A detailed mapping of the readily accessible disulfide bonds in the cortex of wool fibers

Jeffrey Plowman[1], Rachel Miller[2], Ancy Thomas[2], Anita Grosvenor[2], Duane Harland[1], and Santanu Deb-Choudhury[3]

[1]AgResearch Ltd.,
[2]AgResearch Ltd
[3]AgResearch Ltd.

September 11, 2020

Abstract

One way in which trichocyte keratin intermediate filament proteins (keratins) and keratin associated proteins (KAPs) differ from their epithelial equivalents is in their higher levels of cysteine residues. Interactions between these cysteine residues within a mammalian fiber, and the putative regular organization of interactions (i.e., types of disulfide bond) are likely important for defining fiber mechanical properties, and thus biological functionality of hairs. Here we extend a previous study of cysteine accessibility under different levels of exposure to reducing compounds to explore a finer set of levels associated with interactions between keratins and KAPs. We found that most of the cysteines in the KAPs were close to either the N- or C- terminal domains of these proteins. The most accessible cysteines in keratins were present in the head or tail domains indicating their function in readily forming intermolecular bonds with KAPs. Some of the more buried cysteines in keratins were discovered either close to or within the rod region in positions previously identified in human epithelial keratins as being involved in crosslinking between the heterodimers of the tetramer. Our present study therefore provides a deeper understanding of the accessibility of disulfides especially in keratins and thus proves that there is some specificity to the disulfide bond interactions leading to these intermolecular bonds stabilizing the fiber structure.

INTRODUCTION

Trichocyte keratin intermediate filament proteins (keratins) and the keratin associated proteins (KAPs) make up the major part of the protein content of wool fibers [1]. The structural domains of keratins consist of a head (N terminus), central rod and tail (C-terminus). The central rod region is divided into three highly conserved α-helical domains, coils 1A, 1B and 2, each with an underlying structure of heptad repeat sequences separated by short non-helical linker regions, L1 and L12 [2]. Coil 2 also has two additional features: an eleven-residue hendecad repeat region[3] at the start of the domain and a 'stutter' in which three of the residues are deleted from the heptad repeat. The heptad repeats are of the form a-b-c-d-e-f-g in which residues 'a' and 'd' are largely apolar. This arrangement results in a hydrophobic stripe running around the α-helix in a left-handed manner. In the case of the trichocyte keratins two types exist, the acidic type I and the neutral-basic type II, their hydrophobic stripes allowing them to associate to form a coiled-coil heterodimer. This association is also stabilized by electrostatic interactions between acidic or basic residues located at positions 'e' and 'g' in the heptad repeat. In contrast to the central rod region of the keratins, the head domain is cysteine rich, has less sequence regularity and, with the exception of a possible nonapeptide quasi-repeat in type II keratin heads[4], no structure has been identified. The tail of both type I and type

II appear to have limited regions that may include some short α-helix and β-sheet regions, which have the potential to interact with similar regions in some circumstances [4].

The proteins in the hair fiber are notable for their high cysteine content, something that distinguishes them from other intermediate filament forming proteins and also keratins found in epithelial cells[5]. The trichocyte keratins alone are known to have a minimum of 20 cysteine residues, most of which are concentrated in the head and tail domains of the keratins [6]. The amount of cysteine found in the KAPs is even higher, with the so-called high sulfur proteins (HSPs) having between 20-30 moles% cysteine and the ultra-high sulfur proteins (UHSPs) having between 30 and 38 moles% cysteine. It is the interaction between the cysteines in the trichocyte keratins with those of the KAPs that stabilize the three-dimensional structures of the keratin intermediate filaments and distinguishes them from epithelial keratins [2]. These interactions are thought to be the main driver of fiber functionality and important in reduction-then-oxidation processes such as stretch set in the manufacture of wool garments [7]. Removal of cysteine from wool fibers through the conversion of cysteine to cysteic acid, by for example, oxidative bleaching with hydrogen peroxide to whiten wool, is known to impair process like stretch set considerably[7].

Keratin head domains are thought to be important for intermolecular interactions with KAPs, largely assumed to be disulfide bonding. Recent co-immunoprecipitation studies have demonstrated an interaction between the head domain of human K86 with KAP2.1 [8], while Western blot studies have demonstrated an interaction between the head domain of K85 and KAP8.1 [8]. However, studies have indicated that approximately 97.5% of cysteines in the KAPs are involved in intramolecular bonding [9]. Early proteomic studies on wool using two-dimensional electrophoresis suggested that only a few discrete cysteine residues in the KAPs were affected by treatment with hydrogen peroxide [10]. Differential accessibility of the cysteines in single samples of wool has also been investigated by mass spectrometry protocols involving reduction followed by alkylation and then extraction [11]. To further investigate non-random cysteine-accessibility we undertook a study, on which the current study builds, in which wool was subjected to a process of incremental labelling whereby reductants and chaotropes were used in stages to reductively expose and label cysteines in the fiber in a stepwise fashion according to their accessibility, after which the labelled peptides were extracted and identified by proteomics, a process repeated five times at each stage [12]. In this study when we refer to residues, we refer exclusively to these with a non-random accessibility. From this it became apparent that the most readily accessible cysteines were within the head or tail domains of the keratins. Given that at least 50% of the head and tail domains of the trichocyte keratins lie along the inner core of the strongly apolar intermediate filament [13], where reducing agents would be less effective, this would suggest that these accessible cysteines are most likely on the outer surface of the filament. In contrast, no conclusion could be drawn on the cysteine accessibility of the KAPs because, under the conditions investigated, the KAPs did not display the sequential accessibility of the keratins, all repeatably non-random cysteine labelling occurring in one stage/step. One limitation of our earlier study on the effect of reductant on cysteine accessibility was that the lowest concentration of DTT examined was 20 mM[12]. Another study, however, has shown that the various components of the KAPs exhibit a differential extractability when sequentially exposed to concentrations of DTT between 5 and 20 mM[14]. It was for this reason that we extended this approach to study the effect of these lower concentrations of reductant not only to understand the effect on the accessibility of the cysteines in the KAPs in the mature wool fiber but also to examine the accessibility of the cysteines in keratins at these concentrations with the view to gaining an insight as to how the disulfide bridges may form between keratins and KAPs.

METHODS AND MATERIALS

Materials

Ammonium bicarbonate, ethylenediaminetetraacetic acid (EDTA), formic acid, methanol, sodium phosphate and urea were obtained from Merck (Darmstadt, Germany). Dithiothreitol (DTT), iodoacetic acid (IAA),

and iodoacetamide (IAM) were obtained from Sigma (St Louis, MO, USA). Bovine serum albumin (BSA) was obtained from ICPBio, Auckland, NZ). Acrylamide was obtained from Bio-Rad (Hercules, CA, USA). LC/MS-grade acetonitrile was obtained from Fisher Scientific (Pittsburg, USA). TPCK-trypsin was obtained from Promega (Fitchburg, MI, USA). LC/MS-grade water was obtained from Fluka. Empore disks were obtained from Supelco (Bellefonte, PA, USA).

Preparation of merino wool

Merino wool with fibers of ˜18 μm diameter were sourced from a ewe from a farm in mid-Canterbury, New Zealand. The wool was detipped by removing the top third of the fiber and the extraneous dirt and sebum removed by subjection to agitation in an ultrasonic bath in ultrapure water for 15 mins three times. After this the wool fibers were washed sequentially twice in dichloromethane, twice in ethanol, then twice in water before air drying overnight. After this the wool was cut into snippets of less than 1 mm length with scissors.

Sequential alkylation with IAM, acrylamide and IAA

The method of stepwise reduction and labelling of cysteines is summarized in Figure 1, where cysteines in the wool fiber are progressively reduced by increasing the concentration of DTT. This experiment was split into three steps involving the reduction of five independent samples with 5 mM, 10 mM and 15 mM of DTT, followed by alkylation with one of three different reagents: iodoacetamide (IAM), acrylamide and iodoacetic acid (IAA) to determine at which step the cysteine group became accessible to the labelling reagent.

Prior to commencing the experiment, a control run was made with a sample of the snipped wool in the absence of DTT. This was incubated in the base buffer solution only (100 mM sodium phosphate buffer (pH 7.4), 1 mM EDTA) for 24 h followed by labelling with 2 mM IAM in base buffer for 30 min in the dark. The supernatant was collected after the extract was centrifuged at 44,000 g for 15 min. This was Step 0 (Figure 1).

In the first step, 20 mg of cleaned wool was incubated in 1.5 mL of the base buffer containing 5 mM DTT for 24 h followed by labelling with 2 mM IAM in the base buffer for 30 min in the dark. The supernatant was collected after the extract was centrifuged at 44,000 g for 15 min. Residual extraction buffer was removed from the pellet by vortexing for 1 min in wash buffer (10 mM sodium phosphate buffer (pH 7.4), 1 mM EDTA), followed by centrifugation at 44,000 g for 15 min. This was repeated three times and these washes discarded. After this the pellet was suspended in wash buffer and centrifuged. Mass spectrometry was used to detect the presence of residual IAM. This was performed by taking 60 μL of the supernatant from the washed wool pellet and added to 10 μL of 0.2 mg/mL of reduced BSA and vortexed in the dark for 30 min at room temperature. This BSA was digested with trypsin and the resulting digest analyzed for alkylation of the BSA peptides by mass spectrometry. If no alkylation was detected in the BSA peptides, indicating the absence of residual IAM, the washed pellet was then used in the next step for further analysis.

In the second step the washed wool pellet was incubated in the base buffer solution containing 10 mM DTT for 24 h. This was followed by labelling with 350 mM acrylamide in the base buffer for 24 h in the dark. The supernatant was then collected after centrifugation and the pellet washed three times with wash buffer and checked for residual alkylating reagent, as described above, before proceeding to the next step.

In the third step the washed wool pellet was incubated in the base buffer containing 15 mM DTT for 24 h. This was followed by labelling with 2 mM iodoacetic acid in base buffer for 30 min in the dark, after which the extract was centrifuged and the supernatant collected, while the pellet was discarded.

Preparation for Mass Spectrometry

Each extract was digested with TPCK-trypsin at an enzyme-substrate ratio of 1:50 in 50 mM ammonium bicarbonate:AcN (9:1) at 37 °C for 18 h. The peptide digest was vortexed for 3 h in the presence of Empore

disks that had been prewetted with AcN and methanol. Following this the Empore disks were washed with 0.1% TFA and extracted with 75% AcN in 0.1% trifluoroacetic acid. After drying the extracts were resuspended in 50 µL of 0.1% trifluoroacetic acid for further analysis.

Mass spectrometry analysis

LC-MS/MS was performed on a nanoflow UPLC system (Dionex Ultimate 3000) coupled to an Impact II mass spectrometer equipped with a CaptiveSpray source (Bruker Daltonik, Bremen, Germany). For each sample, 1 µL of the sample was loaded on a C18 PepMap100 nano-Trap column (200 µm x 2 cm) at a flow rate of 5 µL/min. The trap column was then switched in line with the analytical column ProntoSIL C18AQ (100 µm ID x 150 mm 3 micron 200 Å). The reverse phase elution gradient was from 2% to 20% to 45% B over 60 min, total 84 min at a flow rate of 1000 nL/min. Solvent A was LCMS-grade water with 0.1% formic acid; solvent B was LC-MS-grade AcN with 0.1% formic acid.

Samples were measured in automatic MS/MS mode, with a mass range of m/z 150–2200. Acquisition speed was 5 Hz in MS and 1-50 Hz in MS/MS mode depending on precursor intensity. Ten precursors were selected in the m/z 300-1250 range, with a preference for doubly or triply charged peptides. The analysis was performed in positive ionization mode with a dynamic exclusion of 60 sec.

Data Analysis

Following the LC-MS run, the QTOF data were searched using Peaks Studio 8.5 (Bioinformatics Solutions Inc). The raw data were refined by a built-in algorithm and protein/peptides were identified using the following parameters: a precursor mass tolerance of 10 ppm; fragment mass tolerance of 0.05 Da, a taxonomy of *Ovis aries* with an in house Sheep_uniprot_ovine_2016_10_plus_keratins database. This database was derived from the Uniprot 2016_10 database, augmented by 54 AgResearch sequences from EST contigs, annotated by BLAST searching against the NCBI database and keratin sequences from the NCBInr non-redundant database. Trypsin was chosen as the proteolytic enzyme and up to 2 mis-cleavages were allowed. Carbamidomethyl (C), propionamide (C) and carboxymethylation (C) were chosen as variable modifications. A false discovery rate (FDR) of 0.5% was used for peptide identification in Peaks. In addition, the Peaks post-translational modification (PTM) A-score was set to 30, de novo only ALC > 85% and only proteins with a minimum of 2 unique peptides identification were included. As the keratin-family of proteins is known to have highly similarly in protein sequences, the quality of identification of all peptides was checked manually.

In the final analysis a peptide sequence containing the same labelled cysteine was only considered to be significant if it was found in four of the five repeats. This was because in our previous study it was established that only peptide sequences with the same labelled cysteine observed in at least four out of five repeats had a significance level of $P < 0.05$ [12]. That is, any peptide with less than four observations was considered to have non-repeatable, and therefore random, accessibility.

RESULTS

Sequential labelling of cysteines

In order to determine the relative accessibility of cysteines in wool the keratins and KAPs were subjected to a process of sequential alkylation as the concentration of DTT was raised in stepwise fashion from 0 to 15 mM.

A total of 38 peptides were detected and identified in the mass spectrum of protein extracts in the control run, involving wool that had not been treated with a reductant. Of these, two were from the epithelial keratin K10 and the rest were from either actin or serum albumin. A number of cysteines in the peptides were labelled with IAM but only in those from serum albumin.

In contrast cysteine labelling of keratins and KAPs was only detected when the wool was incubated with DTT. Among the labelled cysteines were a number identified in our earlier study [12], specifically: in the tail domain of K31, T-42 and T-45; in coil 1B of K31 and K33b, 1B-87; the head domain of K35, H-36; H-24 and H25 in K81 (H-29 and H30 in K83); and H61 and T-24 in K85.

In the first step, when a DTT concentration of 5 mM was used to reduce the proteins, eight peptides from four keratins and three KAPs containing IAM-labelled cysteines were identified in at least four out of the five biological repeats (Figures 2-4, Table 1). Interestingly, all the keratin peptides identified were labelled on either the head or tail domain. In the case of the type I keratins one peptide from the tail region of K31 was identified on which four cysteine residues were labelled. Peptides from three Type II keratins were also found to have IAM-labelled cysteines. For K83 this involved two peptides from the head domain of which three cysteines were found to be labelled. In the case of K85 a peptide containing one cysteine from its tail domain was labelled, while a peptide containing one cysteine from the tail domain of K86 was also labelled. IAM-labelled peptides were identified from three HSPs, specifically cysteines close to the N-terminus of KAP2.3 and the cysteines close to the C-terminus of KAP7.1, KAP11.1 and KAP13.1 (Figure 2, Table 1).

In the second step, proteins were unraveled further by reduction with 10 mM DTT and then labelled with acrylamide (Figures 2-4, Table 2). This resulted in the identification of a total of 62 cysteine-labelled peptides, of which 20 were from keratins and 42 from KAPs. Of the KAPs, 65% of were classed as HSPs, 16% UHSPs and the rest high glycine-tyrosine proteins (HGTPs). In the case of the type II keratins all the newly characterized peptides in this step (and hence cysteines) were either located in the head or tail domains (Figure 4). In contrast, in the type I keratins one acrylamide-labelled cysteine was identified in a peptide from coil 1B and another from coil 2, while an acrylamide-labelled cysteine was identified in a peptide from the L1 linker in K34 (Figure 3). No modified cysteines were found in the hendecad regions [3]. As found previously[12] around 90% of the labelled cysteines in the HSPs and UHSPs are found close to a proline residue in the sequence and over half within 20 residues of the N- and C-terminus of the proteins. Specifically these involved cysteines at the N-terminus of KAP3.3, KAP3.4 and KAP7.1, and cysteines close to the C-terminus of KAP31, KAP3.4, KAP4.2, KAP4.5, KAP9.2, KAP9.8, KAP10.11, KAP10.12, KAP12.2, KAP19.5 and KAP26.1 (Figure 2, Table 2). At the same time a number of cysteines were detected from the more central regions of some of the KAPs, among them KAP2.3, KAP3.4, KAP4.2, KAP6.1, KAP6.4, KAP10.4, KAP11.1,KAP13.1, KAP16.1-like, KAP19.3, KAP24.1 and KAP26.1.

In the third step very few IAA-labelled peptides were identified in the final alkylation step. Of the two keratin peptides identified the one from K31 was labelled in the coil 2 domain while that from K35 was labelled in the head domain (Figure 3, Table 3). Two KAPs were identified in this stage, one being KAP 6.1, an HGTP, and the second being from the HSP KAP13.1-like (Figure 2, Table 3).

Validation of the labelling experiments

To determine if the large increase in peptides identified in step 2 and the subsequent drop in peptide numbers in step 3 was due to the longer labelling time used with acrylamide the entire process was repeated using IAM in all steps. However, a similar pattern was observed when comparing the proportion of peptides identified between each of the steps. This led to the conclusion that the alkylation reagent used was not influencing the efficiency of the labelling and thus peptide identification.

DISCUSSION

New highly accessible disulfides were identified

In combination with our earlier step-based approach to cysteine accessibility, the findings here help give a more complete picture of disulfide accessibility in the wool fiber because between the two studies we have seven levels of reducing and/or chaotropic intensity (Figure 5).

Keratin associated proteins

In our earlier study of cysteine accessibility the cysteines in the KAPs most affected tended to be found in the central region of the proteins and usually close to a proline residue [12]. The HSPs and UHSPs in particular are notable for having a pentapeptide repeat of the structure of the following forms: A1, C-C-Q-P-X or A2, C-C-R-P_X and B, C-C-X-S/T-S/T, in which the presence of a cysteine close to a proline residue was thought to confer a conformation in the protein that may have made these cysteines more accessible [312]. In contrast, in this study at 5 mM DTT the only consistently modified cysteines were located at either the N- or C-termini of the three HSPs. Raising the DTT concentration to 10 mM DTT not only exposed more N- and C-terminal cysteines in the KAPs but also some of the centrally located cysteines in these proteins.

Keratins

This study has provided evidence that the majority of the accessible cysteines were found in the head and tail domains of the keratins at concentrations of DTT lower than the concentration used previously[12]. This is particularly true for the major type II keratins, K81, K83, K85 and K86, where consistently modified cysteines were only found in the head and tail domains (Figure 4, Table 2). In the case of the type I keratins consistently modified cysteines were found in both the head and tail domains in K34 and K35 but only in the tail domain of K31 and K33a, (Figure 4, Table 2). In addition, some cysteines in coil 1B were also accessible in K31 and in the L1 linker of K34 at 10 mM DTT, as well as cysteines close to the tail domain in coil 2 in K31, K33a, K33b, K34 and K35.

Studies of trichocyte keratin structure have indicated that during the process of assembly of keratins into coiled-coil dimers, tetramers, unit-length filaments and intermediate filaments, half of the head and tail domains are thought to protrude from the surface of this filament where they are free to interact with the KAPs [1315]. Thus, the labelled cysteines in the head domains of K34, K35, K81, K83 and K85 are possible sites for keratin-KAP disulfide crosslinks. Likewise, the cysteines in the C-terminus of K31, K33a, K34, K81, K83, K85 and K86 are also possible sites for keratin-KAP disulfide crosslinks, although repeatably accessible cysteines close to the tail domain coil 2B (noted above) are also potential partners.

Keratin rod domain and tetramer linkages

Studies of disulfide crosslinking in human and mouse epithelial keratins have shown the locations of a number of intermolecular crosslinks in the coil 2 region between pairs of heterodimers [16]. For instance in the reduced form of the mouse and human epithelial type I keratin K10, a cysteine in coil 2B at position 2B-62 (position 2-89 in the hendecad-heptad motif of coil 2) has previously been identified as the site of an intermolecular heterodimer disulfide crosslink with another 2B-62 (2-89) cysteine residue of K10 [16]. In this study, an accessible cysteine was found near the stutter in position 2-90 in coil 2 of K31 (Figure 3, Table 2), which would mean that this cysteine is possibly in a good orientation to crosslink with cysteine in the same position in another type I keratin because the A_{22} axial alignment of trichocyte keratins is unchanged between the reduced and oxidized states and also identical to the reduced form of epithelial keratins (Figure 6). However, such an interaction would only be possible if the axial separation of the two cysteines is appropriate and if the stereochemistry is correct. One consequence of this is that it may involve some local unwinding or distortion to allow the energetically favorable disulfide bonds to form[17]. A further readily reducible cysteine was detected at position 2-132 in coil 2 of K31 and K33a (position 2B-105 in the 2A-L2-2B coil system) (Figure 3, Table 3) and this has also been identified as the site of a disulfide bond between 2-132 in K31 and 2-49 in K83 [18]. One cysteine at position T-5 in the tail domain of K83 and K86 is also in a similar position to a cysteine at position T-4 human and mouse epithelial K1 that is involved in keratin-keratin crosslinking with a cysteine close to the C-terminus of the head domain in type I keratins [16], meaning that this could be another site of keratin-keratin crosslinking in trichocyte keratins. Finally, in coil 1 there are two readily reducible cysteines, L1-9 in K34 and 1B-87 in K31. In mouse trichocyte keratins these residues have been identified as possibly being involved in another disulfide crosslinking site between heterodimers

in the tetramer[18] [19]. Thus, the ready accessibility of these cysteines in the keratins suggests that these are potential sites of disulfide breakage as the fibers age and are subject to oxidative damage from such factors as sunlight, heat or chemical treatments.

CONCLUSIONS

This study of the cysteine accessibility in the keratins and KAPs of the wool cortex has revealed the following:

- Repeatably accessible (non-random) cysteines found in the head and tail domains in the type I and II keratins are likely to be on the surface of the intermediate filament and, hence, available for interactions with KAPs.
- Two accessible cysteines were found in coil 2 in the type I keratin K31 which are potentially involved in crosslinks with other keratins in the filament. One cysteine in position 2-90 in a type I keratin is possibly in a good orientation to be involved in a crosslink with another cysteine also at position 2-90 in a second type I keratin, while a second cysteine at position 2-132 in a type I keratin has also been identified as the site of another potential keratin-keratin crosslink.
- One accessible cysteine at position T-5 in the tail domain of K83 and K86 is in a position where it could be involved in crosslinking with a cysteine close to the C-terminus of the head domain of type I keratins.
- Cysteines at positions L1-9 of K34 and 1B-87 of K31 are most probably linked via a disulfide interaction.
- The ready accessibility of the cysteines involved in disulfide interactions between keratins in the intermediate filament point to how oxidative stress on the fiber brought on by environmental factors has the potential to affect the integrity of the filament itself.
- The bulk of the most readily accessible cysteines in the KAPs were found close to either the N- or C-terminal domains in these proteins.

Acknowledgements

The authors are grateful to Erin Lee for preparing the samples for mass spectrometry, Evelyne Maes for helpful comments during the preparation of the manuscript and Professor David Parry of Massey University for helpful discussions. The work was funded by the Wool Research Organisation of New Zealand (WRONZ) and the Ministry of Business Innovation and Employment through the New Materials from New Zealand Wool partnership program via Wool Industry Research Limited.

Figures

Figure 1. Schematic summary of the sample preparation methodology steps.

Figure 2: KAP sequences containing accessible cysteines identified in stage 1 (Bold), Step 2 (Bold underlined) and Step 3 (Bold Italicized).

Figure 3. Keratin type I sequences containing accessible cysteines identified in Step 1 (Bold), Step 2 (Bold underlined) and Step 3 (Bold Italicized), labelled cysteines are in red.

Figure 4. Keratin type II sequences containing accessible cysteines identified in Step 1 (Bold), Step 2 (Bold underlined) and Step 3 (Bold Italicized), labelled cysteines are in red.

Figure 5. A summary of the accessibility of cysteines in type I and II keratins. The two chains (I and II) are indicated by different shades of the same color. K31 and K85 are examples that demonstrate a general trend.

Figure 6. A schematic representation of the disulfide crosslinks between a pair of antiparallel molecules in the oxidized tetramer. The disulfide crosslinks are between cysteines at positions L1-9 and 1B-87 in type I

keratins in the A_{11} alignment and two cysteines at position 2-90 in type I keratins in the A_{22} alignment. Two other cysteines (2-132 and T-5) thought to be involved in crosslinking are also marked on the diagram. The two chains (I and II) are indicated by different shades of the same color.

References

1. Shorland FB, Gray JM. The preparation of nutritious protein from wool. *British Journal of Nutrition.* 1970;24:717-725.

2. Parry DAD, Steinert PM. *Intermediate filament structure.* .Heidelberg, Germany: Springer-Verlag; 1995.

3. Parry DAD. Hendecad repeat in segment 2A and linker L2 of intermediate filament chains implies the possibility of a right-handed coiled-coil structure. *Journal of Structural Biology.*2006;155:370–374.

4. Parry DA, North AC. Hard α-keratin intermediate filament chains: substructure of the N- and C-terminal domains and the predicted structure and function of the C-terminal domains of type I and type II chains. *Journal of Structural Biology.* 1998;122(1-2):67-75.

5. Strnad P, Usachov V, Debes C, Gräter F, Parry DA, Omary MB. Unique amino acid signatures that are evolutionarily conserved distinguish simple-type, epidermal and hair keratins. *Journal of Cell Science.* 2011;124:4221-4232.

6. Plowman JE, Thomas A, Perloiro T, Clerens S, de Almeida AM. Characterisation of white and black merino wools: a proteomics study.*Animal.* 2018:1-7.

7. Maclaren JA, Milligan B. *Wool Science. The chemical reactivity of the wool fibre.* Marrickville, NSW, Australia: Science Press; 1981.

8. Fujikawa H, Fujimoto A, Farooq M, Ito M, Shimomura Y. Characterisation of the human hair keratin-associated protein 2 (KRTAP2) gene family. *Journal of Investigative Dermatology.*2012;132:1806-1813.

9. Fraser BRD, MacRae TP, Sparrow LG, Parry DA. Disulphide bonding in a-keratin. *International Journal of Biological Macromolecules.*1988;10(April):106-112.

10. Plowman JE, Flanagan LM, Paton LN, Fitzgerald AC, Joyce NI, Bryson WG. The effect of oxidation or alkylation on the separation of wool keratin proteins by two-dimensional electrophoresis. *Proteomics.*2003;3:942-950.

11. O'Cualain RD, Sims PFG, Carr CM. Structural analysis of alphahelical proteins from wool using cysteine labelling and mass spectrometry.*International Journal of Biological Macromolecules.*2011;49:323–330.

12. Deb-choudhury S, Plowman JE, Rao K, et al. Mapping the accessibility of the disulfide crosslink network in the wool fiber cortex.*Proteins* 2015;83:224-234.

13. Fraser RDB, Parry DAD. Intermediate filament structure in fully differentiated (oxidised) trichocyte keratin. *Journal of Structural Biology.* 2017;200:45-53.

14. Plowman JE, Deb-Choudhury S, Thomas A, et al. Characterisation of low abundance wool proteins through novel differential extraction techniques. *Electrophoresis* 2010;31:1937-1946.

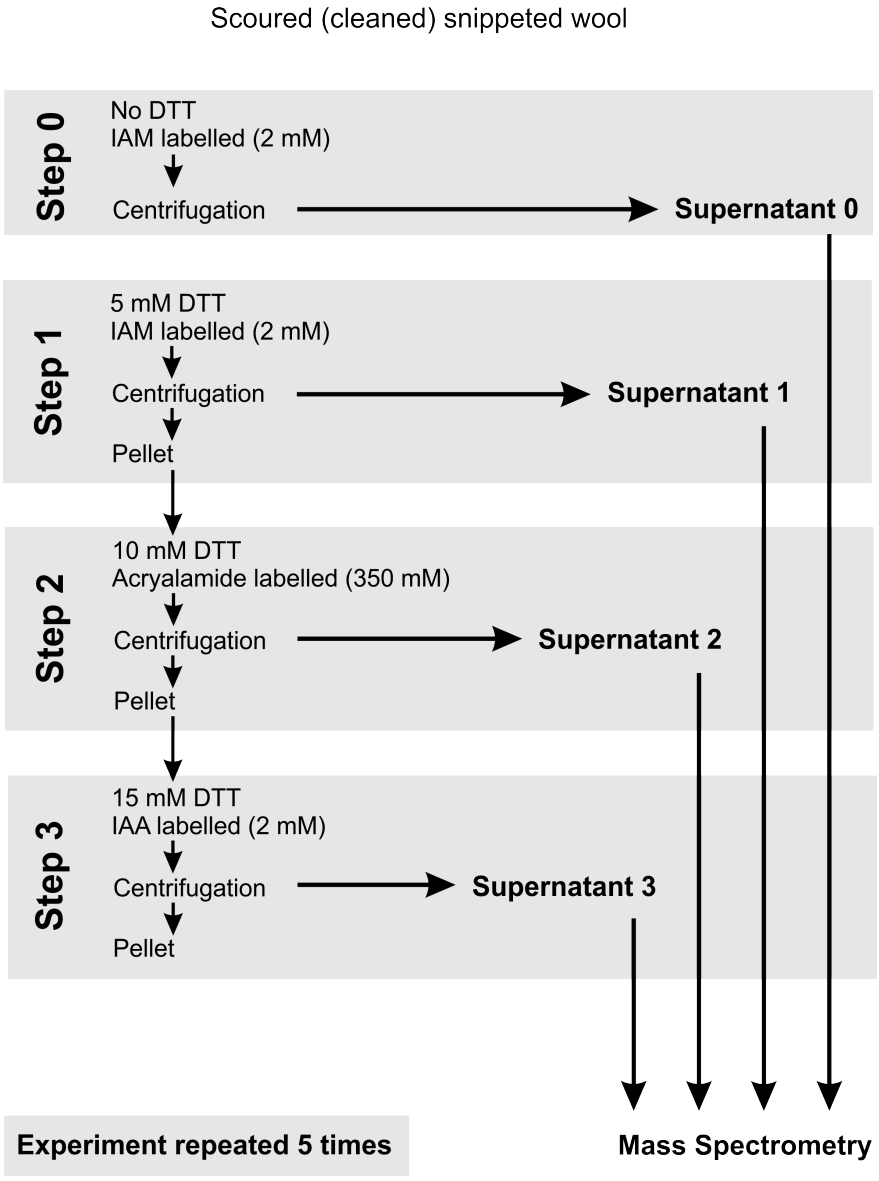
15. Parry DAD, Strelkov SV, Burkhard P, Aebi U, Herrmann H. Towards a molecular description of intermediate filament structure and assembly.*Experimental Cell Research.* 2007;313:2204–2216.

16. Steinert PM, Parry DAD. The conserved H1 domain of the Type II keratin 1 chain plays an essential role in the alignment of nearest neighbour molecules in mouse and human keratin 1/keratin 10 intermediate filaments at the two- to four-molecule level of structure. *Journal of Biological Chemistry.* 1993;268:2878-2887.

17. Fraser RDB, Parry DAD. Structural changes in the trichocyte intermediate filaments accompanying the transition from the reduced to the oxidized form. *Journal of Structural Biology.*2007;159(1):36-45.

18. Fraser RDB, Parry DAD. The role of disulfide bond formation in the structural transition observed in the intermediate filaments of developing hair. *Journal of Structural Biology.* 2012;180:117-124.

19. Wang H, Parry DAD, Jones LN, Idler WW, Marekov LN, Steinert PM. In vitro assembly and structure of trichocyte keratin intermediate filaments: a novel role for stabilization by disulfide bonding.*Journal of Cell Biology.* 2000;151(7):1459-1468.



KAP2.3
TGSCCGPTFS SLSCGGGCLQ PRYYRDPCCC RPVSCQTVSR PVTFVPRCTR
PICEPCRRPV CCDPCSLQEG CCRPITCCPT SCQAVVCRPC CWATTCCQPV
SVQCPCCRPT SCQPAPCSRT TCRTFRTSPC C

KAP2.3 Var 1
TGSCCGPTFS SLSCGGGCLQ PCCYRDPCCC RPVSCQTTVS RPVTFVPRCT
RPICEPCRRP VCCDPCSLQE GCCRPITCCP TSCQAVVCRP CCWATTCCQP
VSVQSPCCRP TSCQPAPCRT TCRTFRTSPC C

KAP2.3 Var 3
TGSCCGPTFS SLSCGGGCLQ PRYYRDPCCC RPVSCQTTVS RPVTFVSRCT
RPICEPCRRP VCCDPCSLQE GCCRPITCCP TSCQAVVCRP CCWATTCCQP
VSVQSPCCRP TSCQPAPCRT TCRTFRTSPC C

KAP3.1
MACCAPRCCS VRTGPATTIC SSDKFCRCGV CLPSTCPHNI SLLQPTCCDN
SPVPCVYPDT YVPTCFLLNS SHPTPGLSGI NLTTFIQPGC ENVCEPRC

KAP3.3
ACCARLCCSV PTSPATTICS SDKFCRCGVC LPSTCPHTVW LLQPTCCCDN
RPPPYHVPQP SVPTCFLLNS SQPTPGLESI NLTTYTQSSC EPCIPSCC

KAP3.4
ACCARLCCSV PTSPATTICS SDKFCRCGVC LPSTCPHTVW FLQPTCCCDN
RPPPCHIPQP SVPTCFLLNS SQPTPGLESI NLTTYTQPSC EPCIPSCC

KAP4.2
VSSCCGSVCS DQSCGRSLCQ ETCCRPSCCQ TTCCRTTCYR PSCGVSSCCR
PICCQPTCPR PTCCISSCSR PSCCVSSCGS SCYRPTSCIS SCRPQCCQPV
CYQPTCSRPT CCISSCYRPS SCGSSCGSSC CRPTCCISSC RPRCCQSVCC
QPSCPRISSC CRPSCCGSSC CRPSCCLRPV CGRVSCHTTC YRPTCVISTC
PRPVSCPSSC C

KAP4.5
VSSCCGSVCS DQSCGRSLCQ ETCCRPSCCQ TTCCRTTCYR PSCGVSSCCR
PVCCQPSCPR ISSCCRPSCY SSSCCRPSCC LRPVCGRVSC HTTCYRPTCV
ISTCPRPVCC PSSCC

KAP6.1
CGYYGNYYGG LGCGSYSYGG LGCGYGSCYG SGFRRLGCGY GCGYGYGSRS
LCGSGYGYGS RSLCGSGYGC GSGYGSGFGY YY

KAP6.4
CGYYGNYYGS LGCGSYGYGG LGCGYGSCYG SGFRRLGCGY GCGYGYGSRS
LYGCGYGCGS GYGSGFGYYY

KAP7.1
TRFFCCGSYF PGYPSYGTNF HRTFRATPLN CVVPLGSPLG YGCNGYSSLG
YGFGGSSFSN LGCGYGGSFY RPWGSGSGFG YSTY

KAP9.2
THSCCSPCCQ PTCCESSCCR PCCPPTCYQT SEHTCCRTTC SKPTCVTTCC
QPACGGSSCC QPCCRPISCQ TTCCRTTCLK PTCVTTCCQP TCCESSCCRP
CCPPTCYQTS ENTCCRTTCS KPTCVTTCCQ PACGGSSCCQ PCCRPISCQT
TCCRTTCLKP VCATTCCQPA CCESSCSQPS CPQTCCQITE TTCCKPTCVT
SCCQPTCCGS SSCGQPCGGS NCCQPASCAP VYCHRTCYHP TCCCLPGCQA
QSCGSSCCQP CSRPVCCQTT CCRTTRCRPS CVSSCCQPSC C

KAP9.8
THSCCSPCCQ PTCCESSCCQ PCCRPISCQT TCCRTTCLKP VCATTCCQPT
CCESSCSQPS CPQTCCQITE TTCCKPTCVT SCCQPTCCAS SSCGQPCGGS
NCCQPASCAP VYCHRTCYHP TCCCLPGCQA QSCGSSCCQP CSCPVCCQTT
CCRTTRCRPS CVSSCCLPSC C

KAP10.4-like Iso X1
ASSALSVCSS DLSYGSRVCL PGSCDSYTGS SWQVDDCPES CCEPPGCAPS
CCTPAPRLTL LCAPVSCESS PCSQPACSSS CLDLYCQQSS CQPSCCTSSP
CQLACCEPVC CTPVCCRPIC CRPAPCSASL CCQPNPCSSV RCRPSSSVSL
LCRPVCRPAC CVPTSSCQPS CCRPASCLSL LCQPKCSHPA CCIRTSAPEP
TIVKTTPALM RGSCCQLSTP TRKGRPGSLL WGQHKPGARY KSPTSPSTLH
TQLIHPHARK PSSSPTTSTM AASTLSICSS DLSYDCPESC CEPPCCAPSC
CTPAPRLTLL CAPVSCESRP CCQPACSSSC PALCCQQSSC QPSCCTSSPC
QQACCEPVCC RPVCCRPVCC TPVCCRPVCC EASPCLAPSS CCRPSSSVSL
LCRPVCRPTC CVPDSSCQPS CCRPASSVSL LCRPVCSRPA CCIPTLAPEP
CC

KAP10.11 Like
AASTLSICSS DLSYDCPESC CEPPCCAPSC CAPAPRLTLL CAPVSCESSP
CCQPACSSSC PALCCQQTSC QPSCCTSSPC QQACCEPVCC RPVCCRPVCC
TPVCCTSVCS EGSPCSAPSS CCRPSSSVSL LCHPVCRPAC CVPASSCQPS
CCRPASSVSL LCRPVCRPAC CVPTSSCQPS CCRPASSVSL LCRPVCRPAC
CVPTSSCQPS CCRPASSVSL LCQPACSRPA CYVPALALEP
CC

KAP10.12 Like Iso X2
AFSNLSACSS DLSYSSRICL PGSCDSCTGS SWQVDDCPES CCEPPCCAPS
CCTSAPRLTL LCAPVSCESS PCCQPACSSS CPASCGQQSS CQSSCCTSSR
CQQACCEPVC CEPVCCRPVC CRPVCCTPVC CTPVCCRPVC CEASPFSTSS
CCQQSSCQPS CCTSSCCQQA CCEPVCCRPV CCTPVCCRPV CCETSPCSAP
SSCCRPSSSV SLLCRPVCRP VCGPTCCVPT SSCKPSFCRP ASSVSLLCRP
ACRLACCVPT SSCQPSCCRP ASSVSLLCRP ACSRLACGIP TSAPELCC

KAP11.1
SYSCSTRNCS SRRIGGEYTV PVVTVSSPDA DCLSGIYLPS SFQTGSWLLD
HCQETCCEPT VCQSTCYQPT PCVSSPVRVT SRQTTCVSSP CSTTCSRPLT
FISSGCQPLS GVSTVCKPVR SISTVCQPVG GVSTICQPTC GVSRTYQQSC
VSSCRRIC

KAP12.2-like
CDNSCPASCQ PTCCVPSSCQ ETCCVPSSCQ PACPALCCPA LGCRNPCGAS
RVALLCRPAC CVPSPCQASC CVPLSCKPVL YVPMCYKHIV YVIPSCQPSR
FCQPSYPSLV CRPVPCLSLK CF

KAP13.1-like
SYNCCSGNFS SRSLQDHLRT PASSFPSNLV YSTDLCSPSS CQLGSSLYSQ
ETCCEPIRTQ TVVSRPCQTS CYRPRTSTFS SPCQTTFPGS LAYRSSSCSS
LSSGSRSCYS VGCGGRGFRR LGYGICGFPV LSSGSGFCRP TFFASR*SCPS*
SCYRPTCGSV FY

KAP15.1
SFNCSTGNFS RSLGGYLGVP VSTCDSFYPS NVVYSPSTFQ LGSTLYSNCQ
ENFFRPVSFQ TPCAVTRFFQ TSCSHPQNFI FRSPCQTIYT GSLGSGNIGF
GSFGCGNTGF QSLGCGSNFC SPTYVSSRSC RSSYY

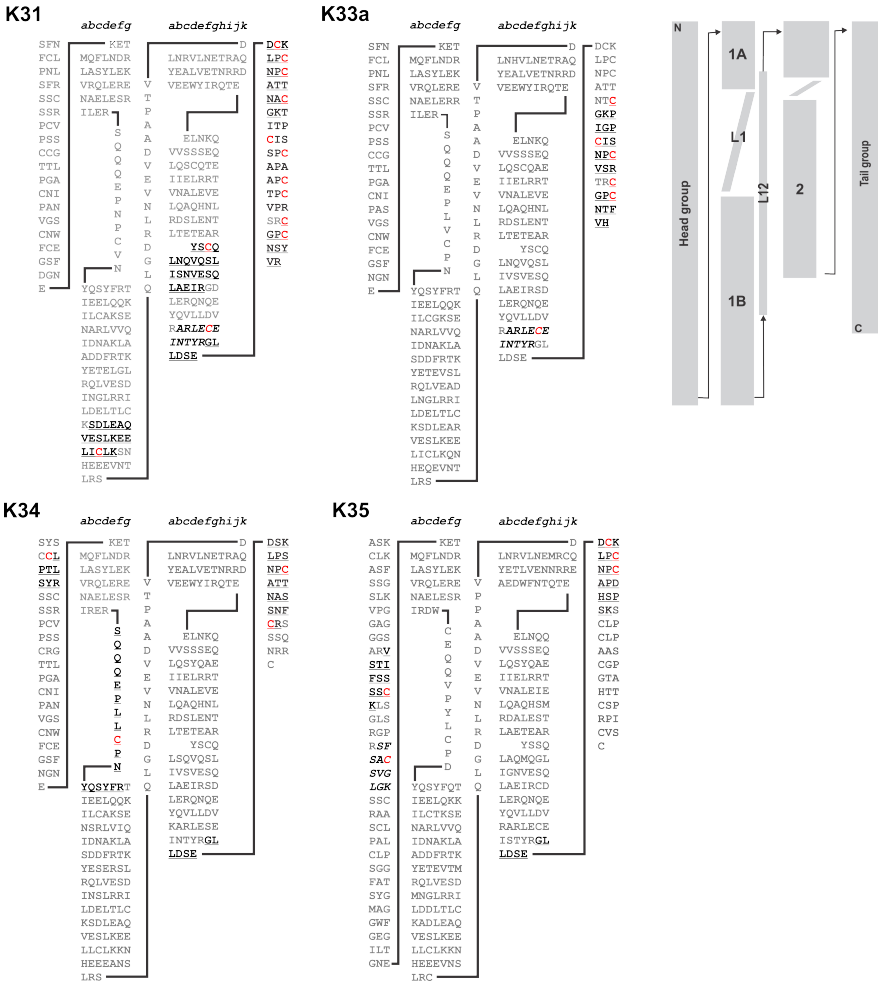
KAP16.1 Iso X1
SGNCCSRKCP SVPAISLCST EVSCRGPVCL PSSCRSQTWQ LVTCQDSCGS
SSCDPQCCEP SCSASSCVQP VCCETTICEP ACPVSSRAQP VSCEATICEP
ACPVSNCAQP VCYKATICEP ACPVSSCAQP VYCEATIYEP ACPVSSCAQP
VCYKATICEP ACPISSCAQP VCCEATICEP SCSVSSYAQP VSYKATICES
ACPVSSCAQP VCEPACPVSS RAQPVSCEAP ICEPACPVSN CAQPVCYKAT
ICEPACPVIS CAQTVCYETT ICEPACPVSS CAQPVYCEAT IYEPACPVSN
CAQPVCYKAT ICEPACPVIS CAQTVCYETT ICEPACPVSS CAQPVYCEAT
ICEPSCSVSS CAQPVCCEAT ICEPSCSVSS YAQPVSYKAT ICESACPVSS
CAQPVYYKAT ICEPSCPVSS CAQPVSCEAT ICEPSCSVSS CAQPVSYKAT
ICEPSCSVSS CAQPVSYKAT ICEPACPVSS CAQPVSCEAT ICEPSCSVSS
YAQPVSYKAT ICEPSCSVSS CAQPVSYKAT ICEPACPVSS CAQPVSYKAT
ICEPSCSVSS CAQPVSYKAT ICEPSCSVSS CAQPVSYKAT ICEPACPVSS
CAQPVSYKPP GQRVFCVPSS CQPILCKPSY CQPVICEPSC YQPVSSGVRC
CPSVCSVANS CQSACCDSSP CEPSCSEPSI CQSATRVSLV CEPICVRPVC
CVSSPCEPPC VSSTCQEPSC CVSSICQPIC SEPSPCLPSV CVPRPCQPTC
YVVKRSRSIS CEPLSCRPLS GRPGSSASAV CQPTCSRTFY IPSSCKQPCT
TSTSYRPICR PICSGPITYR QPYLTSISYR PACYRPFYSI LRRPACIASV
PYRSVCSRLP CADSCKRDCK KSTSSQPDCA DSTPCKTEVS EASPCQPTEA
KPTSPTTREA AVSQPAATKP TNC

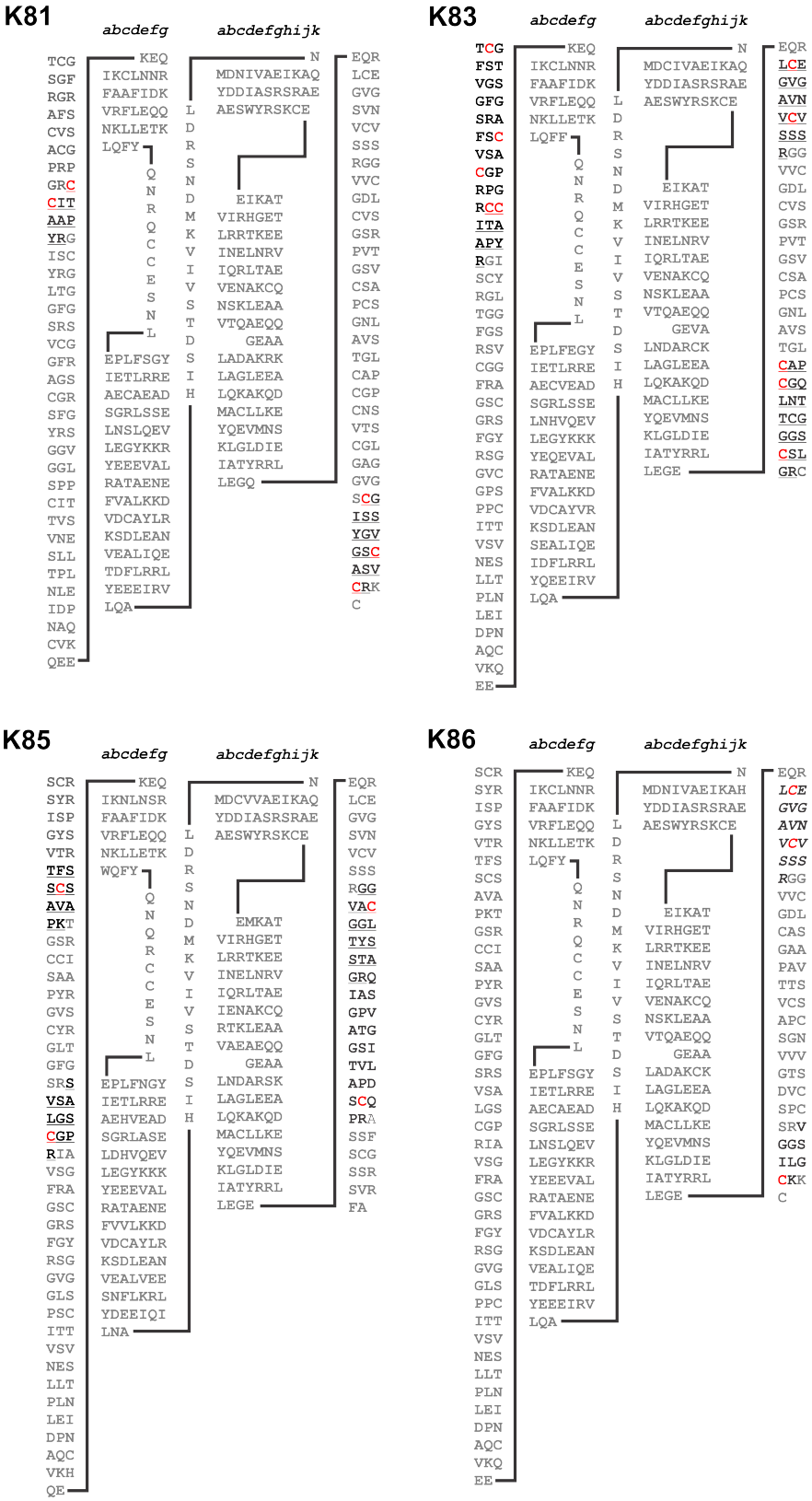
KAP19.3
CHYSNYYSGL GYGYGGFGGL GFGRGCGCGS FRRLGFSTGF GGYGCGSGFG
SFGYGYNRCP SSFGRYGFSS FY

KAP19.5
CHYSNHYRGL GYGYGGFGGL GFGRGCGCGS FRRLGCGSGF RGYGCDSGFG
SFGSGCCHRP LFFRRCGFSS FY

KAP24.1
NIHPGSMAFL GYPGNCSGVS YRTHYYFPVT GSVALCSRHV SPTFGLSLPS
SYHGNLWLLD NCQETCGEAP TCESPCSEPK TCTTTCDQSN SSVPCNSPTG
GQICSARETT NIGPSLSCNQ CPQTKGYVSD GCTPSRHTSK ACQTLGNGFK
CFGQLNCLSK SFQPLSHYRL GSFGYRSYQD LGFIPSGFSA SRYITN

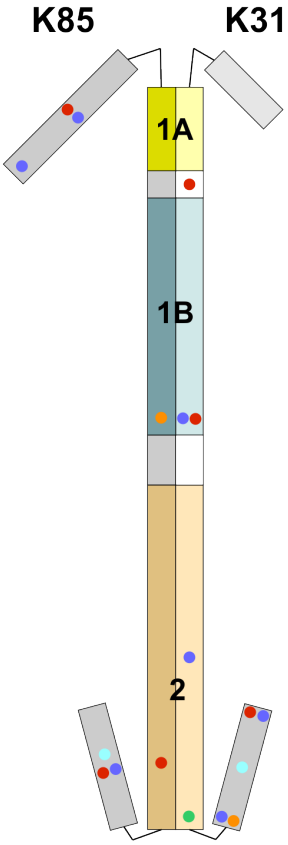
KAP26.1
SCHNYCSGNY TLGSLRSPCH IPVASSVGLY TPSVSVGDGL CLPSSCQDRT
WILSNGQETC SEPTSCQPAN CEPSSCETSS YPSSGCYVPR PSQGTSFLPA
SSYLSGSCLS VSYRPLAYVS SSSRPLSLLP CGYRPSGSLP CGLQPISIVS
SGLRPLRPVF SGCQTLPYVF SPYRPSYSTC GGL

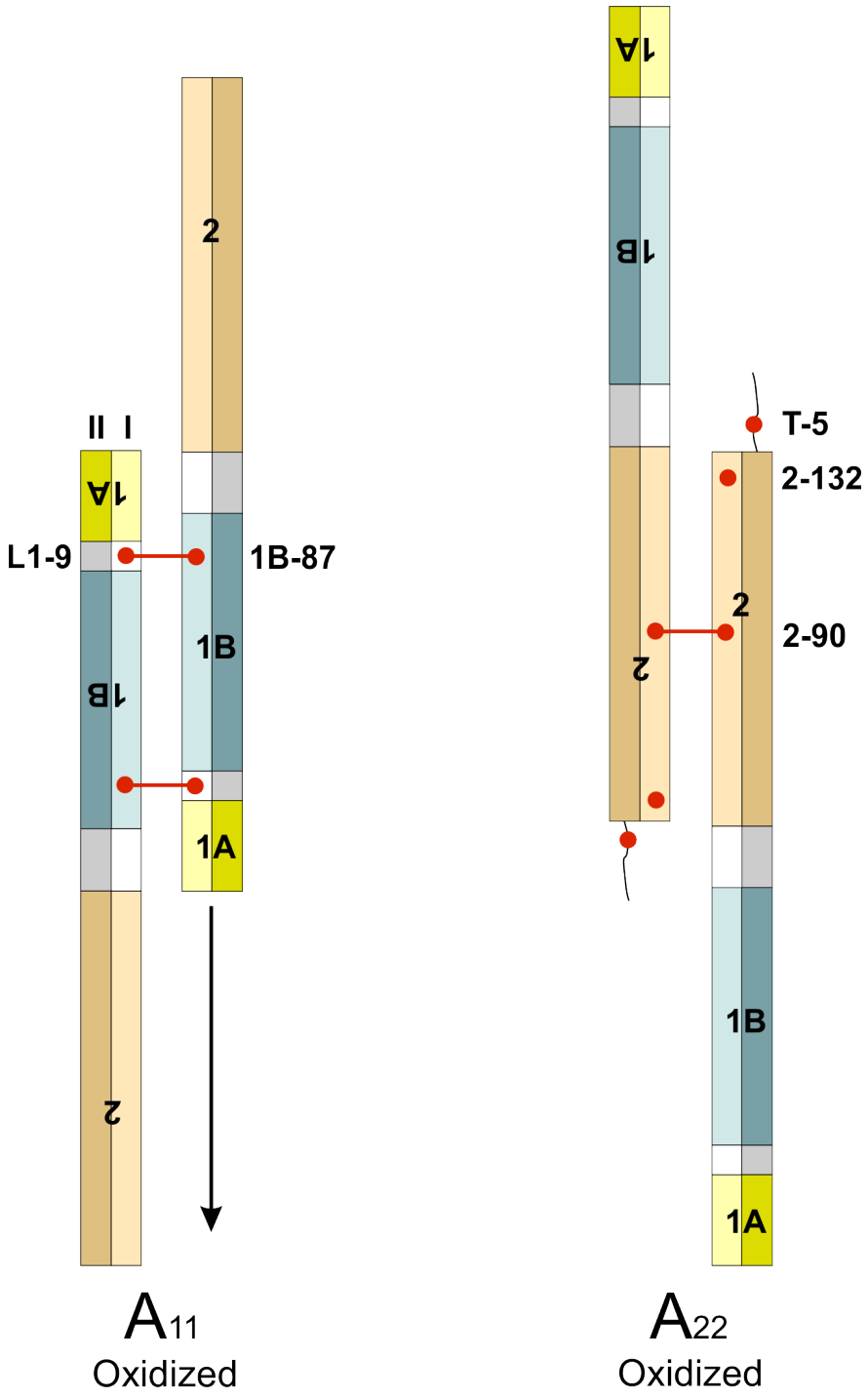




Ordinal label	0	1	2	3	4	5	6
mM concentration DTT	Nil	5	10	15	20	30*	20**
Deb-Choudhury et al 2015	1				2	3	4
This study	0	1	2	3			

* + urea
** + urea + extended incubation





Hosted file

Table 1.docx available at https://authorea.com/users/356211/articles/479161-a-detailed-mapping-of-the-readily-accessible-disulfide-bonds-in-the-cortex-of-wool-fibers

Hosted file

Table 2.docx available at https://authorea.com/users/356211/articles/479161-a-detailed-mapping-of-the-readily-accessible-disulfide-bonds-in-the-cortex-of-wool-fibers

Hosted file

Table 3.docx available at https://authorea.com/users/356211/articles/479161-a-detailed-mapping-of-the-readily-accessible-disulfide-bonds-in-the-cortex-of-wool-fibers