Mapping Protein Cross-Links in Human Hair via Mass Spectrometry

EVELYNE MAES, JOLON M. DYER, SANTANU DEB-CHOUDHURY AND STEFAN CLERENS

Lincoln Research Centre, AgResearch Ltd, Christchurch, New Zealand (E.M., J.M.D., S.D.-C., S.C.)

Riddet Institute, Massey University, Palmerston North, New Zealand (E.M., J.M.D., S.C.)

Biomolecular Interaction Centre, University of Canterbury, Christchurch, New Zealand (E.M., J.M.D., S.C.)

Synopsis

Networks of protein–protein cross-links underpin all the key physicomechanical properties of mammalian fibers and also strongly influence how the fiber responds to treatments and environmental insults. During the last decade, significant improvement has been made in detecting and understanding cross-links, though, to date, mapping where these protein cross-links are within and between proteins has proved to be very challenging. This work reports a proof-of-concept study where mass spectrometry–based proteomics strategies were used to unravel the details of cross-link location between trichocyte keratin proteins in the hair shaft. This work focuses on two cross-links known to exist in hair proteins and that are used as a proxy for the degree of damage in hair proteins: lanthionine and lysinoalanine. Our results demonstrate that mass spectrometric evidence of the existence of both lanthionine and lysinoalanine cross-linked peptides within hair fibers can be found. The approach used also provided the first insight of the exact location of these cross-links within specific keratins. Analysis with respect to the current models of trichocyte intermediate filament organization indicates detected cross-links occur within antiparallel keratin heterodimers. The low abundance of the cross-linked peptides makes this type of evaluation approach difficult, but the results represent a new approach to untangle which cross-links are most essential to defining hair performance.

INTRODUCTION

Human hair is a complex tissue with fibers consisting of three morphological compartments composed of cornified dead cells: the cuticle, the cortex, and the medulla. The medulla is a loosely packed region in the center of the hair shaft, surrounded by the cortex, which represents the major structure in hair. To protect the cortex against the outside environment, a layer of overlapping cells called the cuticle surrounds the cortex (1). The hair cortex is primarily composed of proteins (90–95%), with lipids, water, and trace elements as the other chemical constituents present. The cortical proteome is dominated by two classes of proteins: first, intermediate filament–forming trichocyte (or alpha) keratins, and second, keratin-associated proteins that surround the keratin filaments. The keratins associate

---

together to form coiled-coil dimers, intermediate filaments, and macrofibrils (2). To achieve a rigid hair structure, they form a strong compact network by both intra- and intermolecular interactions such as disulfide bonds, hydrogen bonds, electrostatic salt bonds, and amide bonds (3). Of major interest in hair science are the protein cross-links, such as disulfide bonds, which lead to the formation of covalent linkages between intermolecular amino acid residues and hence determine which regions within peptides and proteins are involved in the formation of protein networks. In addition to these native protein cross-links in hair, artificial cross-links, including lanthionine and lysinoalanine, can also be induced, mostly on insult such as heat, UV, or the use of alkaline products (4).

Networks of protein–protein cross-links underpin all the key physicomechanical properties of mammalian fibers and strongly influence how the fiber responds to treatments and environmental insults. Understanding and controlling these networks is therefore critical to enhance our understanding of fiber science generally and develop new fiber treatments that enhance properties. However, due to the large degree of cross-linking present in mammalian fibers, analysis of these covalent bonds has been very difficult. Although some technologies have successfully led to the detection of cross-links (3), most techniques involve total protein destruction, making mapping where these cross-links are within and between proteins extremely difficult.

In this proof-of-concept study, mass spectrometry–based proteomics strategies are used to unravel the details of cross-link location between trichocyte keratin proteins in the hair shaft. This work focuses on two cross-links known to exist in hair proteins and that are used as a proxy for the degree of damage in hair proteins: lanthionine and lysinoalanine. Both cross-links can be the result of degraded disulfide bonds to form dehydroalanine, or alternatively, of the elimination of water or hydrogen sulfide from serine and cysteine, respectively, to form dehydroalanine. This dehydroalanine intermediate will react with either lysine to form lysinoalanine or with cysteine to from lanthionine. Although mass spectrometry has been used before to elucidate protein cross-links in a food or medical context (5,6), this work, to our knowledge, is the first to decipher cross-links in the complex hair matrix with the aim of finding mass spectrometric evidence of the existence of cross-links within hair fibers and mapping their exact location within the protein *in situ*.

METHODS

SAMPLE PREPARATION

Blended Caucasian virgin hair tresses were obtained from International Hair Importers & Products, Inc. (Glendale, NY, United States). Full-length fibers were cut into snippets of approximately 5 mm in length.

For amino acid analysis, two types of samples were prepared: first, 10 mg of hair snippets were weighted, and second, hair protein extracts were prepared by using 100 mg of hair snippets, which were dissolved in 5 mL of an extraction buffer (7 M urea, 2 M thiourea, 0.05 M Tris, and 50 mM dithiothreitol (DTT), pH 7.5) at room temperature for 18 h using vigorous shaking. For proteomics analysis, 10 mg of hair snippets were dissolved in 1 mL of an extraction buffer (7 M urea, 2 M thiourea, 0.05 M Tris, and 50 mM DTT, pH 7.5) at room temperature for 18 h using vigorous shaking.

PROTEIN DIGESTION

After the protein extraction, an acetone precipitation was carried out to remove the extraction buffer components by adding 5 mL of ice-cold acetone to 1 mL of the sample. The samples were stored at –20°C for 48 h before they were centrifuged at 10,000 g and 4°C for 10 min. The protein pellet was washed twice with ice-cold acetone, air-dried, and resuspended in 200 μL of 10 mM ammonium bicarbonate. The protein concentration per sample was determined by Direct Detect (Millipore, MA, United States) according to manufacturer's instructions. Next, 100 μg of each sample was aliquoted, and proteins were digested by sequencing grade trypsin (Promega) at a 1:50 ratio (trypsin:protein) for 18 h at 37°C at 300 rpm in a thermomixer. Next, the tryptic digest peptide mixtures were stored at –20°C until further analysis.

AMINO ACID ANALYSIS

The hair snippets and the tryptic digests were hydrolyzed with 6 N HCl at 110°C for 24 h. Amino acid analysis was performed using a Dionex HPLC system (Dionex UltiMate 3000 with Fluorescence Detector, Thermo Scientific, San Jose, CA, United States). Free amino acid residues and standards (containing lanthionine and lysinoalanine at a predetermined concentration) were derivatized with an AccQ-Tag reagent (Waters Corporation, Milford, MA, United States) and separated using a Thermo Accucore XL C18 column (4.6 mm i.d. × 250 mm, 4 μm particles) (Thermo Scientific, San Jose, CA, United States) at 37°C using a flow rate of 1.0 mL/min. Eluent A was AccQ-Tag solvent A, and eluent B was 100% acetonitrile. An excitation wavelength of 250 nm and emission wavelength of 395 nm were used for the quantitative analysis using a fluorescence detector (Dionex Ultimate FLD 3000, Thermo Scientific, San Jose, CA, United States).

LIQUID CHROMATOGRAPHY–TANDEM MASS SPECTROMETRY

To detect the cross-links via mass spectrometry, the peptide mixtures were separated on an Ultimate 3000 RSLCnano system (Thermo Scientific, San Jose, CA, United States) using a C18 PepMap100 nano-Trap column (200 μm i.d. × 2 cm) (Thermo Scientific, San Jose, CA, United States) connected to a ProntoSIL C18AQ analytical column (100 μm i.d. × 15 cm, 3 μm particle size, 200 A pore size) (nanoLCMS Solutions, Gold River, CA, United States). For each sample 1 μL (equal to 1 μg) was loaded on the trap column at 3 μL/min. The trap column was then switched in line with the analytical column. Reverse phase liquid chromatography (LC) separation was performed by a linear gradient of mobile phase B (0.1% formic acid in 100% acetonitrile) from 2% to 45% in 60 min, followed by a steep increase to 98% mobile phase B in 6 min, held at 95% mobile phase B for 2 min, returned to 2% mobile phase B over 5 min, and then re-equilibrated at 2% mobile phase B for 15 min resulting in a total run of 88 min used at a flow rate of 1,000 nL/min. The LC was coupled online to mass spectrometry via a Bruker CaptiveSpray ion source (Bruker Daltonics, Bremen, Germany) equipped with a nanobooster device and operated at 1,400 V.

Data were acquired with a Q-TOF Impact II (Bruker Daltonics, Bremen, Germany) mass spectrometer in a dynamic data-dependent auto-MS/MS mode where a full scan spectrum (150–2,200 m/z, 2 Hz) was followed by a dynamic inclusion of collision-induced dissociation

tandem mass spectra (2–32 Hz, depending on intensity) for 3 s. A preference of selection of multiple-charged precursors (2+ to 5+) was made, and an isolation width of 0.03 Th was set. Dynamic exclusion duration was set to 20 s.

PROTEIN IDENTIFICATION

Potential protein–protein cross-links were searched with Kojak v1.6 (7). The following parameters were used: a precursor mass tolerance of 10 ppm was allowed with fragment bin offset $= 0$, fragment bin size 0.03. A custom in-house database was created containing all human keratin protein sequences, keratin-associated proteins, and their reverse complements as decoy proteins. Trypsin was specified as a digestive enzyme, and up to three missed cleavages were allowed. A minimum peptide mass of 245 Da was set. Carbamidomethyl (C), oxidation (M), and deamidation (NQ) were chosen as variable modifications, and a maximum of three modifications per peptide was allowed. A lanthionine bond (Cys–Cys or Ser–Cys) or lysinoalanine bond (Lys–Cys) was specified as a cross-link. For each spectrum, the top 20 scoring peptides were used for cross-link combination. To minimize false positives, peptide spectrum matches exported from Kojak were further validated by Percolator V2.08 (8). The threshold for peptide spectrum matches after validation was set at q value smaller than 0.01.

RESULTS

CONFIRMATION OF CROSS-LINK PRESENCE IN HAIR FIBERS AND THEIR PROTEIN EXTRACTS

Amino acid analysis was used to validate the presence of lysinoalanine or lanthionine cross-links in samples that were used for proteomics cross-link mapping. The amount of lanthionine and lysinoalanine present both in hair fiber snippets (one biological replicate run in duplicate) and hair snippet protein extracts (three separate protein extracts run in duplicate) is shown in Figure 1. Cross-links were detected in both the fiber and in the protein extract from the fiber, though some variability in the samples was noticed.
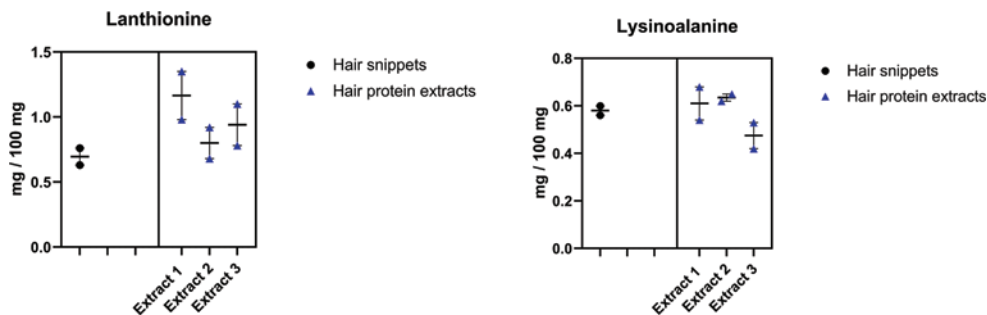

Figure 1. The quantities of lanthionine and lysinoalanine cross-links (mg/100 mg) present in hair snippets (black) and triplicate hair extracts (blue) as determined by amino acid analysis represented with 95% confidence intervals.

LOCATION OF LANTHIONINE CROSS-LINKS WITHIN HAIR PROTEIN EXTRACTS VIA MASS SPECTROMETRY

As cross-linked peptides are very challenging to characterize by mass spectrometry, a data acquisition method that allows for better fragmentation of these low-abundant cross-linked peptide precursors was developed. By allowing a more dynamic sampling of precursors in the mass spectrometer, a higher chance of detecting low-abundant, multiple-charged precursors was created.

To maximize the chance of detecting lanthionine and lysinoalanine cross-links between keratin-related proteins and to minimize the amount of false positive hits within this process, we designed a custom database that consisted of all human keratins and keratin-associated proteins, as it is believed that cross-links formed within these protein families are largely responsible for the hair fiber rigid structure. Using this FASTA file (2019.02), we analyzed for cross-links between 298 proteins (including decoy proteins), resulting in 22,204 *in silico* tryptic peptide combination possibilities, of which 8,759 peptides are linkable.

Table I lists lanthionine cross-links that could be detected in the hair digests. The table lists the two potential peptide sequences that are linked by a lanthionine cross-link as well as their corresponding first hit proteins. It also provides the spectrum number so that further manual validation of the fragmentation spectrum is possible. The spectrum-match validation settings were chosen as very stringent, and only *q* values smaller than 0.05 were considered confident. Because the algorithm only reports *q* values to three decimal places, the *q* values that made it through the filter are reported as 0.

A further look at the fragmentation spectrum of the detected cross-link is depicted in Figure 2. A few observations can be made. First, as lanthionine derives from dehydroalanine intermediates (6), the bond can exist between cysteine–cysteine or serine–cysteine in the cross-linked peptides. All the cross-linked peptides detected consisted of a cysteine–cysteine bond. Second, none of the annotated fragment ions are ions that are directly linked to the Cys–Cys bond, though several other ions are detected. Third, the detected fragment ions are still all very low abundance, yet through our dynamic acquisition method, just detectable. Fourth, the detected b-ions and y-ions originating from the two peptides, the dominant alpha peptide, as well as the beta peptide are indicated.

Table I
List of Lanthionine Cross-Links Found in Human Virgin Hair Shaft Digest

Scan number	*q* value	Peptide	Peptide	Protein 1	Protein 2
19,599	0	.LLEGEEHRLC[57.02] EGVGSVN VCVSSSRGGVSC [57.02]GGLSYSTTPGR	NVSVSPIDIGC[57.02] QPGAEAN IAPM[15.99] CLLANVAHAN [0.98]R	HUMAN_K85	HUMAN_K38
6,471	0	.QVVSSSEQLQ[0.98] SCQAEIIELR	ISSGCGVTRNFSS CSAVAPK	HUMAN_K34	HUMAN_K85
4,490	0	.SQQQEPLVCASYQSYFK	CEIGNVK	HUMAN_K33A	HUMAN_K74
4,479	0	.SQQQEPLVCASYQSYFK	CEIGNVK	HUMAN_K33A	HUMAN_K74
3,540	0	.SQQQEPLLCPSYQSYFK	MSCRSYR	HUMAN_K34	HUMAN_K84

Figure 2. Mass spectrometric evidence of the existence of lanthionine inter–cross-links between two peptides from keratin 38 and keratin 85. The data visualize scan number 19,599 of Table I.

A further look at the localization of the two peptides (red) within the full sequences of the respective proteins is visualized in Figure 3. Interestingly, the peptides of interest are mapped at opposite ends of the total sequence.

An analogous analysis was performed for the other potential cross-links detected. The lanthionine link between keratin 34 and keratin 85 is visualized in Figure 4. Similar to the first example, no direct proof of the cross-link via a fragment ion could be detected, though proof of existence of both peptide sequences (the alpha and beta sequence) within the fragment spectrum can be found (Figure 4). Although the cross-linked peptides between keratin 33A and keratin 74 were found in two spectra, most ion intensities were again very low; this might indicate that the bulk of the ion intensities are still too low to be confidently detectable by our method.

DETECTION OF LYSINOALANINE CROSS-LINKS WITHIN HAIR FIBER PROTEIN EXTRACTS VIA MASS SPECTROMETRY

A similar approach was also taken to explore whether there is mass spectrometric evidence of lysinoalanine bonds between keratins and/or keratin-associated proteins present in the tryptic digest of the extracted hair fiber proteins. Table II lists all lysinoalanine cross-links found with q values smaller than 0.05.

From this list, two cross-links of interest were further investigated on their potential. At first, mass spectrometric evidence was sought for the link between keratin 33B and keratin 80. Figure 5 displays the fragment spectrum with the detected b-ions and y-ions in blue and red, respectively. Figure 6 represents the location of the two peptides (red) within the full sequences of the respective proteins. Similar to the lanthionine data, proof of presence of both peptides was found as ions from both the alpha peptide and beta peptide were detected.

A closer look at the potential lysinoalanine cross-link between keratin 34 and keratin 82, alternatively, shows that most ions are found with a very low signal-to-noise threshold (Figure 7), indicating that the interpretation of these data should be approached with extra caution.

Figure 3. Protein sequences (human) of keratin 38 and keratin 85, with the peptide sequence involved in the cross-link indicated in red, the head domain underlined, and the tail domain in italics.
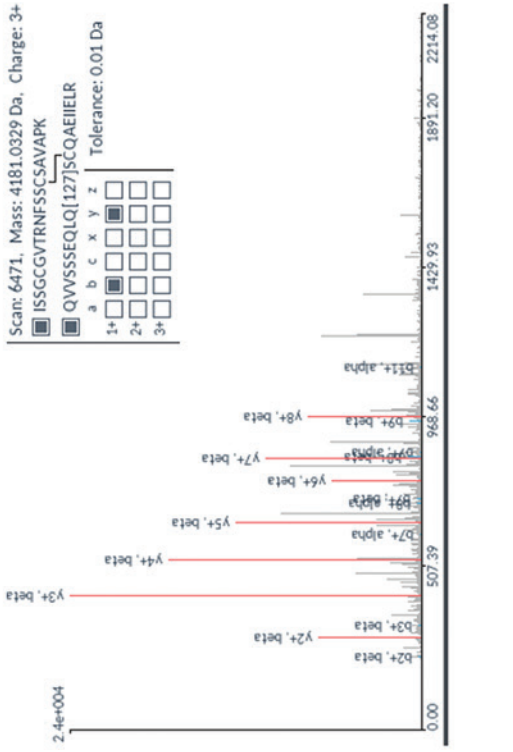
Figure 4. Another example of mass spectrometric characterization of an intermolecule lanthionine cross-link found between keratin 34 and keratin 85 within a tryptic digest of hair shaft-extracted proteins.

Scan number	q value	Peptide 1	Peptide 2	Protein 1	Protein 2
8040	0	.ADLETNAEALVQEIDFL KSLYEEEICLLQSQ[0.98] ISETSVIVKMDNSR	TMQDSVEDF KTKYEEEINK	HUMAN_K4	HUMAN_K82
19330	0	.GPC[57.02] RPGGGRGLRALG CLGSRSLCN[0.98] VGFGRPR	KN[0.98]HEEEVN [0.98]TLR	HUMAN_K82	HUMAN_K34
18602	0	.KDVDCAYLRKSDLEANVE ALVEESSFLR	KVLEEVDQR	HUMAN_K85	HUMAN_K25
17055	0.030	.N[0.98]Q[0.98]YEALVETN [0.98]RREVEQWFA TQTEELNKQVVSSSEQ LQSYQAEIIELR	SQILSVKSHCLK	HUMAN_ K33B	HUMAN_K80
17824	0	.SQYEALVETN RREVEQWFATQ[0.98] TEELNKQVVS SSEQLQSYQAEIIELR	SRQFTCKSGAAAK	HUMAN_ K33A	HUMAN_K71

DISCUSSION

Intermediate filaments are a key repeating structural unit of the human hair shaft, and their stability depends on cross-links both within filaments and between filaments via a network that extends through the interfilament matrix. The premise for this study was that the protein–protein cross-link network underpins health- and consumer-relevant physicomechanical properties of hair and that some cross-links are more important than others in their contribution to these properties. Location of individual cross-links is valuable to better understand the role of specific cross-links on properties and to investigate macromolecular relationships within and between keratin intermediate filaments.

In the last few decades, significant progress has been made in understanding the exact structure of the intermediate filaments through insights gained via amino acid analysis, computer modeling, x-ray fiber diffraction and protein crystallography, electron microscopy, and chemical cross-linking methods (2,9). These methods have helped us understand the structural detail of the hierarchical structure of keratins into heterodimers, tetramers, and 32-chain intermediate filaments. Currently, however, widespread pinpointing of the exact location of cross-links within and between the intermediate filament proteins is still very challenging. In this proof-of-concept study, we have, for the first time, applied mass spectrometric techniques to detect native cross-links present within the human hair cortex and to map the exact location of these cross-links within the proteins.

The use of mass spectrometry in protein cross-links has challenges, including, but not limited to, low abundance of cross-linked peptides, fragmentation of the cross-link peptides by the mass spectrometer, and the lack of automated data analysis for cross-linked peptides (5,6). Therefore, most of the work published on native protein cross-links to date involves a noncomplex mixture of proteins. Extending this technology to a complex hair shaft proteome comes with challenges in data complexity. Although disulfide bonds are the best characterized cross-links between keratin proteins in hair, their analysis in complex samples is complicated because typical proteomics approaches break disulfide bonds to
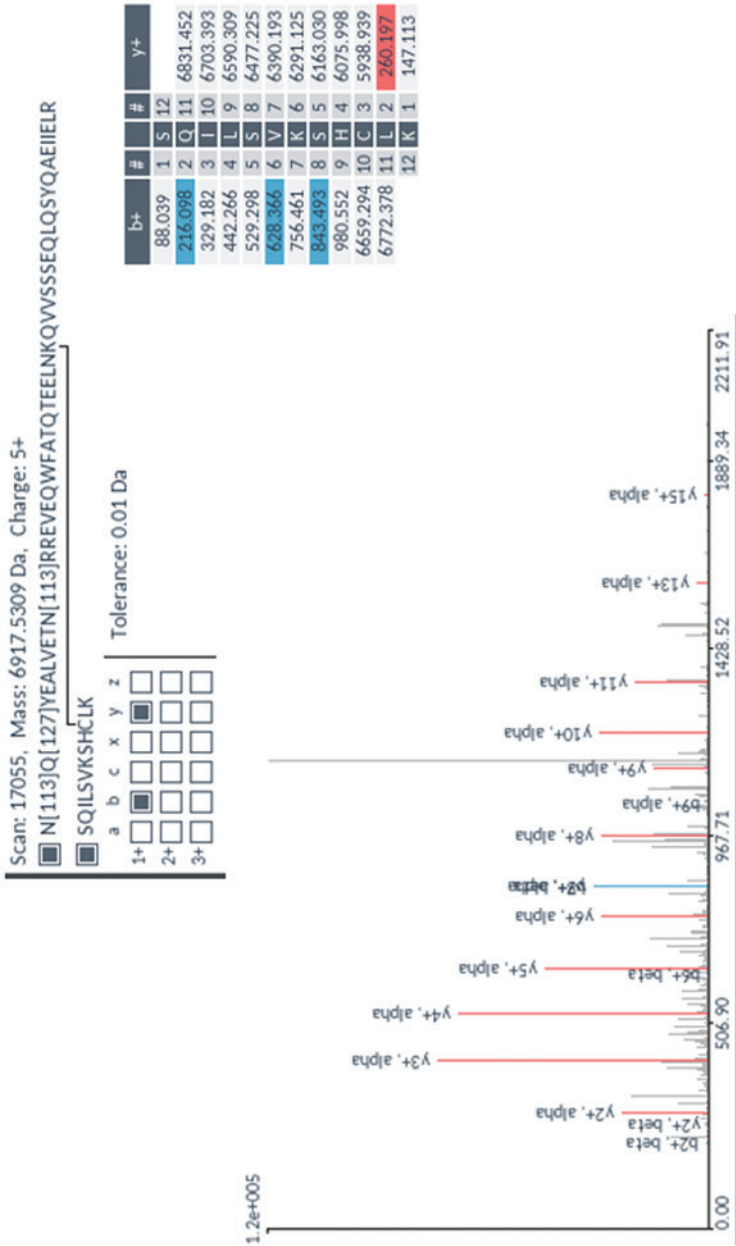
Figure 5. Mass spectrometric characterization of intermolecule lysinoalanine cross-links found between keratin 33B and keratin 80 within a tryptic digest of hair shaft-extracted proteins.

Keratin 33B

```
         10         20         30         40         50
MPYNFCLPSL SCRTSCSSRP CVPPSCHGYT LPGACNIPAN VSNCNWFCEG
         60         70         80         90        100
SFNGSEKETM QFLNDRLASY LEKVRQLERD NAELENLIRE RSQQQEPLLC
        110        120        130        140        150
PSYQSYFKTI EELQQKILCS KSENARLVVQ IDNAKLAADD FRTKYQTEQS
        160        170        180        190        200
LRQLVESDIN SLRRILDELT LCRSDLEAQM ESLKEELLSL KQNHEQEVNT
        210        220        230        240        250
LRCQLGDRLN VEVDAAPAVD LNQVLNETRN QYEALVETNR REVEQWFATQ
        260        270        280        290        300
TEELNKQVVS SSEQLQSYQA EIIELRRTVN ALEIELQAQH NLRYSLENTL
        310        320        330        340        350
TESEARYSSQ LSQVQSLITN VESQLAEIRS DLERQNQEYQ VLLDVRARLE
        360        370        380        390        400
CEINTYRSLL ESEDCKLPSN PCATTNACEK PIGSCVTNPC GPRSRCGPCN

TFGY
```

Keratin 80

```
         10         20         30         40         50
MACRSCVVGF SSLSSCEVTP VGSPRPGTSG WDSCRAPGPG FSSRSLTGCW
         60         70         80         90        100
SAGTISKVTV NPGLLVPLDV KLDPAVQQLK NQEKEEMKAL NDKFASLIGK
        110        120        130        140        150
VQALEQRNQL LETRWSFLQG QDSAIFDLGH LYEEYQGRLQ EELRKVSQER
        160        170        180        190        200
GQLEANLLQV LEKVEEFRIR YEDEISKRTD MEFTFVQLKK DLDAECLHRT
        210        220        230        240        250
ELETKLKSLE SFVELMKTIY EQELKDLAAQ VKDVSVTVGM DSRCHIDLSG
        260        270        280        290        300
IVEEVKAQYD AVAARSLEEA EAYSRSQLEE QAARSAEYGS SLQSSRSEIA
        310        320        330        340        350
DLNVRIQKLR SQILSVKSHC LKLEENIKTA EEQGELAFQD AKTKLAQLEA
        360        370        380        390        400
ALQQAKQDMA RQLRKYQELM NVKLALDIEI ATYRRLVEGE EGRMDSPSAT
        410        420        430        440        450
VVSAVQSRCK TAASRSGLSK APSRKKKGSK GPVIKITEMS EKYFSQESEV

SE
```

Figure 6. Protein sequences (human) of keratin 33B and keratin 80, with the peptide sequence involved in the cross-link indicated in red.

unfold proteins to enable enzymatic digestion and mass spectrometric analysis (10). This proof-of-concept study focused on lanthionine and lysinoalanine cross-links induced by environmental wear and tear. To ensure that the target cross-links were present in our samples, amino acid analysis against lanthionine and lysinoalanine standards measured the absolute quantity of cross-links present in hairs and protein extracts. Although amino acid analysis did not indicate the molecular location of target cross-links, it was essential to creating a robust mass spectrometry and analysis method.

In this work, mass spectrometric data acquisition and further analysis of the hair protein extracts did offer additional insights about the approach we developed for mapping cross-links:

LAL detection between KRT 34 and KRT 82

Scan: 19330, Mass: 4513.2865 Da, Charge: 4+

GPC[160]RPGGGRGLRALGCLGSRSLCN[113]VGFGRPR

KN[113]HEEEVN[113]TLR

Tolerance: 0.01 Da

b+	#		#	y+
3276.672	1	K	11	
3389.731	2	N	10	1238.624
3526.790	3	H	9	1125.565
3655.833	4	E	8	988.506
3784.875	5	E	7	859.463
3913.918	6	E	6	730.421
4012.986	7	V	5	601.378
4126.045	8	N	4	502.310
4227.093	9	T	3	389.251
4340.177	10	L	2	288.203
	11	R	1	175.119

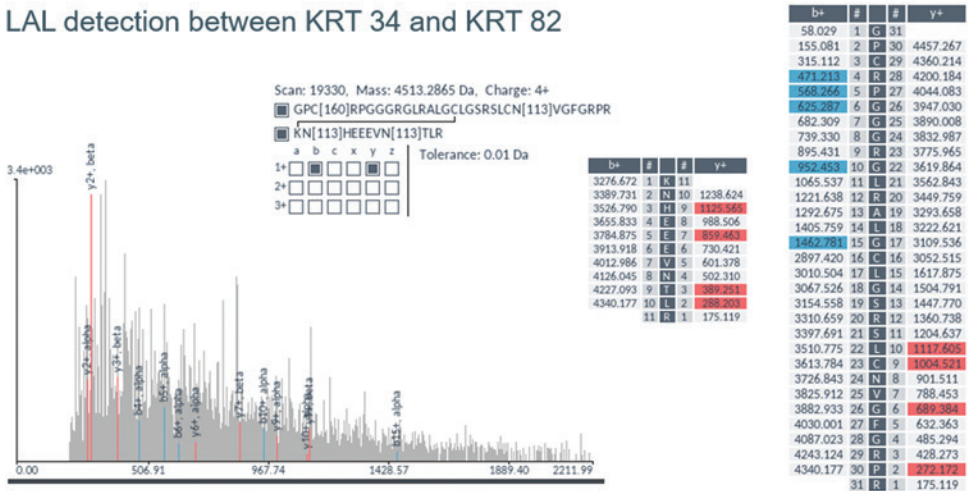| b+	#		#	y+
58.029	1	G	31	
155.081	2	P	30	4457.267
315.112	3	C	29	4360.214
471.213	4	R	28	4200.184
568.266	5	P	27	4044.083
625.287	6	G	26	3947.030
682.309	7	G	25	3890.008
739.330	8	G	24	3832.987
895.431	9	R	23	3775.965
952.453	10	G	22	3619.864
1065.537	11	L	21	3562.843
1221.638	12	R	20	3449.759
1292.675	13	A	19	3293.658
1405.759	14	L	18	3222.621
1462.781	15	G	17	3109.536
2897.420	16	C	16	3052.515
3010.504	17	L	15	1617.875
3067.526	18	G	14	1504.791
3154.558	19	S	13	1447.770
3310.659	20	R	12	1360.738
3397.691	21	S	11	1204.637
3510.775	22	L	10	1117.605
3613.784	23	C	9	1004.521
3726.843	24	N	8	901.511
3825.912	25	V	7	788.453
3882.933	26	G	6	689.384
4030.001	27	F	5	632.363
4087.023	28	G	4	485.294
4243.124	29	R	3	428.273
4340.177	30	P	2	272.172
	31	R	1	175.119

Figure 7. Mass spectrometric characterization of intermolecule lysinoalanine cross-links found between keratin 34 and keratin 82 within a tryptic digest of hair shaft-extracted proteins. The signal-to-noise level of this cross-link is very low, and interpretation of the data should be done with caution as hardly any difference between the target ions and background ions is seen.

1. First, though many cross-linked peptide ions can still not be detected due to the complexity of the sample, this study provides a first proof-of-principle that our approach, which targeted lanthionine and lysinoalanine cross-links, proved successful for mapping locations within human hair proteins. In each case, fragment ions that match both linked peptides were detected.
2. Second, the high homology in protein sequences in the keratin family complicates the identification of cross-linked proteins. We chose to report only the "first hit" protein (i.e., the protein with the highest score) during database matching. However, for some peptides it is possible that the peptide sequence of interest is not unique to just one protein.
3. Third, false positive cross-link identifications during data analysis require careful consideration. While our initial results mapped cross-links to both keratins and keratin-associated proteins, after applying stringent data analysis settings to minimize false positive identifications, in this work only linkages between keratins were confidently identified.

POSITION OF THE CROSS-LINK WITHIN THE FIBER

Next these preliminary mass spectrometric data were compared to the current models of trichocyte intermediate filament organization. Intermediate filament–forming keratin proteins can be classified into two major types: acidic type I keratins and neutral-basic type II keratins. Within the hair cortex, a type I and type II keratin form a coiled-coil heterodimer. To keep this three-dimensional coiled structure of the intermediate filament stable, protein–protein cross-links are crucial. Tetramers are formed when two of these heterodimers are clustered in an antiparallel fashion (9,11,12). Protofilaments cluster different tetramers, which end up in the intermediate filament structure (Figure 8). Three
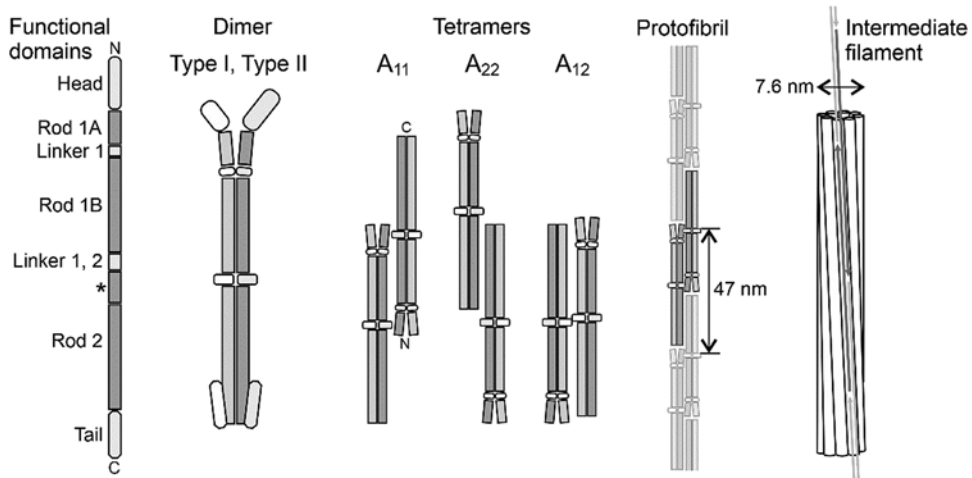

Figure 8. Schematic representation of the trichocyte keratin intermediate filament molecule. Each molecule is a heterodimer containing a type I and type II chain keratin. Tetramers are formed when two of these heterodimers are clustered in an antiparallel fashion. In these tetramers, three modes are common to all classes of intermediate filaments and involve (respectively) the approximate axial alignment of antiparallel 1B segments, antiparallel 2 segments, and antiparallel rod domains. Protofibrils cluster different tetramers, which end up in the intermediate filament structure.

different modes of molecular assembly of intermediate filaments are described. These three modes are common to all classes of intermediate filaments and involve (respectively) the approximate axial alignment of antiparallel 1B segments (model A_{11}), antiparallel 2 segments (model A_{22}), and antiparallel rod domains (model A_{12}) (Figure 8). Previous work from the Parry et al. research groups (11,13–16) do show that newly assembled protofilament molecules associate in A_{11} or A_{22} alignments within the follicle before a profound structural rearrangement occurs when the environment changes from a reducing to an oxidizing one.

We compared these filament models with our experimental data to get a better idea of what these positions of the cross-linked peptides within the protein sequence mean, and to see if our data aligned with the models of trichocyte intermediate filament organization. Important to note, though, is that this work is not intending to make any assumptions on the structural intermediate filament biology aspect.

The lanthionine cross-link found between keratin 38 and keratin 85 matches the models in a number of ways: first, proteins keratin 38 and keratin 85 are, respectively, a type I and type II keratin and could thus form a heterodimer. They are both cortex proteins. Second, the position of the cross-linked peptides (Figure 3) demonstrates that the link is made between proteins in antiparallel positioned dimers forming a tetramer, with an interaction between a head domain of one protein and a tail domain of another. This could be an indication of the suggested model A_{12} in a tetramer but could also be the packaging of adjacent tetramers. Third, the A_{12} arrangement is probably the most prevalent in a mature hair fiber as it lends itself to optimal molecular packing and radial compaction in an oxidized environment (13).

Subsequently, we researched whether the other detected lanthionine cross-links could also be explained by these models. Similar to the previous example, the cross-link between keratin 33A and keratin 74, detected in two spectra (Table I), did show an A_{12} arrangement, with peptide 1 (AA 91–108) from keratin 33A located in linker 1, while peptide 2 from keratin 74 with amino acid position 370–378 in the sequence is part of coil 2. As these two proteins do not tend to be found in the same part of the fiber (i.e., K33a is in the cortex and medulla, while K74 is found in the inner root sheet), it is still a question of where this cross-link would be formed, as both peptides are unique and thus only identified in these respective proteins. The mass spectrometric evidence of the lanthionine cross-link between keratin 34 and keratin 85, a type I and type II keratin, does show a different positioning of the cross-linked peptides. Here, one peptide is found at the N-terminus (head region), while the other peptide is found in the linker 1 region (Figure 5). This lanthionine link would thus suggest that these two proteins are more comparable to the A_{11} model, though the position of the head sequence in the intermediate filament is still questionable.

Similarities between the proposed intermediate filament models are also found when looking at the trichocyte keratin lysinoalanine cross-links within the samples. Mass spectrometric mapping of a lysinoalanine cross-link between keratin 33B and keratin 80 displays an interesting link, as both proteins are present in the medulla. The medulla, however, is known to have a heterogenous structure, and hence, although the position of the peptides of interest shows that these are somewhat comparable to model A_{22}, where the proteins are still present in antiparallel heterodimers within the filament, these may or may not form tetramers in the medulla.

Though some mass spectrometric evidence for the presence of the other detected lysinoalanine cross-links is found, manual inspection of the data (e.g., the cross-link between keratin 34

and keratin 82 [Figure 7]) shows that the peptides are detected with fragments of very low intensity, and a distinction between the fragment ions and the background signals is hard to detect. Therefore, none of the other lysinoalanine data were taken forward for further comparison against the intermediate filament models.

In conclusion, our results demonstrate here, for the first time, the mass spectrometric characterization of native cross-links within hair fibers (lanthionine and lysinoalanine), combined with their mapping within the protein sequence, and their potential mapping within the intermediate filament. These results represent the first steps toward untangling the many unknowns related to cross-links in fibers. However, caution should still be taken, as our results clearly indicate that the low abundance of the cross-linked peptides still makes this type of evaluation approach difficult, even when optimized mass spectrometric data acquisition methods are applied. These preliminary data, however, do show evidence of previous reported heterodimer compositions and do showcase the potential of the technology.

REFERENCES

(1) C. R. Robbins, "Morphological, macromolecular structure and hair growth," *Chemical and Physical Behavior of Human Hair* (Springer, Berlin, Heidelberg, 2012), pp. 1–104.

(2) D. P. Harland, R. J. Walls, J. A. Vernon, J. M. Dyer, J. L. Woods, and F. Bell, Three-dimensional architecture of macrofibrils in the human scalp hair cortex, *J. Struct. Biol.*, **185**(3), 397–404 (2014).

(3) S. Deb-Choudhury, Crosslinking between trichocyte keratins and keratin-associated proteins, *Adv. Exp. Med. Biol.*, **1054**, 173–183 (2018).

(4) A. L. Miranda-Vilela, A. J. Botelho, and L. A. Muehlmann, An overview of chemical straightening of human hair: technical aspects, potential risks to hair fibre and health and legal issues, *Int. J. Cosmet. Sci.*, **36**(1), 2–11 (2014).

(5) E. Maes, J. M. Dyer, H. J. McKerchar, S. Deb-Choudhury, and S. Clerens, Protein–protein cross-linking and human health: the challenge of elucidating with mass spectrometry, *Expert Rev. Proteomics*, 14(10), 917–929 (2017).

(6) H. J. McKerchar, S. Clerens, R. C. J. Dobson, J. M. Dyer, E. Maes, and J. A. Gerrard, Protein–protein crosslinking in food: proteomic characterisation methods, consequences and applications, *Trends in Food Science & Technology*, **86**, 217–229 (2019).

(7) M. R. Hoopmann, A. Zelter, R. S. Johnson, M. Riffle, M. J. MacCoss, T. N. Davis, and R. L. Moritz, Kojak: efficient analysis of chemically cross-linked protein complexes, *J. Proteome Res.*, 14(5), 2190–2198 (2015).

(8) L. Kall, J. D. Canterbury, J. Weston, W. S. Noble, and M. J. MacCoss, Semi-supervised learning for peptide identification from shotgun proteomics datasets, *Nat. Methods*, 4(11), 923–925 (2007).

(9) R. D. B. Fraser and D. A. D. Parry, Structural hierarchy of trichocyte keratin intermediate filaments, *Adv. Exp. Med. Biol.*, **1054**, 57–70 (2018).

(10) E. Maes, S. Clerens, J. M. Dyer, and S. Deb-Choudhury, Improved detection and fragmentation of disulphide-linked peptides, *Methods Protoc.*, 1(3), (2018).

(11) D. A. Parry, L. N. Marekov, and P. M. Steinert, Subfilamentous protofibril structures in fibrous proteins: cross-linking evidence for protofibrils in intermediate filaments, *J. Biol. Chem.*, **276**(42), 39253–39258 (2001).

(12) D. A. Parry, L. N. Marekov, P. M. Steinert, and T. A. Smith, A role for the 1A and L1 rod domain segments in head domain organization and function of intermediate filaments: structural analysis of trichocyte keratin, *J. Struct. Biol.*, **137**(1–2), 97–108 (2002).

(13) R. D. Fraser and D. A. Parry , Structural changes in the trichocyte intermediate filaments accompanying the transition from the reduced to the oxidized form, *J. Struct. Biol.*, **159**(1), 36–45 (2007).

(14) D. A. Parry and P. M. Steinert, Intermediate filaments: molecular architecture, assembly, dynamics and polymorphism, *Q. Rev. Biophys.*, **32**(2), 99–187 (1999).

(15) P. M. Steinert, L. N. Marekov, R. D. Fraser, and D. A. Parry, Keratin intermediate filament structure. crosslinking studies yield quantitative information on molecular dimensions and mechanism of assembly, *J. Mol. Biol.*, **230**(2), 436–452 (1993).

(16) H. Wang, D. A. Parry, L. N. Jones, W. W. Idler, L. N. Marekov, and P. M. Steinert, In vitro assembly and structure of trichocyte keratin intermediate filaments: a novel role for stabilization by disulfide bonding, *J. Cell Biol.*, **151**(7), 1459–1468 (2000).