrix and IFs must be linked by disulphide bonds (Figure 2.15) it can be calculated that the greater proportion of the bonds are *within* the HS molecules (Fraser et al., 1988).

Figure 2.14 The distribution of cysteine ("half-cystine") residues in different regions of Type I and Type II a-keratin filaments. Each cysteine ("half-cystine") residue is denoted by a dot. Source: Parry and Steinert, (1995).
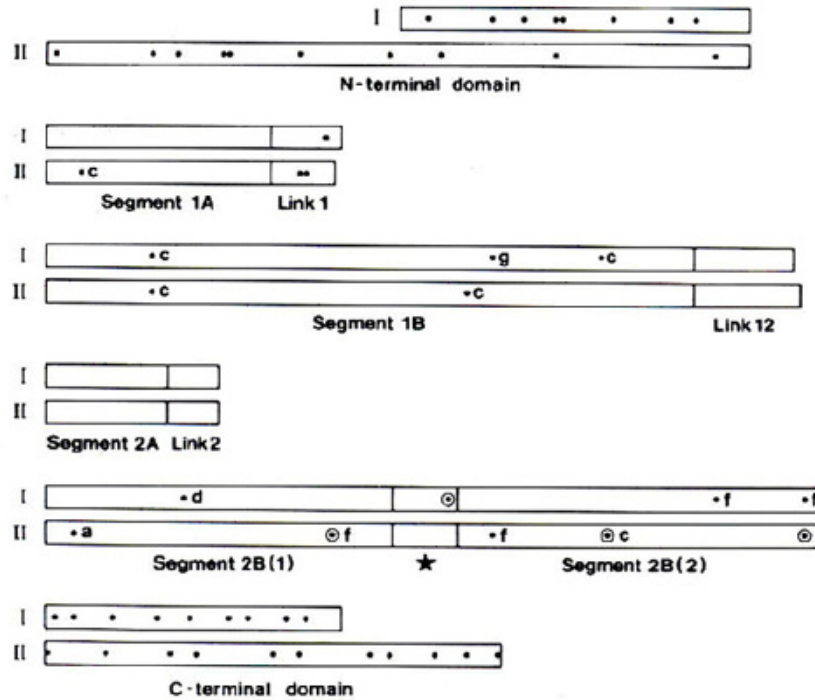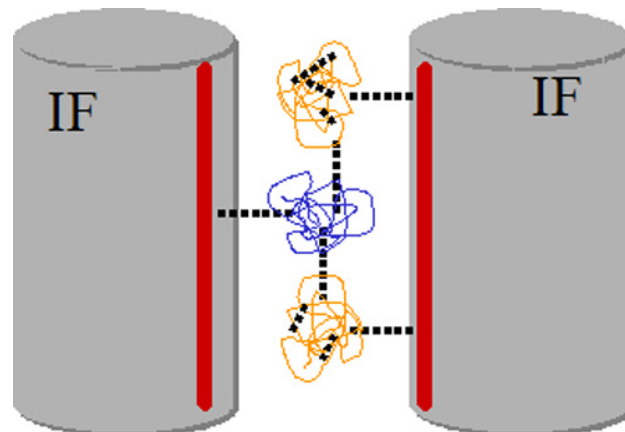

**Figure 2.15 A model of segments of two IFs and interposed HS and HGT matrix molecules indicating possible covalent interactions between the components. Two HS molecules (yellow) with internal disulphide cross-linking (dot lines) and cross-linking to a HGT protein molecule (blue) and a protofibril (brown) in each of the IFs.
Source: Jones and Rogers (2006).**

The laminated structure of the cuticle of wool fibres as seen in the electron microscope is described in Topic 2, Lecture 1. The main layers are the exocuticle, endocuticle, A-layer and an outermost layer called the epicuticle (Figure 2.16).

Figure 2.16 Diagram of the different laminae inside a scale cell. The outermost layer is a lipid layer of fatty acids of which approximately half is 18-methyleicosanoic acid an unusual fatty acid. Source: Zahn et al., (2005).
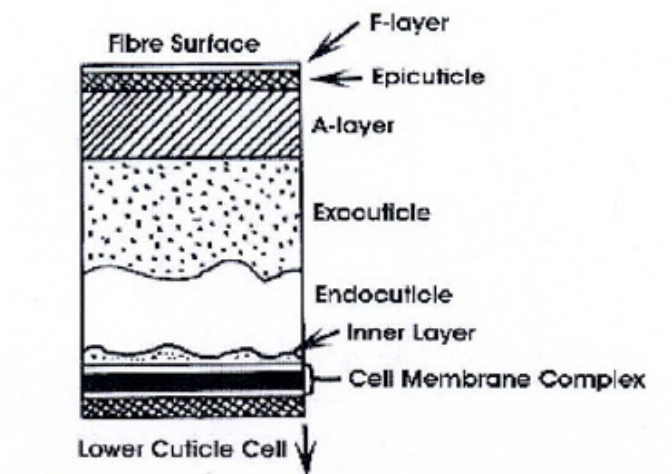

The number of proteins that compose the exocuticle is uncertain but two sulphur rich KAP proteins, KAP5 and KAP10, have been identified by gene cloning (MacKinnon et al., 1990). The exocuticle is a tough protective layer that is highly cross-linked by disulphide bonds and it is removed when wool is extracted with keratin solubilizing solutions described above in 2B.1. It is resistant to proteases but is digested by them when they act in the presence of a reducing agent such as dithiothreitol (Swift and Bews, 1976), further evidence for the presence of disulphide bond cross-linking in the cuticle. The endocuticle is readily digested by proteases without chemical pre-treatment and probably consists of denatured proteins remaining from the earlier cytoplasm. This layer swells in the presence of water. The A-layer was identified by TEM in sections of sections of wool and hair in 1959 (Rogers, 1959a) and it was suggested that it has a sulphur content higher than the exocuticle below it. Recent analyses substantiate that claim and it has been suggested that the responsible protein is possibly loricrin, a protein that is found in the cell envelope proteins of keratinising cells generally (Zahn et al., 2005). Further research is required to certainly identify the protein components of the scale cells and especially to determine the identity of the protein(s) to which the fatty acids in particular 18-methyleicosanoic acid are linked to form the outermost layer (Jones and Rivett, 1997). They are linked through their carboxyl groups to a protein(s) by bonds to SH groups, thioester bonds that is R-CO-S-Pr, where R stands for the carbon chain of the fatty acid and Pr stands for protein molecule. This lipid layer of fatty acids is responsible for the hydrophobic properties of the surface of wool and hair fibres..

Another covalent bond found in hair is the isopeptide bond. This bond (Figure 2.17) is established between closely adjacent glutamine and lysine side chains in proteins through the activity of transglutaminases (Stryer, 2002). In wool (and hair) the isopeptide bond was found in the protein of the medulla (Harding and Rogers, 1972) when these cells are present as a cellular core in the centre of the fibre cortex. The protein of the medulla cells in a mature hair is derived from trichohyalin granules in the developing hair through the activity of an enzyme called peptidylarginine deiminase . It converts arginine residues to citrulline and changes the conformation of the trichohyalin molecule from helical to random thereby opening up the molecule for cross-linking by transglutaminase (Fietz et al., 1993). Trichohyalin has a high content of lysine and glutamine residues that then become cross-linked into isopeptide bonds by the action of the transglutaminase during growth of the fibre. The granules fuse into an amorphous mass but IFs are not present. The formation of isopeptide linkages also plays a major role in the differentiation of the IRS of hair follicles. In the IRS cells (compared to the medulla) specific intermediate filaments are formed during differentiation and the trichohyhalin is converted to a matrix that finally cross-links to itself and to the filaments by process analogous to what occurs in the fibre cortex but is biochemically distinct. For a detailed description see Powell and Rogers (1997).

Figure 2.17 The isopeptide bond (e-[g-glutamyl lysine]) formed from a glutamine and lysine residue between two polypeptide chains by the action of a tranglutaminase. Source: Stryer, (2002).

Glu -CO – NH- Lys
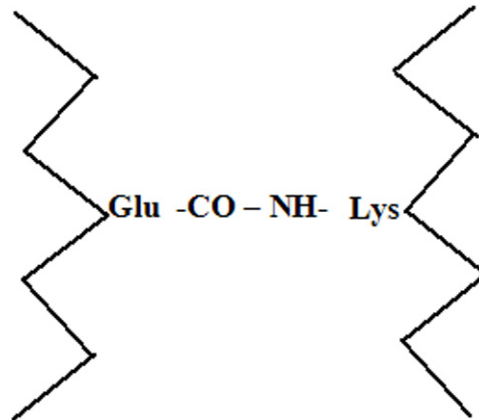
The structure of the cell membrane complex (CMC) that is found between cortical cells and cuticle cells are discussed in Lecture 2A however, it is pertinent to mention here that isopeptide bonds appear to be at least partly responsible for the stability of the CMC and some of the proteins in the cuticle, possibly only the epicuticle, also are cross-linked by isopeptide bonds in addition to disulphide bonds. It is known that the cuticle contains HS proteins related to the keratin associated proteins of the cortical matrix.

The relationship between wool structure and properties

In the present discussion of wool composition it is clear that the disulphide bond is central for the stable conformation of the proteins of wool. Several industrial processes that wool undergoes involve the properties of the bond and their influence on the resistance to physical and chemical insults. For example, a process used for wool such as permanent set depends on the breaking. The formation of disulphide bonds and the process of scouring for removing lipid material from wool depends on the relative stability of the keratin fibre compared to dissolution of the lipid. The elastic behaviour a wool fibre when it is stretched in water is shown in Figure 18. In broad terms the elongation of the fibre involves the unfolding of the a-helices of the coiled-coils of the cortex (the yield region) and the limitation to that extension by the disulphide bond linkages between IF dimers and between matrix molecules and IFs (post-yield region).

Figure 2.18 A typical stress strain curve of a wool fibre extended in water. Source: Fraser et al., (1972).
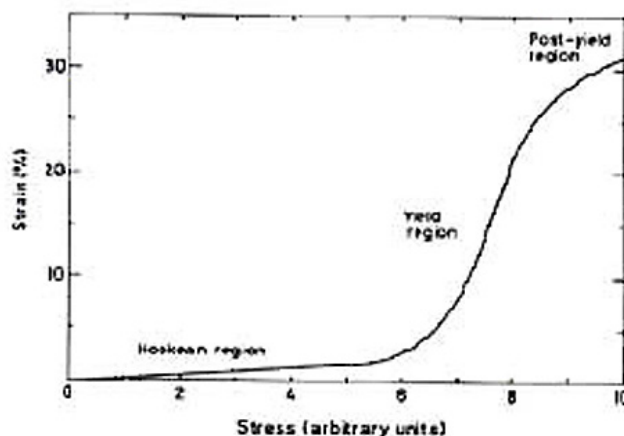
The elastic behaviour is a summation of the separate influences of the cortex as the main player, the cuticle and the cell membrane complexes that surround all of the cells. Very little is known as to the relative quantitative and qualitative contributions of these minor components. Insofar as the cortex is the dominant one, very little is known about precisely where the disulphide linkages are located between and within the filaments and the matrix molecules. This is an area for future research and with the development of proteomics an understanding on this question should be achievable.

The advent of producing transgenic sheep (the injection of DNA into pronuclei of fertilised ova followed by implantation) has made it possible to introduce new or novel genes that in principle would produce changes in the properties of wool. It is already known that the availability of essential amino acids to the sheep for wool growth influences growth rate and the quality of wool. It has been shown that when keratin genes are over-expressed in transgenic sheep major changes in wool structure and properties can be produced (Powell and Rogers, 1997; Rogers, 1990). The commercial usefulness of transgenesis for improving wool properties will only be realised when the efficiency of transgenesis is improved and the level of expression of newly introduced genes can be controlled.

Readings [3]

The following readings are available on CD:

1. Jones, L.N. and Rivett, D.E. 1997. 'The role of 18-Methyleicosanoic acid in the structure and formation of mammalian hair fibres. Micron, vol 28, pp 469-485.
2. Zahn, H., Wortmann, F. and J. and Hoecker, H., 2005. 'Considerations on the occurrence of Loricrin and Involucrin in the cell envelope of wool cuticle cells.' International Journal of Sheep and Wool Science, vol 53, pp 1-13.
3. Steinert, P.M. and Freedberg, I.M., 1991. Chp1. Epidermal Structural Proteins. Molecular proteins and cellular biology of keratins. In Physiology, biochemistry and molecular biology of the skin. Vol 1 (ed) Goldsmith, L.A. Oxford University Press, New York, pp 120-129.

Activities   @ Available on WebCT

Multi-Choice Questions   @ Submit answers via WebCT

Useful Web Links   @ Available on WebCT

Assignment Questions   @ Choose ONE question from ONE of the topics as your assignment. Short answer questions appear on WebCT. Submit your answer via WebCT

References

Answers.com - available from http://www.answers.com/topic/alpha-helix (public domain).

Fietz, M.J., C.J. McLaughlan, M.T. Campbell, and G.E. Rogers. 1993. Analysis of the sheep trichohyalin gene: potential structural and calcium-binding roles of trichohyalin in the hair follicle. *Journal of Cell Biology, vol 121, pp 855-865.*

Filshie, B.K., and G.E. Rogers. 1961. The fine structure of a-keratin. *Journal of Molecular Biology*. vol 3, pp 784-787.

Fraser, R.D.B., T.P. MacRae, and G.E. Rogers. 1972. Keratins. Their Composition, Structure and Biosynthesis. Charles C. Thomas, Springfield, Illinois.

Fraser, R.D.B., T.P. MacRae, L.G. Sparrow, and D.A.D. Parry. 1988. Disulphide bonding in a-keratin. *International Journal of Biological Macromolecules, vol 10, pp 106-112.*

Gillespie, J.M. 1991. The structural proteins of hair: isolation, characterization and regulation of biosynthesis. *In* Physiology, Biochemistry and Molecular Biology of the Skin. Vol. I. L.A. Goldsmith, editor. Oxford University Press, Oxford. 625-659.

Goddard, D.R., and L. Michaelis. 1935. Derivatives of keratin. *Journal of Biological Chemistry*. vol 112, pp 361-371.

Harding, H.W.J., and G.E. Rogers. 1972. The occurrence of the e-(g-glutamyl) lysine crosslink in the medulla of hair and quill. *Biochim. et Biophys. Acta*. vol 257, pp 37-39.

Hesse, M., A. Zimek, K. Weber, and T.M. Magin. 2004. Comprehensive analysis of keratin gene clusters in humans and rodents. *European Journal of Cell Biology,* vol 83, pp,19-26.

John Hopkins Medicine - the Coulombe Laboratory, Department of Biological Chemistry - available from http://www.hopkinsmedicine.

org/CoulombeLabPage/intermediate.htm.

Jones, L.N., and D.E. Rivett. 1997. The role of 18-methyleicosanoic acid in the structure and formation of mammalian hair fibres. *Micron*. vol 28, pp 469-485.

Jones, L.N., M. Simon, N.R. Watts, F.P. Booy, A.C. Steven, and D.A. Parry. 1997. Intermediate filament structure: hard alpha-keratin. *Biophysical Chemistry,* vol 68, pp 83-93.

MacKinnon, P.J., B.C. Powell, and G.E. Rogers. 1990. Structure and expression of genes for a class of cysteine-rich proteins of the cuticle layers of differentiating wool and hair follicles. *Journal of Cell Biology,* vol 111, pp 2587-600.

Parry, D.A.D., and P.M. Steinert. 1995. Intermediate filament structure. Springer-Verlag, Heidelberg.

Powell, B.C., and G.E. Rogers. 1997. The role of keratin proteins and their genes in the growth, structure and properties of hair. *In* Formation and structure of hair. P. Jolles, H. Zahn, and H. Hocker, editors. Birkhauser Verlag, Basel. vol 59, pp148.

Rice, R.H., V.J. Wong, and K.E. Pinkerton. 1994. Ultrastructural visualization of cross-linked protein features in epidermal appendages. *Journal of Cell Science.*:1985-1992.

Rogers, G. 1990. Improvement of wool production through genetic engineering. *Trends in Biotechnology*. vol 8, pp 6-11.

Rogers, G.E. 1959a. Electron microscope studies of hair and wool. *Annals of the New York Academy of Sciences,* vol 83, pp 378-399.

Rogers, G.E. 1959b. Electron microscopy of wool. *Journal of Ultrastructure Research,* vol 2, pp 309-330.

Steinert, P.M., and I.M. Freedberg. 1991. Molecular and cellular biology of keratins. *In* Physiology, biochemistry, and molecular biology of the skin. Vol. 1. L.A. Goldsmith, editor. Oxford University Press, New York. vol 113, pp 147.

Steinert, P.M., L.N. Marekov, R.D.B. Fraser, and D.A.D. Parry. 1993. Keratin intermediate filament structure: crosslinking studies yield quantitative information on molecular dimensions and mechanism of assembly. *Journal of Molecular Biology*. vol 230, pp 436-452.

Stryer, L. 2002. Biochemistry. W. H. Freeman, New York.

Swift, J.A., and B. Bews. 1976. The chemistry of human hair cuticle-III: the isolation and amino acid analysis of various subfractions of the cuticle obtained by pronase and trypsin digestion. *Journal of the Society of Cosmetic Chemists,* vol 27, pp 289-300.

Ward, K.A., and S.E. Kasmarik. 1980. The isolation of wool keratin messenger RNA from sheep. *Journal of Investigative Dermatology,* vol 75, pp 244-248.

Ward, K.A., M.J. Sleigh, B.C. Powell, and G.E. Rogers. 1982. The isolation and analysis of the major wool keratin gene families. *In* Proceedings of the 2nd World Congress of Genetics Applied to Livestock Production. Vol. 6. E. Garsi, editor, Londres 17, Madrid-vol 28, pp 146-156.

Zahn, H., F.-J. Wortmann, and H. Hoecker. 2005. Considerations on the occurrence of loricrin and involucrin in the cell envelope of wool cuticle cells. *International Journal of Sheep and Wool Science,* vol 53, pp 1-13.

Glossary of terms

Cytokeratins	The group of proteins that give rise to the keratin intermediate filaments of the epidermis and distinguishable

	from the hair keratin intermediate filaments
Cytoskeleton	The network of structural proteins that contribute to cell shape and include actin, microtubules and intermediate filaments
Denaturation	The process in which a protein in its normal configuration is unfolded and altered into a conformation with different properties
Disulphide bond	A covalent bond formed between two sulphur atoms of two cysteine molecules in a protein chain(s)
Intermediate filaments	The 8-10nm diameter filaments that are found in virtually all tissues of vertebrates and include desmin, vimentin, neurofilaments, lamins and keratins
Isopeptide bond	A covalent bond formed between a glutamyl (glutamic acid) residue and a lysyl (lysine) residue in a protein chain(s)
Protein family	A group of proteins of closely-related amino acid sequences that arise from gene duplication during evolution
Keratin protofilament	The structure envisaged when two dimers made up of one Type I and one Type II keratin intermediate filament chains aggregate to give a tetramer
TEM	Transmission electron microscopy